Multi-Level Matching: Hallberg & Cook

• Present 2 WSCs where the intervention at the school level, and students are nested within schools.

• Issue is how well a 2 level model works to reduce bias vs single school or student level, and how well school works alone vs student level.

• “good” set of variables at each level, incl 5 prior time points at school level and 2 at student level, plus lots of others.

• Intervention 1 is teacher feedback in Indiana Year 1.

• Intervention 2 is a science and literacy intervention in Fla.

• Outcome is performance in ELA and math.
Presentation

• Begin with Balance at each level and with (a) no covariates; (b) matching at each level, and then © matching at both levels.

• We use nearest neighbor matching

• First in Indiana and then in Florida

• Finally we produce estimates of correspondence in causal estimates by design, and do so as a function of

• Covariates at one level only

• Matching schools 1 to 1 and 1 to 4

• Regression or matching at the school level
Figure 1. School level matching approaches – Absolute standardized mean difference in school level covariates – Indiana dataset
Figure 2. School level matching approaches (4 school match) – Absolute standardized mean difference in student level covariates – Indiana dataset.
Figure 3. School level matching approaches (1 school match) – Absolute standardized mean difference in student level covariates – Indiana dataset
Figure 4. Student level matching approaches – Absolute standardized mean difference in school and student level covariates – Indiana dataset
Figure 5. School level matching approaches – Absolute standardized mean difference in school level covariates – P-SELL dataset
Figure 6. School level matching approaches (4 school match) – Absolute standardized mean difference in student level covariates – P-SELL dataset
Figure 6. School level matching approaches (1 school match) – Absolute standardized mean difference in student level covariates – P-SELL dataset
Figure 8. Student level matching approaches – Absolute standardized mean difference in school and student level covariates – P-SELL dataset
Figure 9. Math standardized treatment effect in the quasi-experiment relative to the benchmark = 0.
Figure 10. Reading standardized treatment effect in the quasi-experiment relative to the benchmark $= 0$. 

![Graph showing treatment effect in the quasi-experiment relative to the benchmark. The graph compares different matching strategies: unadjusted, 1 school match, 4 school match, 1 school match - student regression, 4 school match - student regression, 1 school match - student match, 4 school match - student match, multi-level student match, and student-level student match. The graph also includes data points for Indiana and P-SELL.]
Conclusions re Multi-Level Matching

• School level matching alone achieved balance on both school and student characteristics
• Student level matching alone achieved balance on student but not school characteristics
• Matching schools 1:4 (nearest neighbor on est PS) did marginally better than 1:1 both in achieving balance and also in reducing bias
• Bias was consistently close to zero across studies when either school matching alone occurred 1:4 or when both levels were matched including 1:1
• Recommendation: If intervention at school level be sure to match there; do so 1:k; and add student level though likely only to be minor insurance.
Big Picture 1

• In Ed 32 WSCs to date we have been able to count, most pretty recent
• Nearly all show results “close” to RCT, but not meta-analyzed to date
• Data analysis not matter much
• NECGDs show importance of (a) knowledge of selection process; (b) intact comparison group selection for overlap; © local matching of individual cases; (d) pretests and maybe rich covariates
• but caution about rich covariates alone
• Importance of school level matches with school level interventions
Big Picture 2

• Link to theoretical claims – Rubin model – know selection, measure accordingly and top up with many multi-dimensional reliable covariates – better phrase than “rich”

• Link to historical claims – Heckman and action specifics

• Link to Bigger Data – attributes of better QEDs and growth in dimensions of bigger data? (a) does bigger data = more RA or RD?; (b) does bigger data mean more time points in past, more dense data collection facilitating more local comparisons; more data on successive cohorts within organizations; more heterogeneous constructs; more measures of same construct; more empirical descriptions of selection process?

• If we are right, does bigger data = more opportunities for better Qes?
The Workhorse Design inherently limited

• Why not escape into CITS?
• Why not improve on the design in other ways
• We will discuss these other ways
• Begin with Coady
• Tomorrow Will and I join him to discuss alternatives