Evaluation Study Design –
A pluralist approach to evidence

Dr Allyn Fives
Dr John Canavan
Prof Pat Dolan

UNESCO Child & Family Research Centre
School of Political Science and Sociology
NUI Galway, Ireland
Overview

• What counts as evidence?
  – Hierarchical approach (RCTs at the apex)
  – Pluralist approach (Matrix) offered instead

• Two lines of argument

  1. Methodological aptness
     Against the arguments in a polarized debate on RCTs

  2. Experimental & non-experimental research required
Randomized controlled trials (RCTs)

• What is an RCT?

A study design where treatments are assigned randomly and treatments are withheld from participants.

• Participants do not all receive the same treatments

• Treatments are not assigned on the basis of criteria used in everyday professional practice, in particular the professional’s judgment of the recipient’s needs.
What type of evidence is required to underpin policy & practice?

Hierarchical approach

- RCTs with a very low risk of bias provide the best evidence
- Why? Because they rule out confounding explanations, they allow causal inferences to be drawn

Not all RCTs are equal: Bias may result from attrition or treatment diffusion

Causal evidence *PLUS* manuals to ensure consistent replication
# Hierarchy: an example (N.I.C.E. 2005)

<table>
<thead>
<tr>
<th>Level of evidence</th>
<th>Type of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1++</td>
<td>High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias</td>
</tr>
<tr>
<td>1+</td>
<td>Well-conducted meta-analyses, systematic reviews of RCTs, or RCTs with a low risk of bias</td>
</tr>
<tr>
<td>1-</td>
<td>Meat-analyses, systematic reviews of RCTs, or RCTs with a high risk of bias*</td>
</tr>
</tbody>
</table>
| 2++               | High-quality systematic reviews of case-control or cohort studies  
High-quality case-control or cohort studies with a very low risk of confounding, bias, or chance and a high probability that the relationship is causal |
| 2+                | Well-conducted case-control of cohort studies with a low risk of confounding, bias, or chance and a moderate probability that the relationship is causal |
| 2-                | Case-control or cohort studies with a high risk of confounding, bias, or chance and a significant risk that the relationship is no causal* |
| 3                 | Non-analytic studies (e.g. case reports, case series) |
| 4                 | Expert opinion, formal consensus |
What type of evidence is required to underpin policy & practice?

Pluralist approach (Evidence Matrix)

Wide variety of research questions (not just “Does X work?”)
Wide variety of study designs (not just RCTs)

- Question of effectiveness answerable by a range of study designs:
  RCT, QES, Cohort study
- Whether a proven programme works here:
  Theory of change, norm referenced, benchmark
- Implementation & Exploratory Questions:
  Documentary analysis, qualitative studies, surveys
- “How does X work?”
  RCT also can address
### Pluralist Approach (Evidence Matrix)

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Study Design (Evidence Type)</th>
<th>Meta-analyses of RCTs</th>
<th>RCTs</th>
<th>Repeated case studies (n=1)</th>
<th>QESs</th>
<th>Case-control studies</th>
<th>Cohort studies</th>
<th>Theory of change studies</th>
<th>Norm referenced studies</th>
<th>Benchmark studies</th>
<th>Documentary studies</th>
<th>Qualitative studies</th>
<th>Survey</th>
<th>Expert knowledge studies</th>
<th>Reflective practice</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exploratory Studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are needs being met by existing services?</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are practitioners satisfied with services?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are users satisfied with services?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is the current service cost-effective?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What alternative would be appropriate?</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does the workforce have necessary skills to implement alternative service?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are financial &amp; other resources available to implement alternative service?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Will users take up alternative service?</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How does proposed alternative fit with policy / service priorities?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Implementation and Outcomes Studies</strong></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does X work more than Y?</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does X do more good than harm?</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does a proven program work here?</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How does X work?</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is it meeting its target group?</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are users satisfied &amp; will they take up?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are practitioners satisfied?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was it implemented with fidelity?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is it cost-effective?</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is it socially valuable?</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. An Evaluation Evidence Matrix: A pluralist guide to appropriate study designs.

Note: RCT = Randomized Controlled Trial; QES = Quasi Experimental Study
Arguments for Pluralism (Matrix)

First line of argument:

The argument from methodological aptness
Different types of research question are best answered by different types of study

Against the two extremes in a polarized debate:
1. That RCTs are the gold standard (& they are appropriate)
2. That RCTs are in most cases to be avoided (& are inappropriate)

They are both based on commonly held but questionable arguments
The arguments in the RCT debate

1. That RCTs are the “the gold standard” in social evaluations

   Argument 1a: RCTs ethically permissible as duties of care are irrelevant
   Argument 2a: Study conditions must be (and will be) equal at baseline
   Argument 3a: Only RCTs provide evidence of what works

2. That RCTs are in most cases to be avoided in social evaluations

   Argument 1b: RCTs in conflict with duties of care
   Argument 2b: Study conditions must be (but will not be) equal at baseline
   Argument 3b: RCTs cannot be scientifically promising in social evaluations
Against the polarised positions in RCT debate

Qualified defense of RCTs:

RCTs may be scientifically promising
   Even without baseline equality
   Even when there are threats to internal & external validity

RCTs may be ethically justified
   Even when the professional community sees no need for causal evidence

But:

Ethical objections may arise
   Because the duty of care remains, even when all consent

Alternative study designs may be chosen for methodological reasons
   e.g. QES, single case study design (n=1)
Arguments for Pluralism (Matrix)

Second line of argument:

Evaluations often require experimental & non-experimental research

1. Adaptation
Evidence-based practices (EBPs) may need to be adapted: this requires practice-based evidence (PBE)

2. Mixed methods studies
Causal evidence not by definition superior
Other evidence types help answer ‘how’ and ‘why’ programmes work
A Matrix guide to appropriate study designs

What a Matrix can do:

a) Provide guidance on exemplary questions
b) Rule out an evidence type as inappropriate
c) Indicate what evidence type may be appropriate
d) Show if a study design can answer more than one question
e) Show if more than one study design can answer same question
f) Help make ethically-informed decisions

What a matrix cannot do:

a) Resolve moral dilemmas
b) Distinguish between strong & weak evidence types
c) Show how to combine evidence types
Relevant publications


