Using covariates to sharpen bounds

The research reported here was partially supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D150040, and the Spencer Foundation through the "Heterogeneous Treatment Effect" project. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education or the Spencer Foundation.

Our Team

Avi Feller (Berkeley)

Jane Furey (Abt Associates)

Todd Grindal (Abt Associates)

Luke Miratrix (HGSE)

Lindsay Page (University of Pittsburgh)

Marie-Andree Somers (MDRC)

Rebecca Unterman (MDRC)

Early College High Schools

Autonomous small schools

managed by local school districts

in partnership with two- or four-year colleges

Students...

- begin taking college-level courses in ninth grade
- are expected to graduate high school with an associates' degree and/or two years of transferable college credit

Goal: improve high school graduation and college going.

Hoped-for Mechanisms: environment promotes college going; powerful teaching and learning; small school size (400 students max); positive, supportive relationships between students & staff; effective leadership

And they were studied!

The **ECHS Study** is a randomized control study.

44 lotteries conducted at

19 schools in

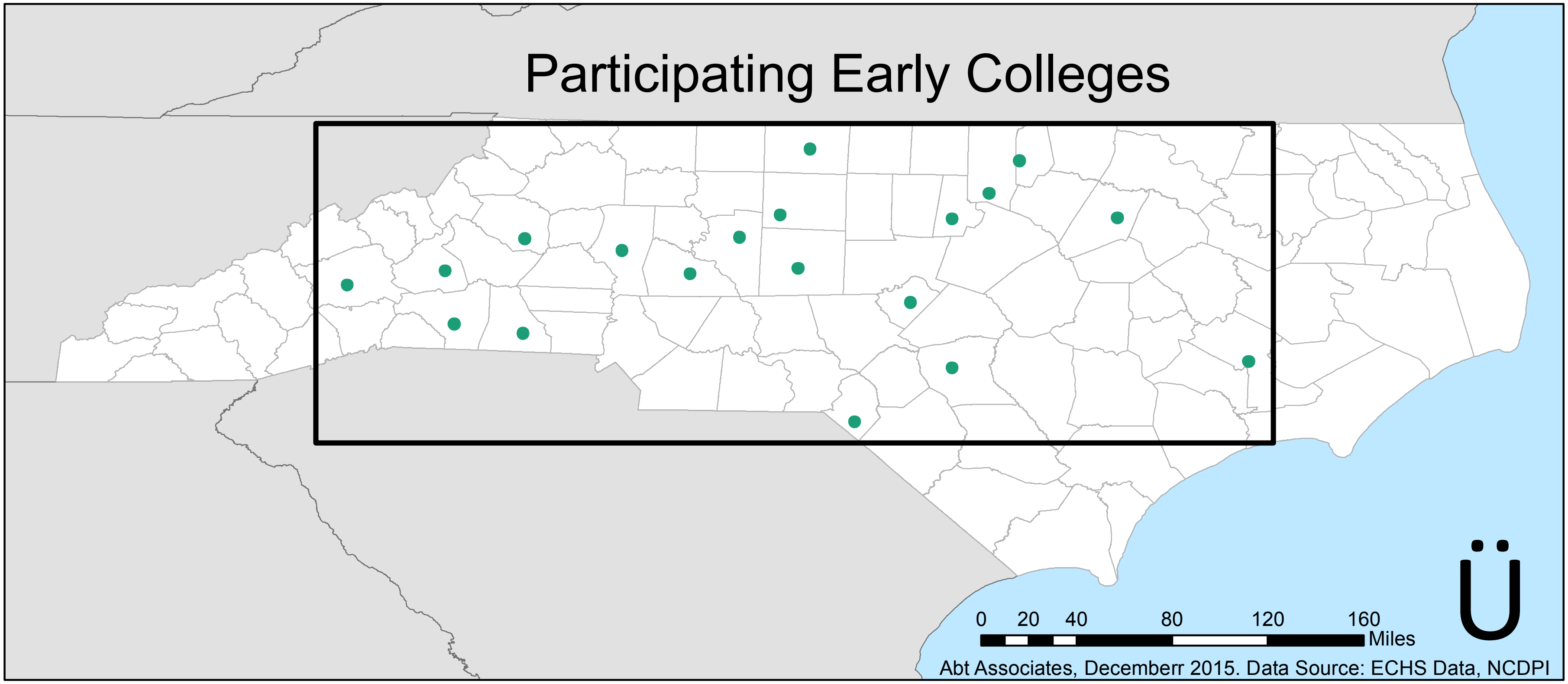
6 years

First cohort entered 9th grade in 2005

Data sources:

- North Carolina Dept. of Public Instruction (NCDPI)
- NC Community College System
- National Student Clearinghouse

Schools, scattered across the state



Each site x year combo is its own study.

We might naturally expect substantial variation in those stories.

Existing research: Positive ECHS treatment effects

Outcome	N	Unadjusted Means		Adjusted ITT Estimate
		T	C	
Grade 9 On Track	3855	97.1%	90.4%	5.9*
College Credits Earned in HS	3402	21.7	2.5	19.3*
Graduation	2941	87.3%	81.1%	4.6*
Earned Associate's Degree by Grade 14	2941	24.9%	2.5%	22.1*

*$p < 0.05$

Edmunds, J., Unlu, F., Glennie, E. , Furey, J. & Arshavsky, N. (2015). "Scaling up Early Colleges: Implementation and Impacts across Settings." Society for Research on Educational Effectiveness (accepted). Washington, DC. March 2016.

Existing research: Positive ECHS treatment effects

Outcome	N	Unadjusted Means		Adjusted ITT Estimate
		T	C	
Grade 9 On Track	3855	97.1%	90.4%	5.9*
College Credits Earned in HS	3402	21.7	2.5	19.3*
Graduation	2941	87.3%	81.1%	4.6*
Earned Associate's Degree by Grade 14	2941	24.9%	2.5%	22.1*

*$p < 0.05$

Our focus

Edmunds, J., Unlu, F., Glennie, E. , Furey, J. & Arshavsky, N. (2015). "Scaling up Early Colleges: Implementation and Impacts across Settings." Society for Research on Educational Effectiveness (accepted). Washington, DC. March 2016.

Our principal strata: quality of counterfactual school

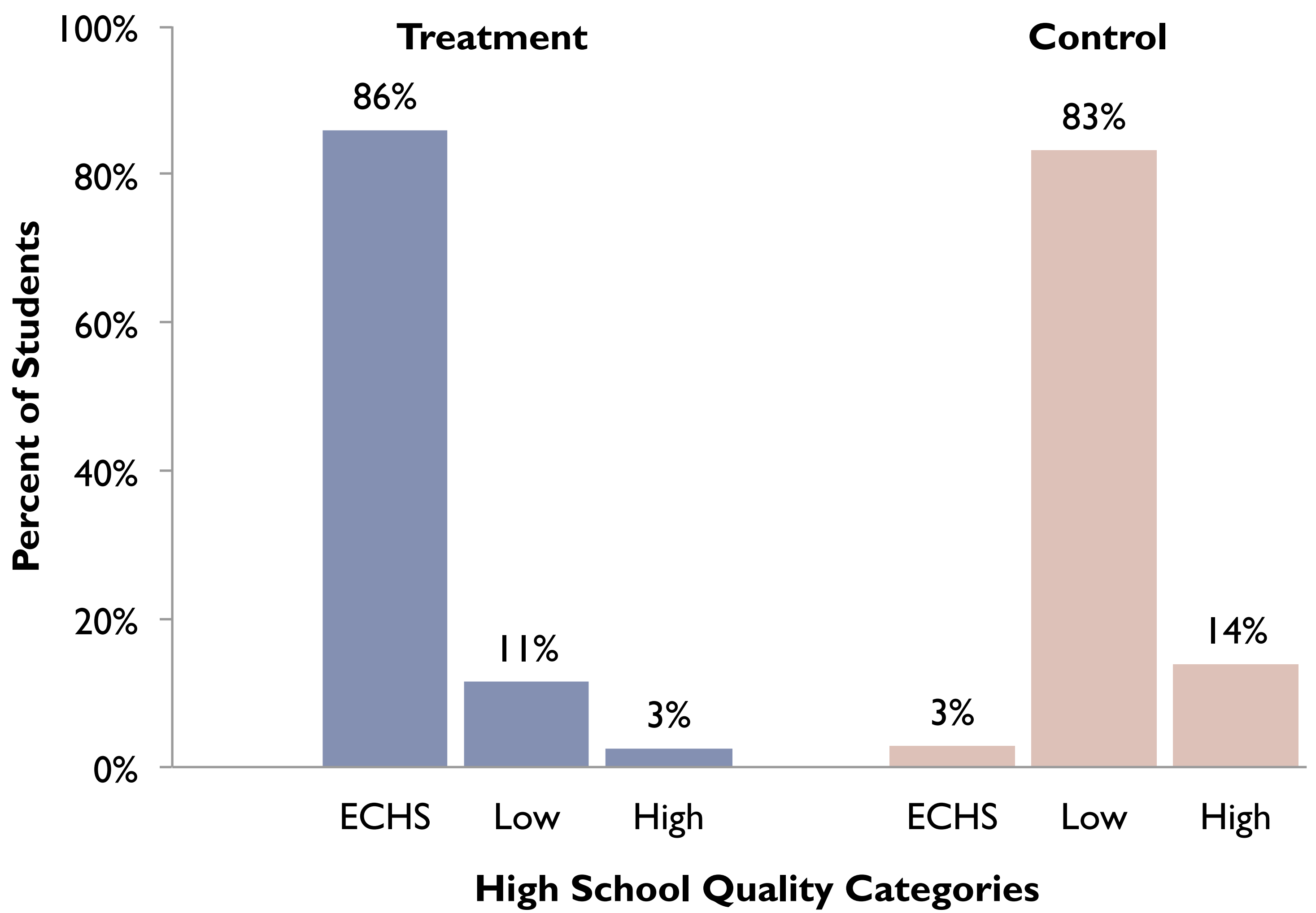
There are three types of school:

