Can harnessing causal inference with large-scale data enable the health economist to reach the policy-making summit?

Richard Grieve

richard.grieve@lshtm.ac.uk
Van-HEM: thanks are due
Getting (the right) health econ into policy..

Why’s it like mountaineering

The 3 rules of mountaineering
‘It's always further than it looks. It's always taller than it looks. And it's always harder than it looks.’

‘Each climber loses one finger or toe once in a while. This is a small but important reason for Polish climbers’ success. Western climbers haven't lost as many fingers or toes. **Wanda Rutkiewicz**’
Structure of talk

- **Context:** Policy, CEA, RCTs and population
- **Framework:** Providing treatment effects for populations
- **Why causal inference and big data can help when**
  - populations differ
  - comparators differ
  - preferences differ
- **Areas further research**
Context

• Cost-effectiveness used in policy
• NICE; CADTH; PBAC; many others
• Decision-modelling, mixed treatment comparisons, utilities
• Clear mechanism for health econ into decision-making
• BUT decision-makers want cost-effectiveness for target population, subpopulations and individual patients
RCTs: the problem?

• Heavy investment in trials, pharmaceutical companies, public funders
• Methods guidance clear: estimate relative effectiveness RCTs
• Standard approach use trials, network meta-analyses within models
• **Fundamental problem: mismatch design and decision**
  – 1. populations 2. treatments 3. preferences
• CEA, assume external validity without justification
• What are we assuming?
• How can we test the underlying assumptions?
• Danger: providing policy-makers with inaccurate evidence
Problem 1: selection into trial

- Excluded from RCT
- In RCT but non compliant
- In RCT but loss to followup
- In RCT incomplete data
- In RCT full data
External validity not dealt with

- 2015: 70 RCTs in NEJM, JAMA Lancet
- 34 (49%) any assessment external validity
- Most assessments were qualitative.
- Asserted patients ‘representative’ from baseline characteristics
- 5 (7.1%) studies any quantitative data
- None studies adjusted outcomes accordingly
- Health economics evaluation, similar (Georghe et al, 2013)
Towards Better use of big data.

Pulmonary artery catheterization (PAC)

- Common Invasive device monitoring flow Intensive care Units (ICU)
- Routine practice limited evidence (but strong beliefs)
- Highly influential observational study: PAC increase mortality,
- **UK multicentre RCT:** PAC no effect on survival, and not cost-effective
- Concern seminal RCT lacked external validity
- Prospective non-randomised study (NRS)
- Accessed UK intensive care database over 2 million admissions
- Included data from 50 centres, where patients had PAC (or no PAC)
- NRS same protocol, casemix, resource use and endpoints RCT
## PAC patients in RCT: covariates and outcomes

<table>
<thead>
<tr>
<th></th>
<th>RCT PAC (n=506)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Age</td>
<td>64.2</td>
</tr>
<tr>
<td>% Elective surgical</td>
<td>6.3</td>
</tr>
<tr>
<td>% Emergency surgical</td>
<td>28.1</td>
</tr>
<tr>
<td>% Ventilated admission</td>
<td>88.9</td>
</tr>
<tr>
<td>% Teaching hospital</td>
<td>21.7</td>
</tr>
<tr>
<td>% In hospital Mortality</td>
<td>68.4</td>
</tr>
<tr>
<td>Mean hospital cost (£)</td>
<td>18,612</td>
</tr>
</tbody>
</table>
## Selection PAC patients
### RCT vs NRS

<table>
<thead>
<tr>
<th></th>
<th>RCT PAC (n=506)</th>
<th>NRS PAC (n=1,051)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Age</td>
<td>64.2</td>
<td>61.9</td>
</tr>
<tr>
<td>% Elective surgical</td>
<td>6.3</td>
<td>9.3</td>
</tr>
<tr>
<td>% Emergency surgical</td>
<td>28.1</td>
<td>23.1</td>
</tr>
<tr>
<td>% Ventilated admission</td>
<td>88.9</td>
<td>86.2</td>
</tr>
<tr>
<td>% Teaching hospital</td>
<td>21.7</td>
<td>42.5</td>
</tr>
<tr>
<td>% In hospital Mortality</td>
<td>68.4</td>
<td>59.3</td>
</tr>
<tr>
<td>Mean hospital cost (£)</td>
<td>18,612</td>
<td>19,577</td>
</tr>
</tbody>
</table>
Selection into trials: why?

There could not be worse experimental animals on earth than human beings; they complain, they go on vacations, they take things they are not supposed to take, they lead incredibly complicated lives, and, sometimes, they do not take their medicine*.

Reliance on trial-based CEA: example of the streetlight effect
Insufficient response 1

• Decision models

Whither trial-based economic evaluation for health care decision making?

