# SUMMARY OF RIGOROUS IMPACT EVALUATION METHODOLOGIES

| | Identifying Assumption | Data Requirements | Advantages | Disadvantages |
|---|---|---|---|---|
| Randomized Controlled Trial | No systematic differences in unobservable characteristics between treated & control groups (bolstered by showing that there are no systematic differences in observable characteristics between treated & control groups) | At least post-intervention data on randomly assigned members of treated & control groups (baseline helps make identifying assumption plausible) | • Weakest identifying assumption of 4 methods (as long as randomization successful)<br>• When budget prevents program from being provided to everyone simultaneously, randomizing (either timing or in the absolute sense of who does and does not get the program) is sometimes viewed as the fairest way to allocate scarce resources | • Spillovers can invalidate control group<br>• Might be politically difficult to argue in favor of randomization<br>• External validity can be questionable (was the study population representative of the population served at scale?) |
| Difference-in-Difference | Whatever happened to the control group over time is what would have happened to the treated group in the absence of the program | At least pre- & post-intervention cross sections including both treated & control (additional pre-intervention data helps make identifying assumption plausible) | Might be possible to do with existing survey data | Identifying assumption is relatively strong (need to consider all other possible differences between the groups that could have led to the observed outcomes) |
| Regression Discontinuity | Those just above the threshold (the treated) and those just below (the control) are otherwise identical | Data on threshold qualification (already exists) & outcomes | When threshold is built into program design there is no need to do anything special for evaluation | Only identifies program impact for those at the threshold |
| Matching | After controlling for observables, treated & control are not systematically different | Rich data on as many observable characteristics as possible, large sample size (so that it is possible to find appropriate match) | Might be possible to do with existing survey data | Strong identifying assumption (even if they're otherwise identical, why did some get program and others not?) |