- **ECHS** (E) - the “treatment” schools
- **High-quality** (HQ) - high schools that are rated as being of good quality.
- **Low-quality** (LQ) - high schools that are not so rated.

Our Research Question

To what extent do the effects of Early College High Schools (ECHSs) vary according to the quality of the school students would have otherwise attended?

Students attend different types of high schools



Most students are on track in 9th grade

	On Track	Not On Track
Treatment	94%	6%
Control	88%	12%
Total	91%	9%

Remember this tidbit:

Being near boundaries increases the chance of bounds being viable.

The school-attendance behavior gives our boxes

If we assign to **Treatment**, they go to...

If we assign to **Control**, they go to...

	ECHS	Low-Quality School	High-Quality School
ECHS			
Low-Quality School			
High-Quality School			

Next, eliminate or structure boxes

If we assign to **Control**, they go to...

If we assign to **Treatment**, they go to...	ECHS	Low-Quality School	High-Quality School
ECHS	EAT	LQC	HQC
Low-Quality School		LQAT	
High-Quality School			HQAT

Next, eliminate or structure boxes

If we assign to **Control**, they go to...

If we assign to **Treatment**, they go to...

	ECHS	Low-Quality School	High-Quality School
ECHS	EAT	LQC	HQC
Low-Quality School	✗	LQAT	
High-Quality School	✗		HQAT

Defiers

Next, eliminate or structure boxes

If we assign to **Control**, they go to...

If we assign to **Treatment**, they go to...

	ECHS	Low-Quality School	High-Quality School
ECHS	EAT	LQC	HQC
Low-Quality School	✗	LQAT	✗
High-Quality School	✗	✗	HQAT

Irrelevant Alternatives

Defiers

What we estimate, and what is mixed up

To get our **complier average treatment effects** we need:

- The mean outcomes of our LQC and HQC groups under control
- The mean outcomes of our LQC and HQC groups under treatment

For **control side:**

- Our LQCs are mixed with our LQATs, but we see our LQATs on the treatment side
- Same for our HQCs

For **treatment side:**

- Our LQCs and HQCs and EATs are all mixed up!
- We can see EATs on the control side, but we have no way to separate treated LQCs and HQCs
- We turn to bounds to assess what possible mean outcomes there could be for these two groups.

Bounds and always-takers

Under the **exclusion restriction** we can observe Always-Takers and know their mean outcome under either treatment arm

This means we can “subtract them out.”

The presence of always-takers does not impact the bounds, only the estimation of the bounds.

Example of "Subtracting Out" to get the Control LQC mean

$\bar{Y}_{1 l_q}$ ← we see this directly in our data

$$\bar{Y}_{0 l_q} = \frac{\pi_{l_q c}}{\pi_{l_q c} + \pi_{l_q at}} \mu_{0 l_q c} + \frac{\pi_{l_q at}}{\pi_{l_q c} + \pi_{l_q at}} \mu_{0 l_q at}$$

$$\Rightarrow \mu_{0 l_q c} = \frac{\pi_{l_q c} + \pi_{l_q at}}{\pi_{l_q c}} \bar{Y}_{0 l_q} - \frac{\pi_{l_q at}}{\pi_{l_q c}} \mu_{0 l_q at}$$

$$= \frac{p_{0 l_q}}{p_{1 l_q} - p_{0 l_q}} \bar{Y}_{0 l_q} - \frac{p_{1 l_q}}{p_{1 l_q} - p_{0 l_q}} \bar{Y}_{1 l_q}$$

Our problem (without always takers)

We have

- Low-Quality Compliers (LQC) and
- High-Quality Compliers (HQC)

We see

- Mean outcomes and proportions under the control condition
- One amorphous blob under the treatment condition

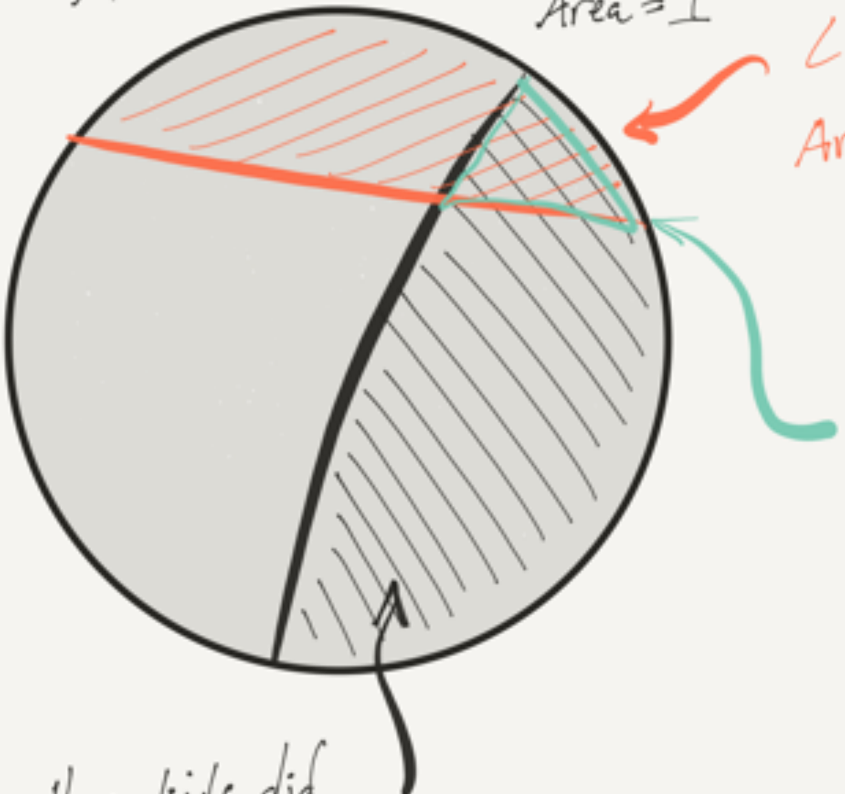
We want

- The mean outcome under the treatment condition

Calculating the bound (geometrically)

All our HQC + LQC

Area = 1



these are LQC kids

Area $= \dfrac{\pi_{lqc}}{\pi_{hqc} + \pi_{lqc}} \equiv f_{lqc}$

these kids did well ($Y=1$)

Area $= \bar{Y}_{1e}$

(assuming no EATs)

Let $q \equiv$ fraction of kids who are LQC <u>and</u> $Y=1$. $\quad \mu_{1lqc} = q / f_{lqc}$

