Bias and evidence hierarchies in the evaluation of smarter choices

Dr Steve Melia
Senior Lecturer, Transport & Planning
Motivations for this Review

1. Controversy around the robustness of smarter choices (or ‘VTBC’) evaluations
2. Growing influence of evidence hierarchies favouring experimental methods
3. DfT decision to downplay smarter choices in WebTag
Aims

1. To examine claims of invalidity or bias in evaluations of smarter choices
2. To examine the case for evidence hierarchies (Do RCTs offer a solution to ‘low quality’ transport research?)
3. To consider implications of applying evidence hierarchies to transport research generally
Controversy: Alamein Travelsmart Evaluation

Morton & Mees (2010)

Alleged sources of bias in the evaluation:
• Expectation bias
• Good subject effect
• Non-response bias

Special edition of Transport Policy 16(6) on evaluation of smarter travel (Chatterjee 2009)
## The Challenge of Triangulation

<table>
<thead>
<tr>
<th></th>
<th>Household surveys</th>
<th>Traffic counts</th>
<th>Possible explanations for trends</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trips per person</td>
<td>Distance per person</td>
<td>Overall change</td>
</tr>
<tr>
<td>National trend</td>
<td>-1.2%</td>
<td>-0.9%</td>
<td>-0.5% (car traffic) -0.7% (all vehicles)</td>
</tr>
<tr>
<td>Sustainable Travel Towns</td>
<td>-9%</td>
<td>-5%~ -7%</td>
<td></td>
</tr>
<tr>
<td>Darlington</td>
<td>-7%~ -10%</td>
<td>-6%~ -7%</td>
<td>-2.4% to -3.2% (all vehicles)</td>
</tr>
<tr>
<td>Peterborough</td>
<td>-8%~ -10%</td>
<td>-7%~ -10%</td>
<td>-2.4%</td>
</tr>
<tr>
<td>Worcester</td>
<td>-8%~ -10%</td>
<td>-3%</td>
<td>Growth until 2006/07, then fall; -1.9% to -2.6%</td>
</tr>
</tbody>
</table>

Sustainable Travel Demonstration Towns, from Sloman et al (2010) (two intermediate columns removed)
Triangulation on a ‘Like for Like’ Basis
Dungarvan (Irish Sustainable Travel Town) – see Melia (2013)
Experimental Methods

Before and after studies

Approaches to Research Design

1. ‘Horses for courses’ e.g. Tavistock Institute & AECOM (2010)

2. Hierarchical
Hierarchies of Evidence?

Applied to smarter choices by:

- Graham-Rowe et al. (2011),
- Möser and Bamberg (2008) – cited in DfT WebTAG decision
- To school travel plans by Rowland et al. (2003)

From Leigh (2009)
Meta studies of randomised trials

Randomised trials

Meta studies of natural experiments

Natural Experiments (quasi-experimental or theory based)

Before and after studies

From Leigh (2009)

Hierarchies of Evidence?

Relatively few of these in the literature
Hierarchies of Evidence?

Under what circumstances can we state that RCTs generate more robust evidence for policy – based on a comparison of methods alone?

From Leigh (2009)
If All Are *Fully* Satisfied

1. The main focus of the research is to test (but not explain) a hypothesised cause-effect relationship.
2. A representative study population of a sufficient size can be obtained from the target population to whom the intervention would be applied.
3. The intervention can be applied selectively to an experimental group within the study population.
4. No other factors with a significant influence on the outcome would impact the experimental and control groups differently during the experiment.
5. Wider application of the intervention would replicate the causal relationships which applied during the experiment.

*Experimental methods will yield more robust results*
If Criteria Are *Partially* Satisfied

1. The main focus of the research is to test (but not explain) a hypothesised cause-effect relationship.
2. A representative study population of a sufficient size can be obtained from the target population to whom the intervention would be applied.
3. The intervention can be applied selectively to an experimental group within the study population.
4. No other factors with a significant influence on the outcome would impact the experimental and control groups differently during the experiment.
5. Wider application of the intervention would replicate the causal relationships which applied during the experiment.

Comparison becomes an empirical question
Key Issues: Scale & Social Interaction

During the experiment:

* Interactions must either be identical – or else have no impact on outcomes

How significant are these influences?

Implementation:

Wider Society

Target Population
Example – Role of Cycling Infrastructure in Modal Choice

**Quasi-experimental evidence**: no significant modal shift e.g. Brand et al. (2014)

**Historical/descriptive evidence**. Symbiotic relationship e.g. Pucher et al. (2010), Melia (2015)
Which Types of Question Do We Ask?

Quasi-experimental evidence:

Conclusions

1. There are reasons for some concern about potential bias in the evaluation of smarter choices (or any other transport intervention involving human behaviour)
2. There are no easy answers, but self-contained areas may offer one means of strengthening triangulation
3. Experimental methods are only suited to a relatively narrow range of transport questions
4. Hierarchies of evidence risk choice of question bias, misapplication of experimental methods and misleading findings for policymakers
5. Researchers need to do more to educate policymakers about the limitations of experimental methods and of quantification and certainty in findings.
References