## Figure 3    Hierarchy of evidence

| Type | Design features | Evidence quality |
|---|---|---|
| **Systematic reviews and meta-analysis** | Use established approach to synthesize all quality research evidence (esp., "upper tier" studies) on a specific issue. | Strongest evidence but only as strong as underlying evidence. |
| **Upper Tier – Individual studies with randomization/credible source of exogenous variation** | | |
| **Randomized experiments** | Well-designed with sufficient sample size. | Very strong evidence. |
| **Natural experiments** | High quality source of exogenous variation generating comparison group that provides credible approach to estimating counterfactual. Well-designed pre-post measures of outcomes and well-measured and appropriate data with large sample. Employs techniques such as regression discontinuity, instrumental variables, difference-in-differences, or propensity score matching. | Very strong evidence if the source of exogenous variation is credible and if appropriate econometric/statistical technique(s) is (are) employed to extract the information from the data. |
| **Middle Tier – Limited or no source of exogenous variation, but with credible comparison group/counterfactual** | | |
| **Some control in the assignment of treatment** | Limited source of exogenous variation or some control of selection process (e.g., program administrator, perhaps non-randomly, assigns treatment; different sites follow different procedures; or individuals select into limited range of options). Well-designed pre-post outcome measures; dynamic pre-treatment measures; well measured, appropriate data with large sample. Employs techniques such as difference-in-differences and/or propensity score matching, or an appropriate regression technique. | Studies in this tier produce evidence that ranges from *very strong* and *strong* to *moderate* depending on specific design features. All other things being equal, studies with some control in assignment of treatment are generally ranked higher than studies without control. |
| **Correlational studies including studies relying on selection on observables and case studies with a comparison group** | Reasonable approach to estimating counterfactual; well-designed pre-post measures of outcomes; large sample and rich set of covariates. Quality of the comparison group is critical. Employ techniques such as difference-in-differences; population correlation designs; propensity score matching, hierarchical linear modeling, structural equation modeling, and OLS regression. Longitudinal designs may use techniques like fixed effects. | Studies without any exogenous variation but with a credible comparison group/counterfactual generally produce evidence that ranges from *very strong* and *strong* to *moderate* depending on specific design features. |
| **Lower Tier – Studies without measured comparison groups/counterfactuals** | | |
| **Studies without comparison group** | Credible case selection, explicit causal logic model and analytical strategy, understanding of the process, quality outcome measures. | Evidence should be considered suggestive. |
| **Participant satisfaction** | Collect feedback from participants on value of intervention. Better quality studies ask about "value added" or change in relevant outcomes following from treatment, rather than only eliciting measures or opinions regarding satisfaction, inputs, outputs, processes, or outcomes. | Care needs to be taken to understand potential biases and interpret the findings accordingly. |
| **Expert opinions** | Respected organizations or individuals, explicit rationale for opinion. | |
| **Exploratory case studies** | Less credible/explicit: case selection criteria, theory of change, analytical strategy, or outcome measures. Does not have good quality (or any) outcome measures. May rely on measures of inputs or outputs. | Evidence should be considered suggestive. |

*Caveat emptor*: Some studies claim, a high quality source of exogenous variation or rich covariates for matching, but the reader must decide whether these claims are credible. A study using an instrumental variable approach with a poor source of exogenous variation may be best discarded. Even the most sophisticated technique cannot extract information from data that is not there to start with.