How big? $\min(f_{lqc}, \bar{Y}_{1e})$

How small? $\max(0, (1-\bar{Y}_{1e}) - (1-f_{lqc}))$

$= \max(0, f_{lqc} - \bar{Y}_{1e})$

Finally $\mu_{1lqc} \le \min(1, \bar{Y}_{1e}/f_{lqc})$

$\mu_{1lqc} \ge \max(0, 1 - \bar{Y}_{1e}/f_{lqc})$

Our final bounds

We have

$$\mu^{low}_{lqc,1} - \mu_{lqc,0} \leq ITT_{lq} \leq \mu^{high}_{lqc,1} - \mu_{lqc,0}$$

with

$$\mu^{low}_{lqc,1} = \max\left(0, \frac{1}{\pi_{lqc}}\bar{Y}_{1e} - \frac{1-\pi_{lqc}}{\pi_{lqc}}\right)$$

$$\mu^{high}_{lqc,1} = \min\left(\frac{1}{\pi_{lqc}}\bar{Y}_{1e}, 1\right)$$

$$\mu_{lqc,0} = \bar{Y}_{0lq}$$

Our final bounds in terms of observed quantities

We have

$$\mu_{lqc,1}^{low} - \mu_{lqc,0} \leq ITT_{lq} \leq \mu_{lqc,1}^{high} - \mu_{lqc,0}$$

with

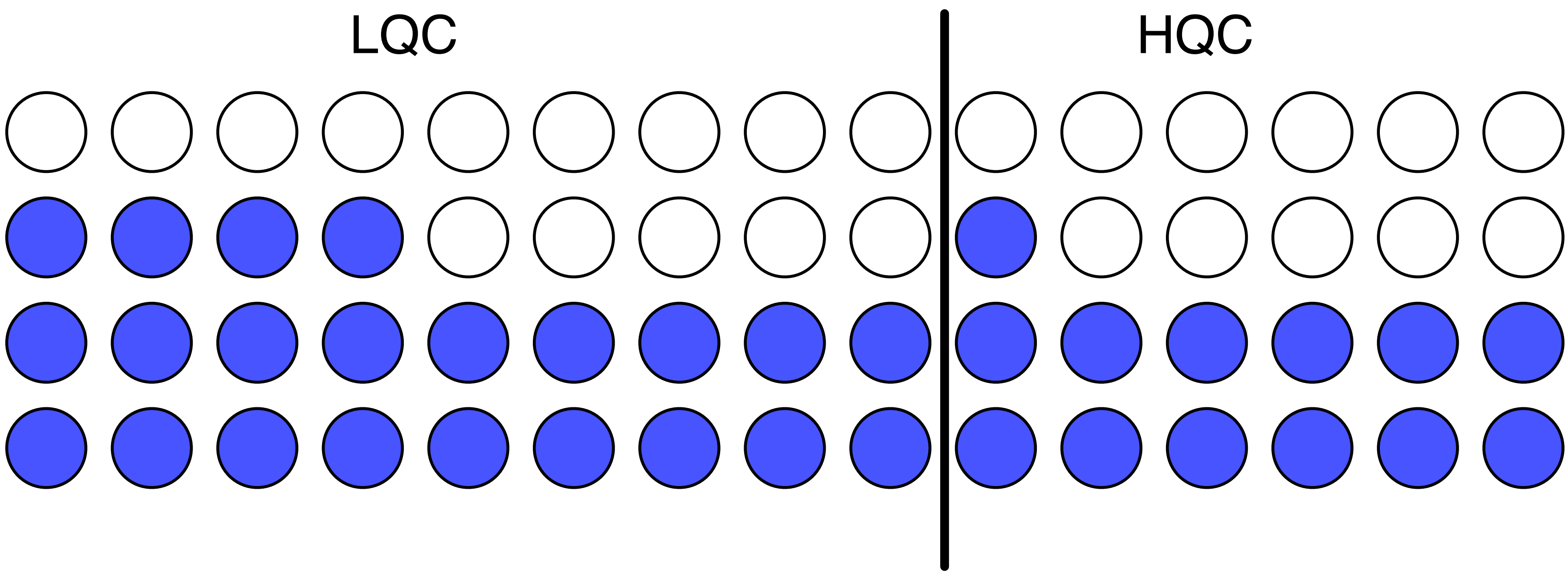
$$B_{lqc} = \frac{1}{p_{0lq}} \bar{Y}_{1e}$$

$$\mu_{lqc,1}^{low} = \max\left(0, B_{lqc} - \frac{1 - p_{0lq}}{p_{0lq}}\right)$$

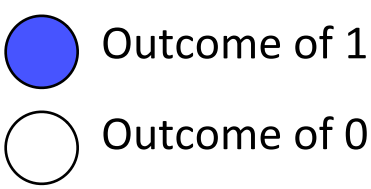
$$\mu_{lqc,1}^{high} = \min\left(B_{lq}, 1\right)$$

$$\mu_{lqc,0} = \bar{Y}_{0lq}$$

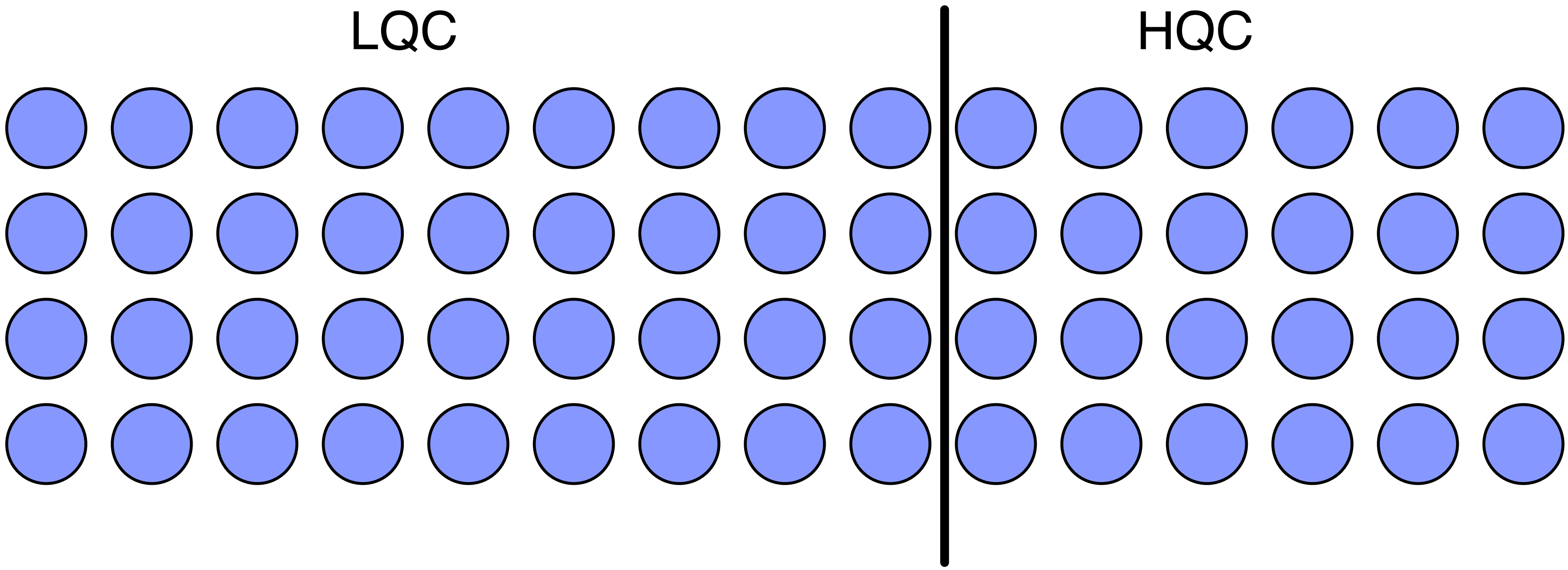
A hypothetical population



This is a "perfect information" picture.
We don't actually know which people are on which side.