Mark J. Sculpher*, Karl Claxton*, Mike Drummond* and Chris McCabe*

*Centre for Health Economics, University of York, UK
*SchARR, University of Sheffield, UK

Summary

The randomised controlled trial (RCT) has developed a central role in applied cost-effectiveness studies in health care as the gold standard for analyses. This paper considers the role of trial-based economic evaluation in this era of explicit decision making. It is argued that any framework for economic analysis can only be judged insofar as it can inform two key decisions and be consistent with the objectives of a health care system subject to its resource constraints. The two decisions are, firstly, whether to adopt a health technology given existing evidence and, secondly, an assessment of whether more evidence is required to support this decision in the future. It is argued that a framework of economic analysis is needed which can estimate costs and effects, based on all the available evidence, relating to the full range of possible alternative interventions and clinical strategies, over an appropriate time horizon and for specific patient groups. It must also enable the accumulated evidence to be synthesised in an explicit and transparent way in order to fully represent the decision uncertainty. These requirements suggest that, in most circumstances, the use of a single RCT as a vehicle for economic analysis will be an inadequate and partial basis for decision making. It is argued that RCT evidence, with or without economic content, should be viewed as simply one of the sources of evidence, which must be placed in a broader framework of evidence synthesis and decision analysis. Copyright © 2006 John Wiley & Sons, Ltd.
Insufficient response 2

• Poorly designed Observational studies

The New York Times

• Hormone Replacement Study a Shock to the Medical System
• By GINA KOLATA with MELODY PETERSEN
• Published: Wednesday, July 10, 2002
• “The announcement yesterday that a hormone replacement regimen taken by six million American women did more harm than good was met with puzzlement and disbelief by women and their doctors across the country.”

• “A rigorous study found that the drugs, a combination of estrogen and progestin, caused small increases in breast cancer, heart attacks, strokes and blood clots. Those risks outweighed the drugs' benefits: a small decrease in hip fractures and a decrease in colorectal cancer. Many of the 16,000 women in the study, supported by the National Institutes of Health, opened letters yesterday telling them to stop the drugs.”
Recalling Advice from a friend..
The mountain still has to be climbed

RCTs
Problems: selection/wrong treatments/preferences

Solutions:
Observational data (combined with RCTs)
Causal inference
Assumptions and transparency
The framework

From sample average treatment effect to population average treatment effect on the treated: combining experimental with observational studies to estimate population treatment effects

Erin Hartman,
University of California at Berkeley, USA

Richard Grieve
London School of Hygiene and Tropical Medicine, UK

Roland Ramsahai
University of Cambridge, UK

and Jaajeste S. Seikhan
University of California at Berkeley, USA

[Received December 2013. Final revision August 2014]

Summary. Randomized controlled trials (RCTs) can provide unbiased estimates of sample average treatment effects. However, a common concern is that RCTs may fail to provide unbiased estimates of population average treatment effects. We derive the assumptions that are required to identify population average treatment effects from RCTs. We provide placebo tests, which formally follow from the identifying assumptions and can assess whether they hold. We offer new research designs for estimating population effects that use non-randomized studies to adjust the RCT data. This approach is considered in a cost-effectiveness analysis of a clinical intervention: pulmonary artery catheterization.

Keywords: Causal inference; Cost-effectiveness studies; External validity; Observational studies; Placebo tests; Randomized controlled trials

1. Introduction

Randomized controlled trials (RCTs) can provide unbiased estimates of the relative effectiveness of alternative interventions within the study sample. Much attention has been given to improving the design and analysis of RCTs to maximize internal validity. However, policy makers require evidence on the relative effectiveness and cost-effectiveness of interventions for target populations that usually differ from those represented by RCT participants (Hoeh et al., 2002; Mit and Indurkhya, 2005; Mojtabai and Zivin, 2003; Nixon and Thompson, 2005; Willan et al., 2006). A key concern is that estimates from RCTs and meta-analyses lack external validity (Alcott and Mullainathan, 2012; Deaton, 2009; Heckman and Urzua

An Approach to Assess Generalizability in Comparative Effectiveness Research: A Case Study of the Whole Systems Demonstrator Cluster Randomized Trial Comparing Telehealth with Usual Care for Patients with Chronic Health Conditions