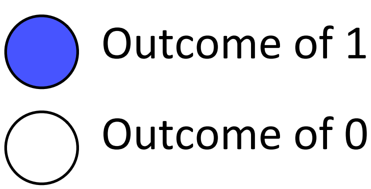


We see a mixture

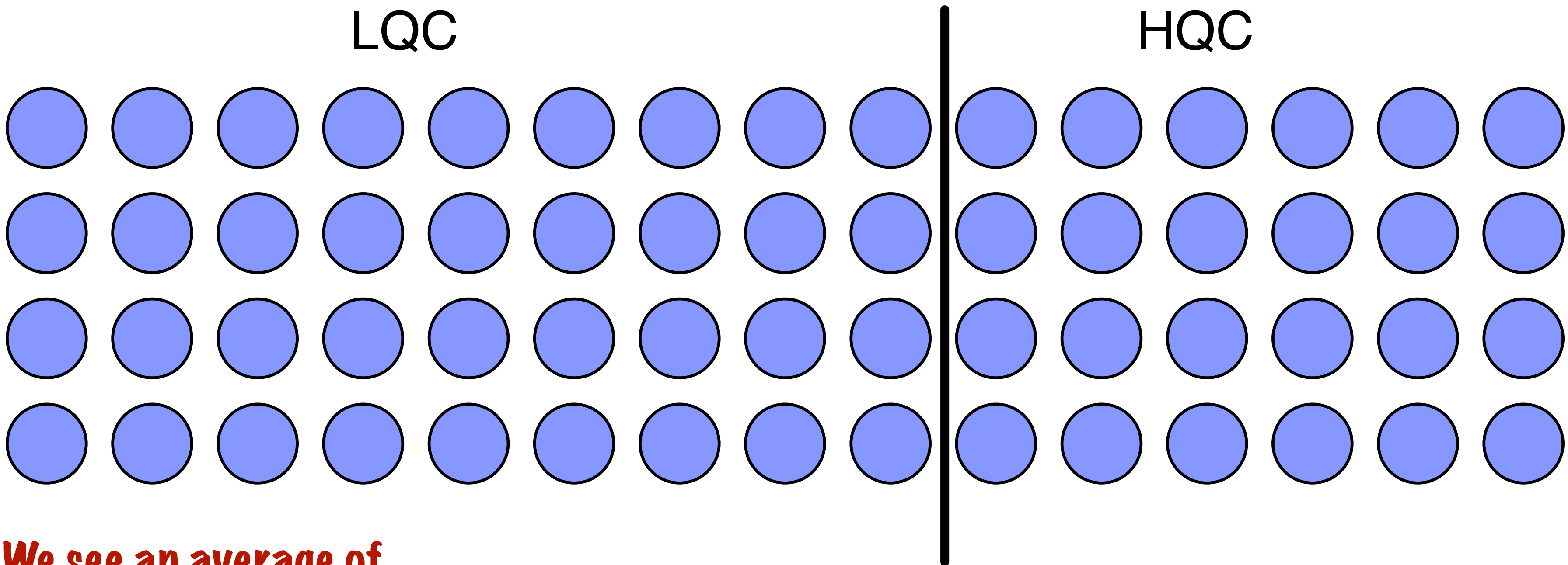


Since we do not know who is a LQC or HQC, we could imagine them all being an average outcome...

This would give an overall, pooled estimate.



We see a mixture



We see an average of 35/60 = 58%

Proportion of LQC is 36/60 = 60%

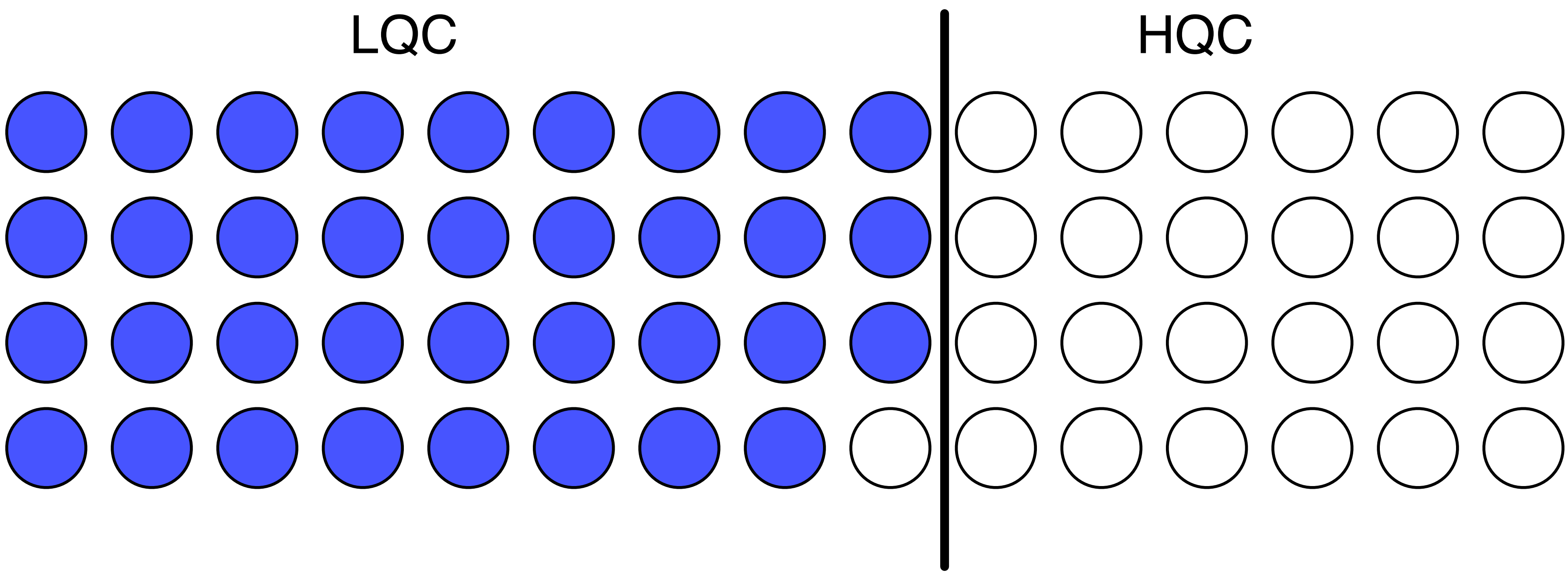
Since we do not know who is a LQC or HQC, we could imagine them all being an average outcome...

This would give an overall, pooled estimate.

Outcome of 1

Outcome of 0

Or we can think "worst case"



Here we try and see how well of the LQC could be, best case scenario.

This is also how badly off the HQC could be. Their worst case scenario.

Outcome of 1

Outcome of 0

Our actual results

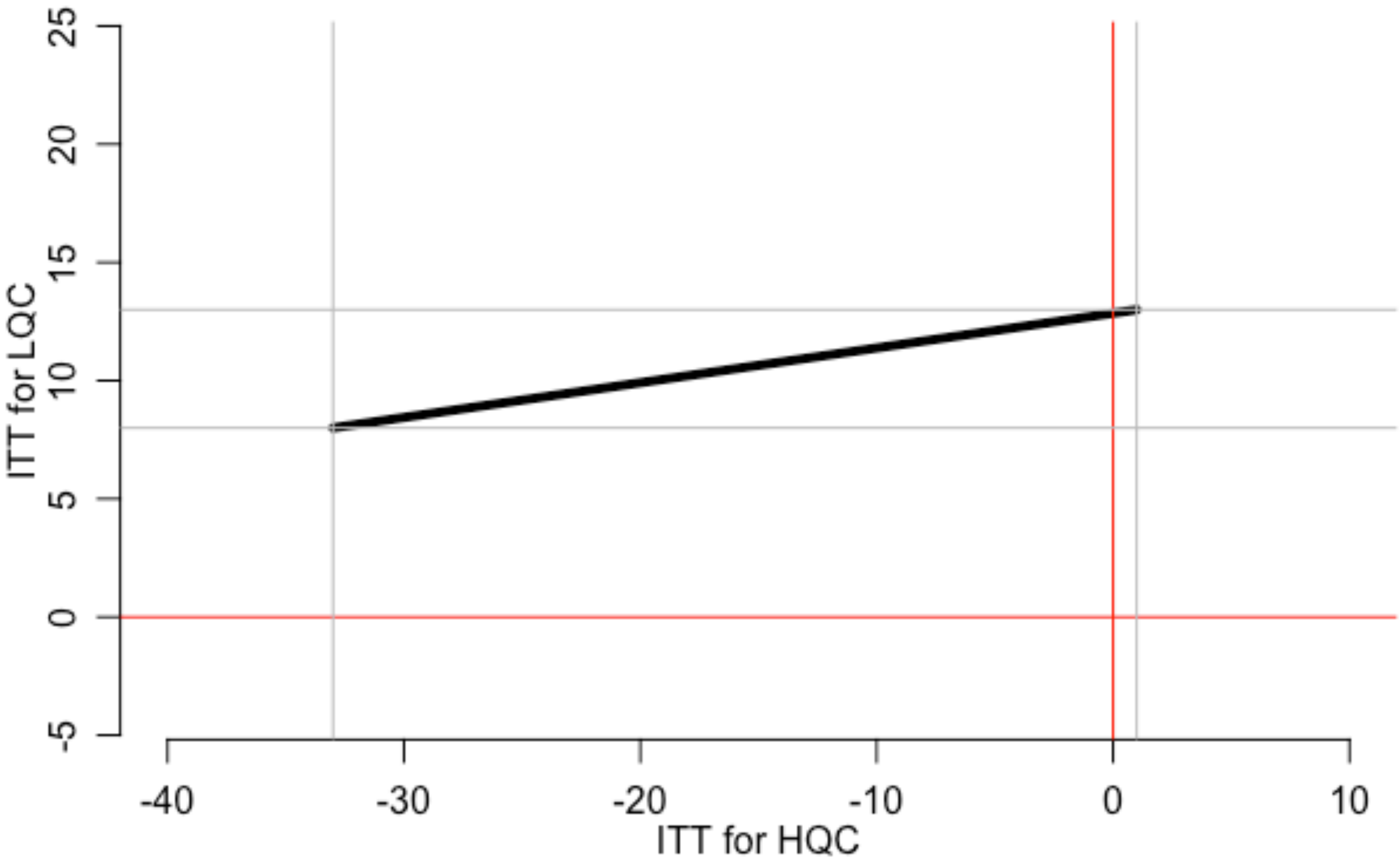
ECHS Always Takers EATs	Low-quality compliers LQCs	Low-quality compliers LQCs	Low-quality Always Takers LQATs	High-quality Always Takers HQATs
p = 3% Yt = Yc = 100%	p = 72% Yt = (95% - 100%) Yc = 87%	p = 72% Yt = (95% - 100%) Yc = 87%	p = 12% Yt = Yc = 84%	p = 3% Yt = Yc = 90%

To get our estimates, we subtract the two means:

$$ITT_{lqc} = +8 - +13\% \text{ effect}$$

$$ITT_{hqc} = -33 - +1\% \text{ effect}$$

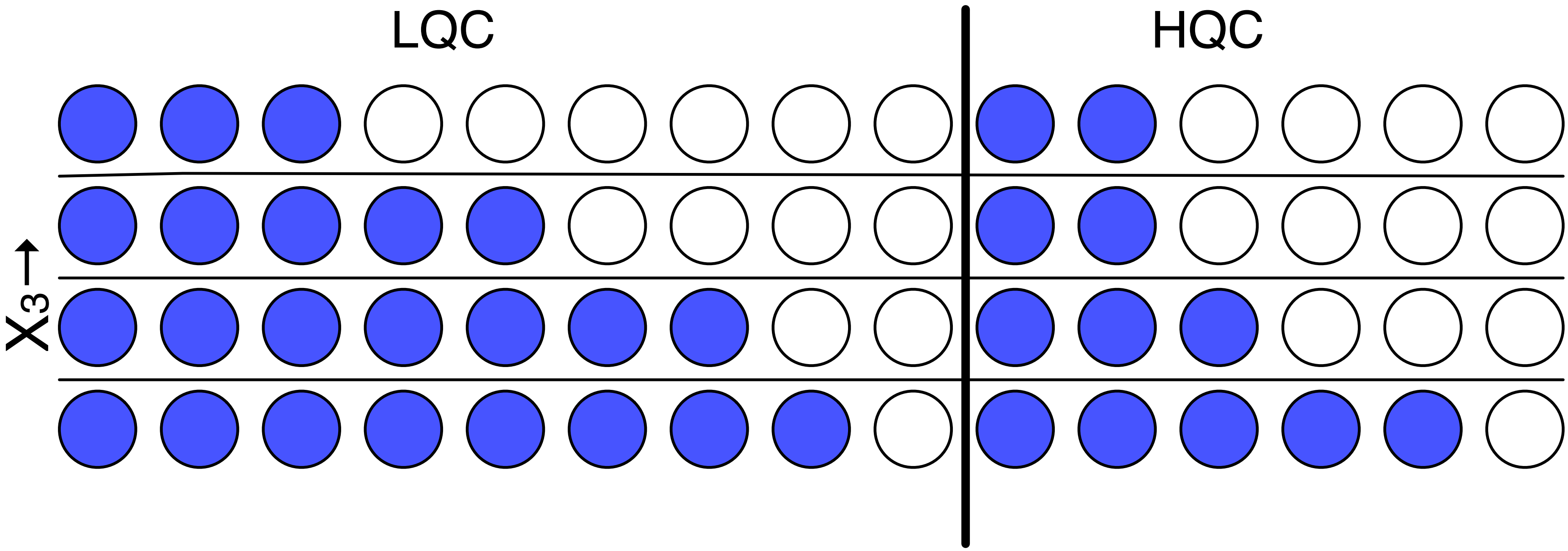
The tradeoff of treatment effects



Comparing different stratifications

	High Quality Compliers				Low Quality Compliers			
	Control Mean	Bounds on Tx Mean	Impact (ITT)	Bounds width	Control Mean	Bounds on Tx Mean	Impact (ITT)	Bounds width
No Strat	98.8	65.5 - 100	-33.3 - 1.2	34.5	86.6	94.5 - 100	7.9 - 13.4	5.5
by achievment	98.3	76.7 - 100	-21.6 - 1.7	23.3	86.6	94.5 - 100	7.9 -13.4	5.5
by middle school	98.4	86.8 - 100	-11.6 - 1.6	13.2	87	94.5 - 96.9	7.5 - 9.9	2.4
by both	97.8	87.2 - 100	-10.6 - 2.2	12.8	86.9	94.2 - 95.9	7.3 - 9.0	1.7

Now say we had a "prognostic covariate"



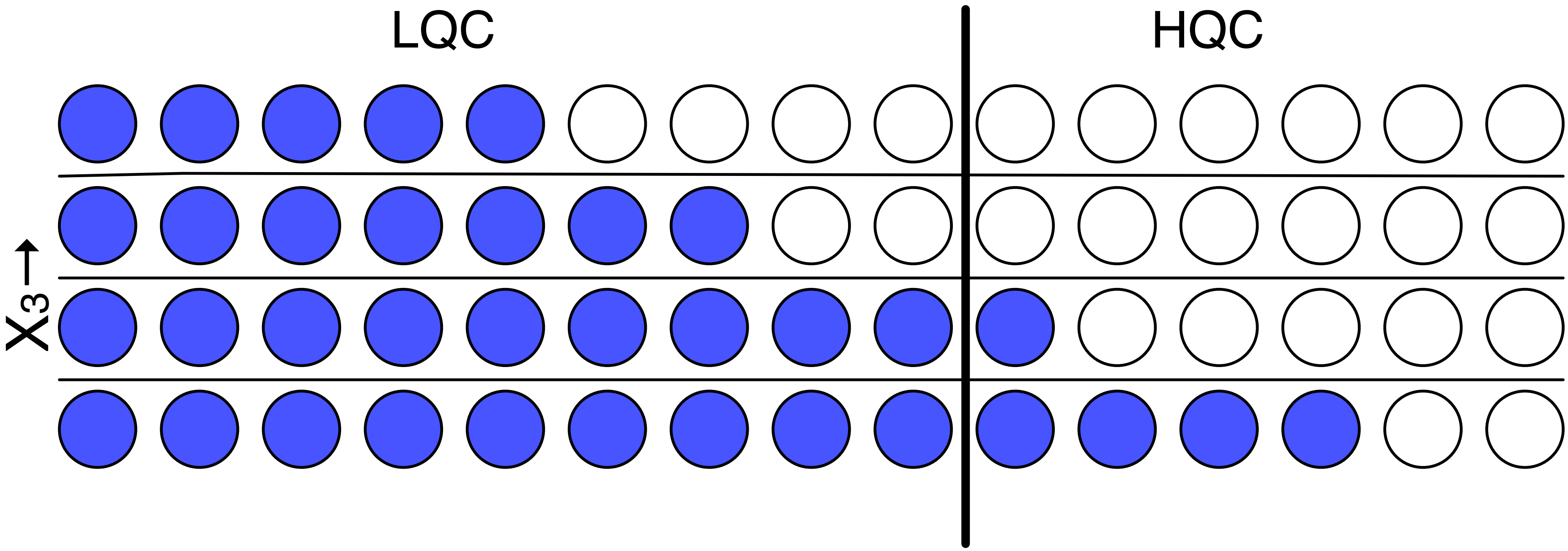
For ECHS we used grade 8 test scores.

We use it to create four equal sized slices of our population.

Inside each slice we can identify the proportion of LQC and HQC (in this example the proportions turn out to be the same)

We then do our same bounding exercise within each slice...

Create our worst case (for HQC) by calling as many successes as we can LQCs



Our first two slices are just as bad as our original for HQC (a 0 lower bound)

But our next two are better!

We then combine (average) to get an overall bound.

Some calculations

HQC: $0 + 0 + \frac{1}{6}\frac{1}{4} + \frac{4}{6}\cdot\frac{1}{4} = \frac{5}{24} \cong 0.21$

LQC: $\frac{5}{9}\frac{1}{4} + \frac{7}{9}\frac{1}{4} + 1\cdot\frac{1}{4} + 1\cdot\frac{1}{4} = \frac{30}{36} \cong 0.83$

so HQC lower bound → 0.21 from $\underline{\underline{0}}$

LQC upper bound → 0.83 from $\underline{\underline{0.97}}$

Some calculations

$$HQC:\quad 0 + 0 + \frac{1}{6}\frac{1}{4} + \frac{4}{6}\cdot\frac{1}{4} = 5/24 \cong 0.21$$

$$LQC:\quad \frac{5}{9}\frac{1}{4} + \frac{7}{9}\frac{1}{4} + 1\cdot\frac{1}{4} + 1\cdot\frac{1}{4} = \frac{30}{36} \cong 0.83$$

So HQC lower bound $\rightarrow$ 0.21 from $\underline{\underline{0}}$

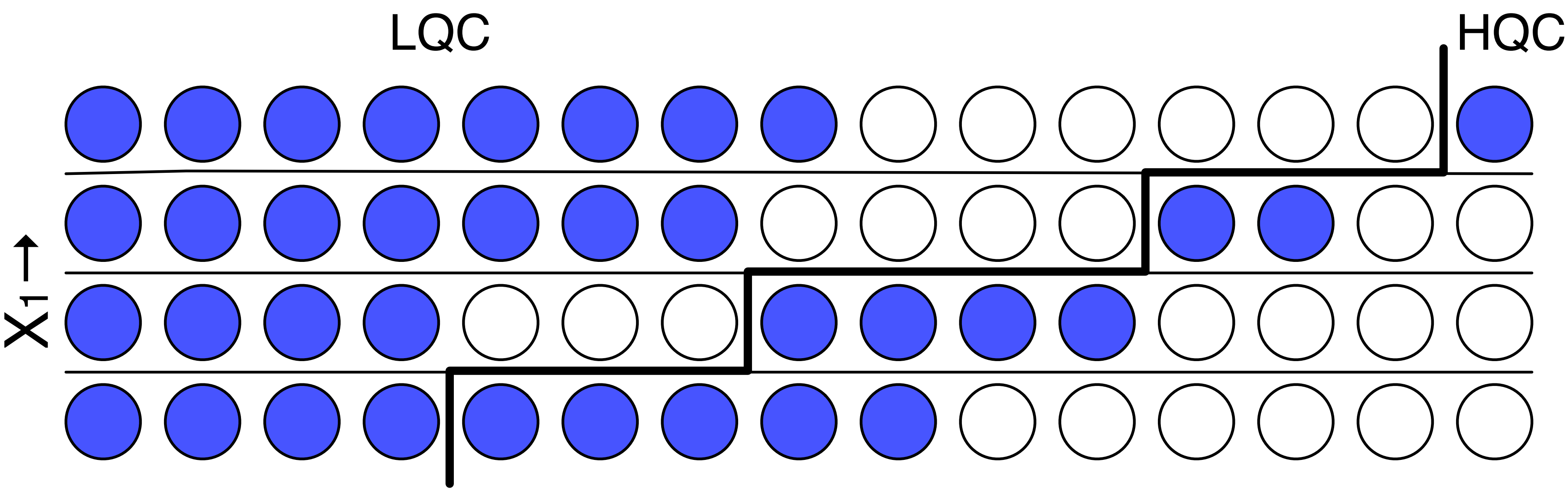
LQC upper bound $\rightarrow$ 0.83 from $\underline{\underline{0.97}}$

This is some serious improvement!

(There is nothing like an artificial example.

What other kinds of variables might we consider for slicing up our data?

Here we have a “propensity score” covariate that predicts complier type.

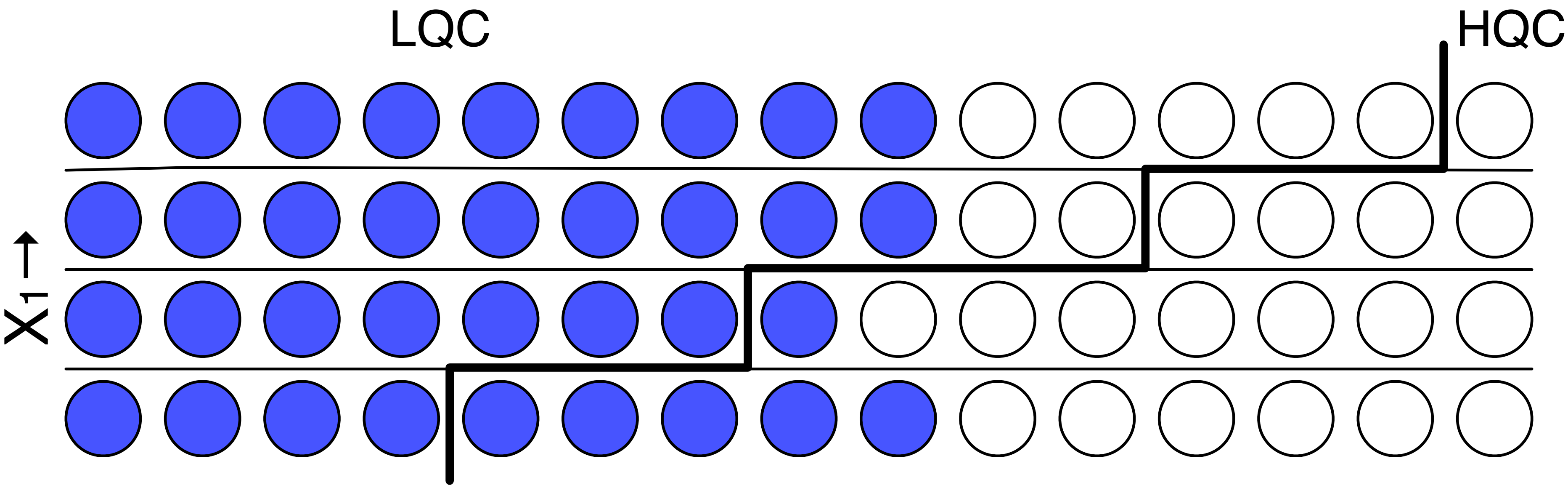


Slice based on a covariate that is predictive of compliance type.

High values of X_1 indicate a higher proportion of LQC students.

In the ECHS study, this was the proportion of a student’s middle school cohort who went to a LQ high school.

We again look at worst-case within each slice of our sample



Again, our bounds are tightened in some of our slices and not others.

But notice how the relatively bigger slices are the ones with the tighter bounds!

Averaging here is a bit trickier... Let's talk about it

Some further calculations

For HQC we weight by # folks: $N_{HQC} = 24$

$$\mu \geq 0\frac{1}{24} + 0\frac{4}{24} + \frac{1}{8}\frac{8}{24} + \frac{5}{11}\frac{11}{24} = \frac{6}{24} \Rightarrow \frac{3}{12} = 0.25$$

LQC

$$\mu \leq \underbrace{\frac{9}{14}\frac{14}{36}}_{\text{big strata, tight bound}} + \frac{9}{11}\frac{11}{36} + 1\frac{7}{36} + 1\frac{4}{36} = 29/36 \hat{=} 0.81$$

Comparisons
No bounds:
HQC > 0   LQC < 0.97

Prognostic bounds:
HQC > 0.21   LQC < 0.83

Some further calculations

For H... 24

$\mu \geq$ ... $\frac{3}{12} = 0.25$

LQC

$\mu \leq \frac{9}{14}$ ... 81

big

tight bound

Full disclosure:
I kinda cheated. I tweaked the variables so that this covariate was better by a little bit than last.

(More exciting.)

But this does raise a question...

HQC > 0 LQC < 0.97

Prognostic bounds:
HQC > 0.21 LQC < 0.83

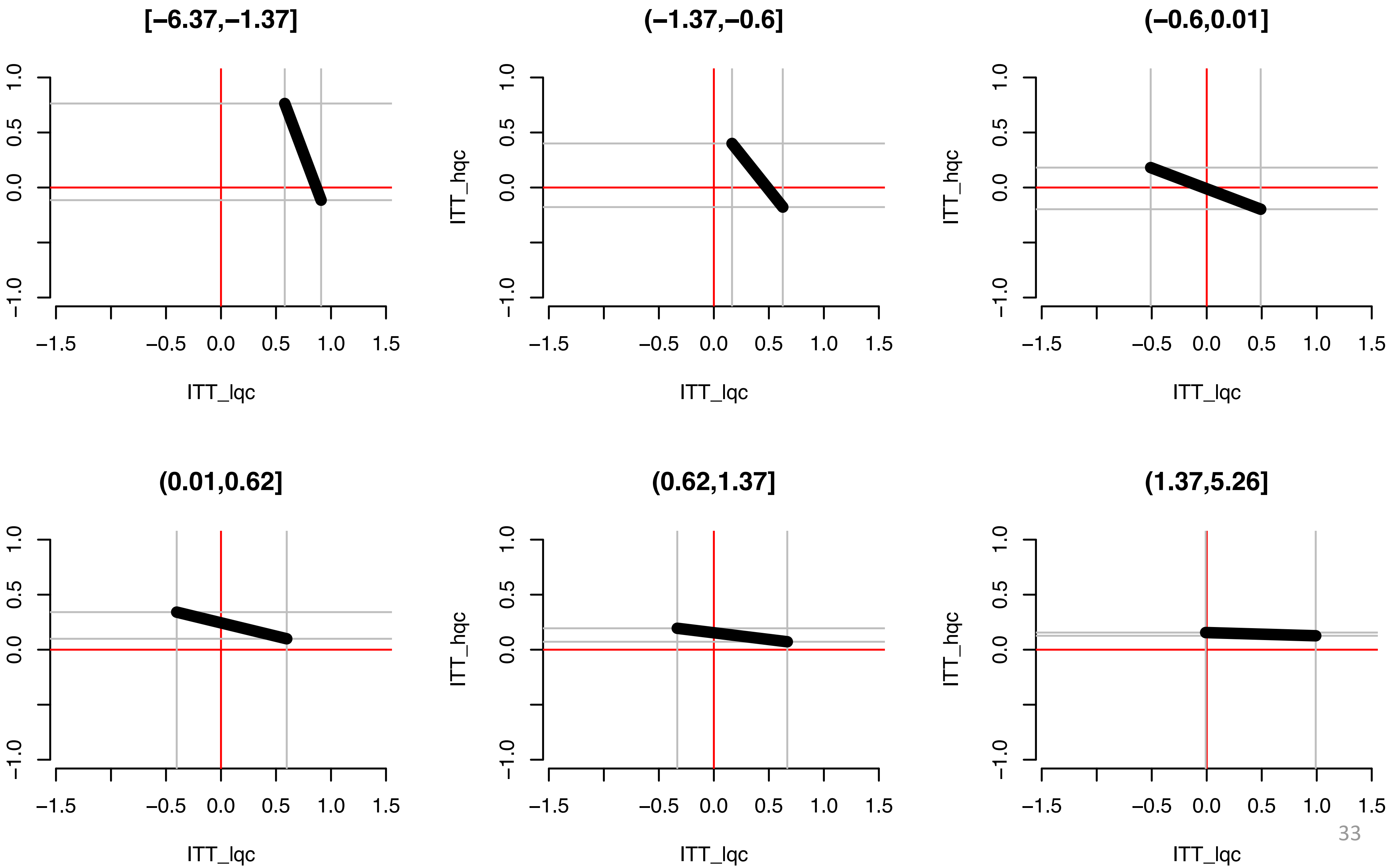
Choices, choices...

How do we decide what to stratify on?

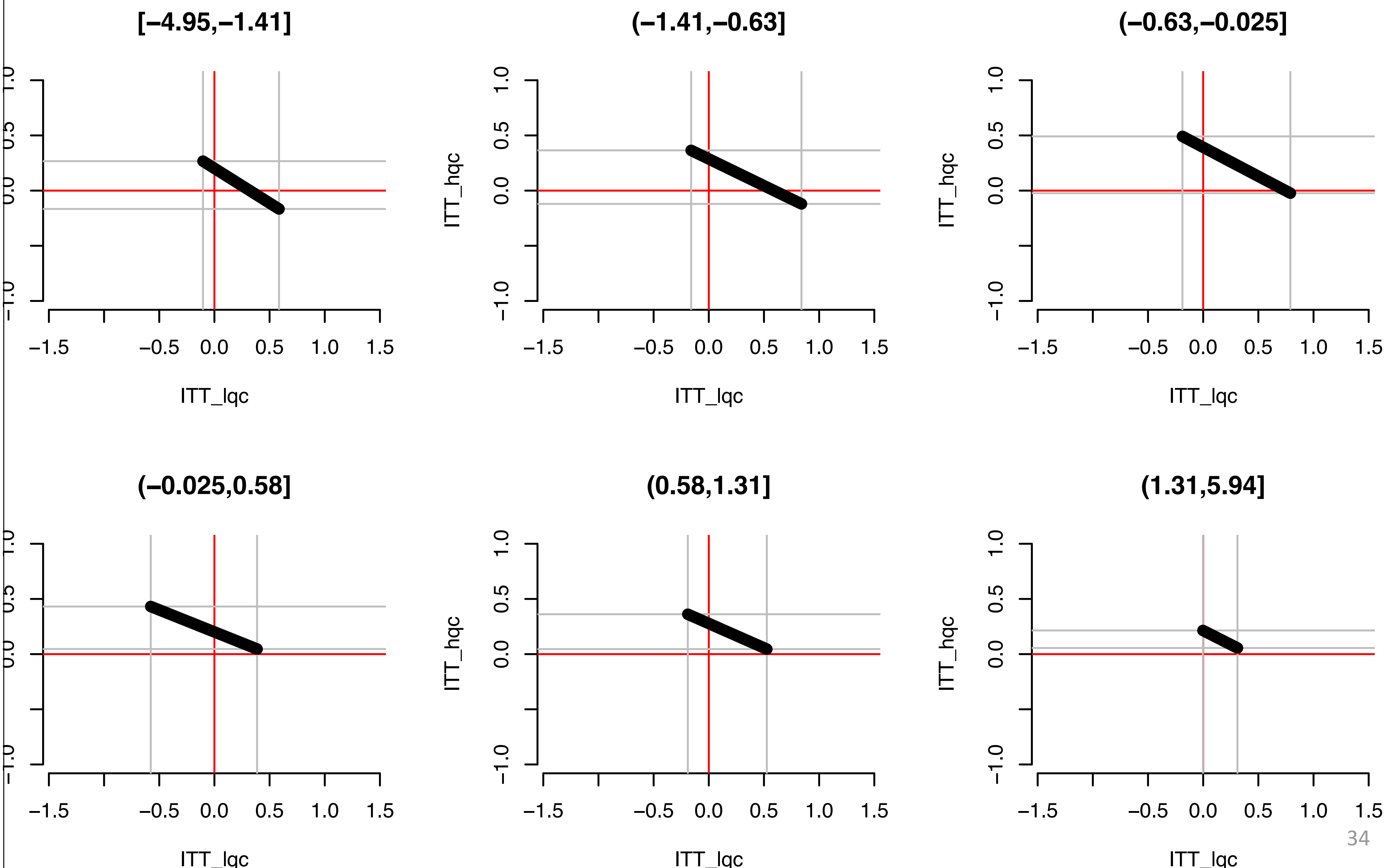
Two primary choices:

- "prognostic score" - covariate predictive of outcome.
- "propensity score" - covariate predictive of compliance status.

Stratifying by Propensity Score
Different strata are differently effective for LQC and HQC



Stratifying by Prognostic Score: Tighter bounds at the margins



A simulation study

We will generate fake data indexed by some latent traits that govern school quality, compliance, and outcome.

We will then stratify these covariates and look at the resulting bound widths.

This could shed light on when covariates are powerful, or not.

The data generation process

$$U_{1i}, U_{2i}, U_{3i} \sim N(0,1) \text{ (all independent)}$$

$$H_i = \begin{cases} hq & U_{1i} > \Phi^{-1}(\gamma_0) \\ lq & else \end{cases}$$

$$S_i(0) = \begin{cases} e & U_{2i} > \Phi^{-1}(1 - \pi_{eat}) \\ H_i & else \end{cases}$$

$$S_i(1) = \begin{cases} e & U_{2i} > \Phi^{-1}(\gamma_1(H_i)) \\ H_i & else \end{cases}$$

$$Y_i(z) = logit^{-1}(\beta U_3 + \omega(S_i(z)))$$

The data generation process

$$U_{1i}, U_{2i}, U_{3i} \sim N(0,1) \text{ (all independent)}$$

$$H_i = \begin{cases} hq & U_{1i} > \Phi^{-1}(\gamma_0) \\ lq & else \end{cases}$$

$$S_i(0) = \begin{cases} e & U_{2i} > \Phi^{-1}(1 - \pi_{eat}) \\ H_i & else \end{cases}$$

$$S_i(1) = \begin{cases} e & U_{2i} > \Phi^{-1}(\gamma_1(H_i)) \\ H_i & else \end{cases}$$

$$Y_i(z) = logit^{-1}(\beta U_3 + \omega(S_i(z)))$$

The gamma and omega parameters govern marginal means, etc. They are functions of desired strata proportions and effect sizes. Beta governs strength of U_3 and the outcomes.

What we observe

Generate a population as above

We then observe *noisy versions* of the latent variables:

$$X_{it} = U_{it} + \varepsilon_{it} \text{ with } \varepsilon_{it} \sim N(0, \sigma_t^2)$$

We then randomize our units to treatment and control, and observe the relevant potential outcomes to get our data:

	X1	X2	X3	Z	S	Yobs
1	-2.27	-1.68	-1.38	0	lq	0
2	0.56	-0.17	-0.55	1	e	1
3	1.21	1.38	0.59	0	hq	0
4	1.01	2.15	1.58	1	e	1

What we observe

Generate a population as above

We then observe *noisy versions* of the latent variables:

$$X_{it} = U_{it} + \varepsilon_{it} \text{ with } \varepsilon_{it} \sim N(0, \sigma_t^2)$$

We then randomize our units to treatment and control, and observe the relevant potential outcomes to get our data:

Sample data

	X1	X2	X3	Z	S	Yobs
1	-2.27	-1.68	-1.38	0	lq	0
2	0.56	-0.17	-0.55	1	e	1
3	1.21	1.38	0.59	0	hq	0
4	1.01	2.15	1.58	1	e	1

What we observe

Generate a population as above

We then observe *noisy versions* of the latent variables:

We vary these to get more or less useful Xs.

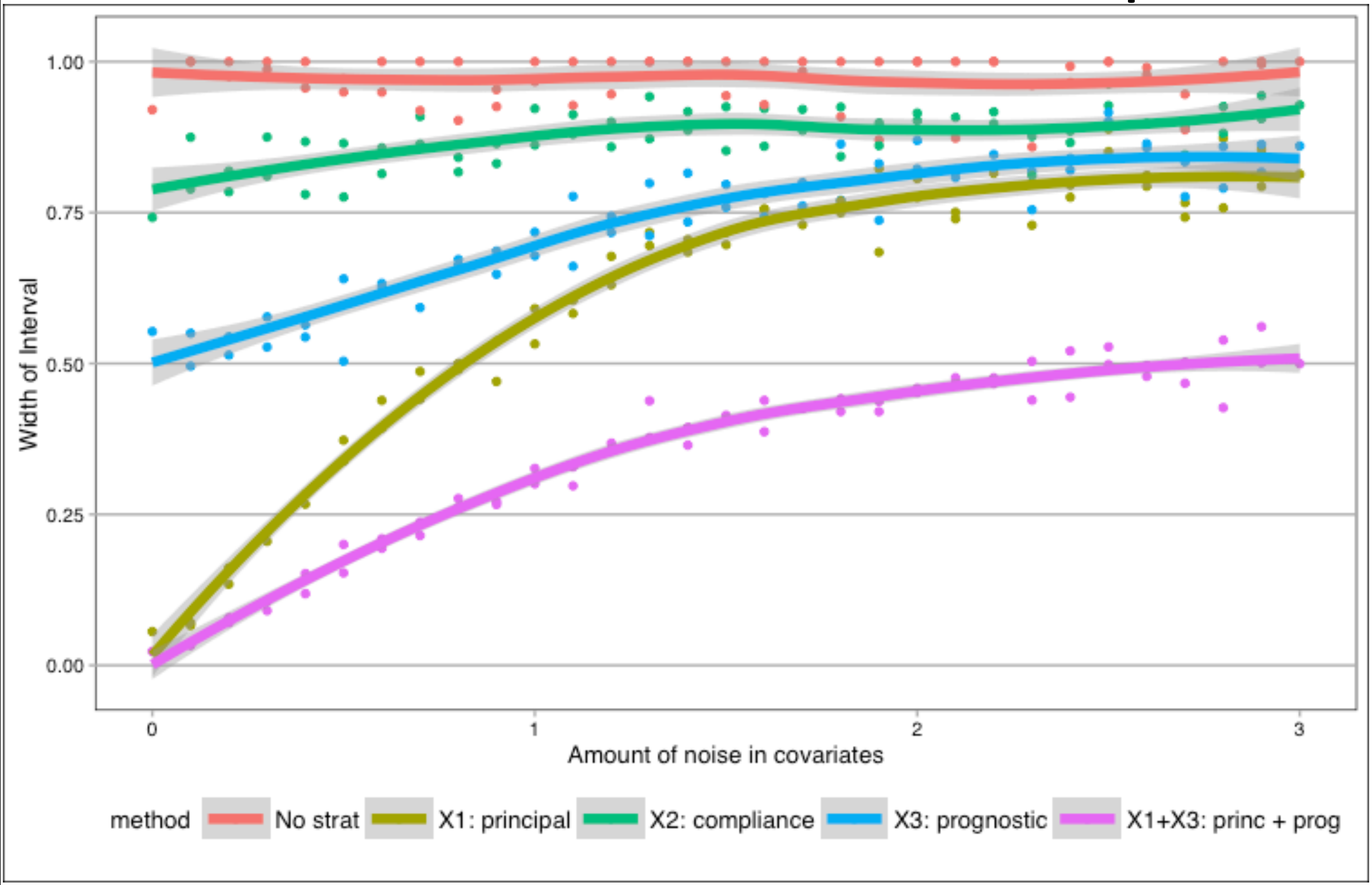
$$X_{it} = U_{it} + \varepsilon_{it} \text{ with } \varepsilon_{it} \sim N(0, \sigma_t^2)$$

We then randomize our units to treatment and control, and observe the relevant potential outcomes to get our data:

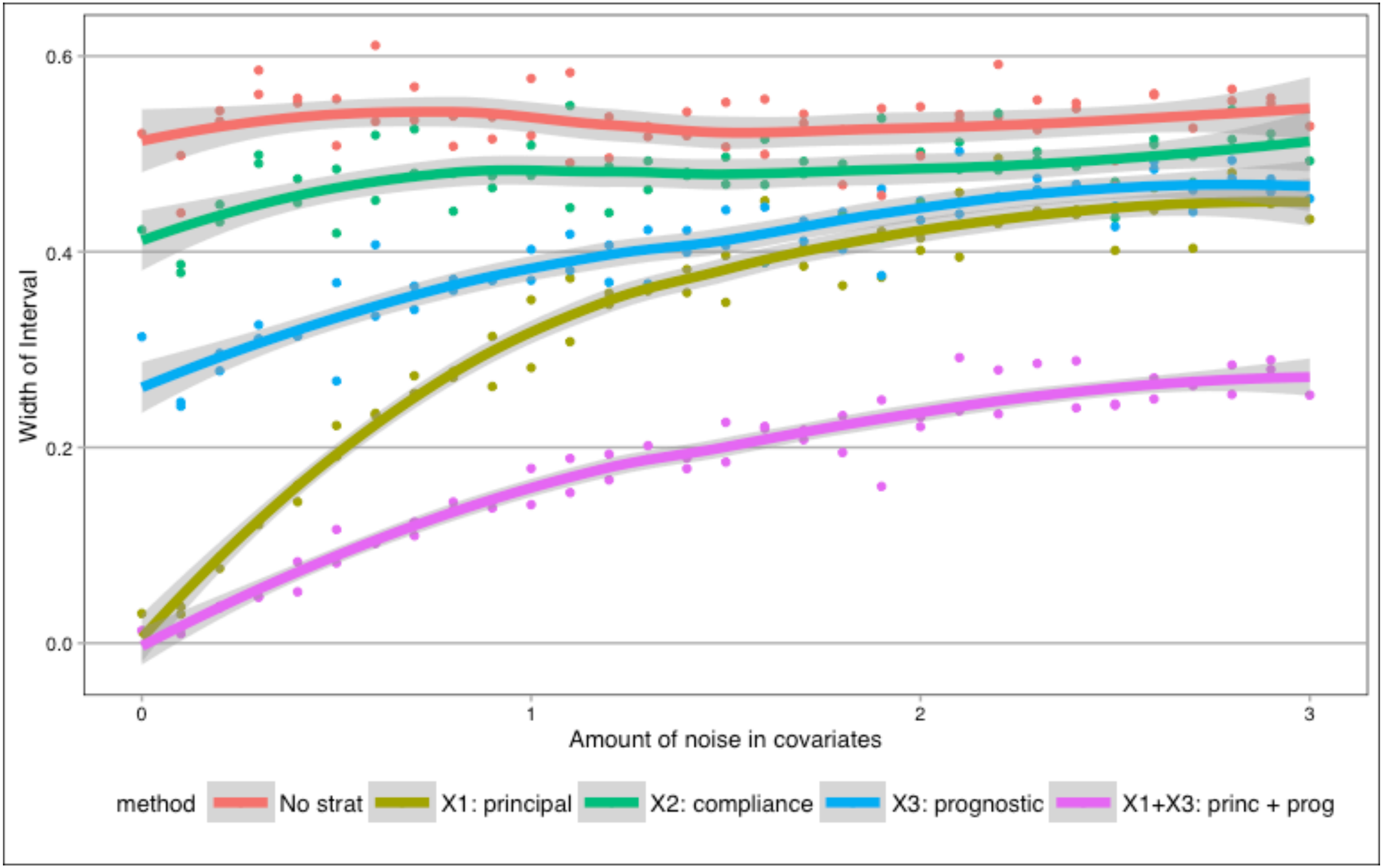
Sample data

	X1	X2	X3	Z	S	Yobs
1	-2.27	-1.68	-1.38	0	lq	0
2	0.56	-0.17	-0.55	1	e	1
3	1.21	1.38	0.59	0	hq	0
4	1.01	2.15	1.58	1	e	1

Bound Width for LQC Group



Bound Width for HQC Group



Conclusions

Covariates can substantially sharpen bounds

- Both simulation studies and empirical examples (not shown) show substantial impact of using covariates.

Propensity scores seem to be more effective than prognostic scores

- In simulation studies, propensity scores "won"
- A caveat: We do not have easy way of "comparing predictiveness" so comparisons might not be fair.
- Cross-cutting data and using both does seem to be even more effective, however.

Thank you!