Adam Stevenhagen, MA, Richard Grieve, PhD, Martin Bardsley, PhD

Background. Policy makers require estimates of comparative effectiveness that apply to the population of interest, but there has been little research on quantitative approaches to assess and extend the generalizability of randomized controlled trial (RCT)-based evaluations. We illustrate an approach using observational data. Methods. Our example is the Whole Systems Demonstrator (WSD) trial, in which 3330 adults with chronic conditions were assigned to receive telehealth or usual care. First, we used novel placebo tests to assess whether outcomes were similar between the RCT control group and a matched subset of nonparticipants who received usual care. We matched on 60 baseline variables obtained from the electronic medical record. Second, we conducted sensitivity analysis to consider whether the estimates of treatment effectiveness were robust to alternative assumptions about whether “usual care” is defined by the RCT control group or nonparticipants. Thus, we provided alternative estimates of comparative effectiveness by contrasting the outcomes of the RCT telehealth group and matched nonparticipants. Results. For some endpoints, such as the number of outpatient attendances, the placebo tests passed, and the effectiveness estimates were robust to the choice of comparison group. However, for other endpoints, such as emergency admissions, the placebo tests failed and the estimates of treatment effect differed markedly according to whether telehealth patients were compared with RCT controls or matched nonparticipants. Conclusions. The proposed placebo tests indicate those cases when estimates from RCTs do not generalise to routine clinical practice and motivate complementary estimates of comparative effectiveness that use observational data. Future RCTs are recommended to incorporate these placebo tests and the accompanying sensitivity analyses to enhance their relevance to policy making. Key words: causal inference; external validity; generalizability; randomized trials; telehealth; chronic health conditions. (Med Decis Making 2015;35:1363-1386)

Well-conducted randomized controlled trials (RCTs) can ensure high levels of internal validity because the treatment groups are balanced. However, a major concern with RCT evidence is that criteria and compares the intervention with usual care, the trial may exclude important subgroups of patients and context.
General approach

1. Target population defined from observational data

2. Estimate Treatment effectiveness in RCT

3. Use observational data to reweight RCT to target population

4. Assess external validity with Placebo test: reweighted RCT vs target population

5. If test passes, treatment effectiveness after reweighting to target population

6. If test fails..present treatment effectiveness from observational data
Causal inference (the map)

• Aims to learn about cause-effect relationships
• Suggest novel statistical methods (and data) required lead to lead to more reliable causal inferences
• Here quantitative approach, combines RCT and NRS
• Program evaluation: Heckman, Vytlacil 2008; Imai et al 2008;
• Defines underlying assumptions
• Reweights RCT estimates for target population
• Tests underlying assumptions
• Aims to give unbiased estimates for the target population
What do we want to estimate?
(see Imai et al, 2008)

Population versus sample effects

• Population average treatment effects
  – e.g. treatment effects in target population (e.g. for treated, PATT)
• Sample average treatment effect (e.g. for treated, SATT)
  – e.g. treatment effects for treated in RCT

• Patient and contextual characteristics differ across settings
• These characteristics may modify treatment effects
• Treatment or control regimens may differ by setting
• If heterogeneity or treatments differ, $SATT \neq PATT$
Identifying PATT from RCT

Key assumptions

1. Treatment invariant to sample assignment (consistency)
   Individual's potential outcomes for RCT or target population, e.g. for $t=1$
   \[ Y_{i01} = Y_{i11} \]

2. Strong ignorability of sample assignment (for treated)
   Potential outcomes independent for same $W$
   \[ (Y_{o1}, Y_{11}) \perp S \mid (W, T = 1) \]
Testing assumptions
Placebo tests (Jones, Health Econ 2007)

- Test assumptions by comparing outcomes population vs RCT
- Are outcomes RCT after reweighting same as target population
- Placebo tests: assess if recover zero effect with current model
- Null Hypothesis: data are inconsistent with valid research design
- Will accept null if:
  a) Observe outcome differences between settings
     i) selection into RCT conditional on potential outcomes
     ii) treatment (or control) differs between settings
  b) lack of power
- i.e. A good result is a **small** mean difference and **low** P Value
Large scale observational data initiatives (the food and drink)

Artificial Intelligence, Big Data, and Cancer

Even though the original computers were designed in 1946, computers have become part of the social and professional fabrics of our lives only since the mid-1980s, enhancing workplace and individual productivities. Computers are still evolving, and so are the ways we use them. Artificial intelligence (computer) and big data (incorporating large volumes of information into computers) are gaining wide acceptance in different fields. Conventional “Deep Blue,” an IBM-trained chess computer that beat Gary Kasparov, the chess world champion. Another IBM-trained Watson System won the Jeopardy! game against champions. After 12 years of refinements, poker computers now consistently win against poker world champions.

Having proven the concept in chess, Jeopardy!, and poker, this technology is being brought to the real world. This new form of artificial intelligence, or “cognitive computing,” learns in ways similar to humans’ learning. With proper “training,” they address human-like situations and tasks that could be targeted with new or older drugs. The American Society of Clinical Oncology has developed CancerLinQ, which proposes to incorporate data of patients with cancer in the United States into 1 large database. This would capture cancer data on 100% of patients with cancer rather than the 3% who are entered on clinical trials, and thereby accelerate new information, knowledge, and discoveries.

Cancer centers like MD Anderson and Memorial Sloan Kettering are launching single-institutional large database endeavors aimed at enhancing cancer care and research. The MD Anderson initiative, referred to as the Oncology Expert Advisor (OEA) will incorporate all information related to the more than 1 million patients with cancer treated over the lifetime of the institution to generate a novel support system for research and patient care.

Novel cognitive computers may include all the clinical and laboratory information available in different cancer centers. The New England Journal of Medicine, 365, 1587-1588.
Implementing approach in PAC case study to identify population average treatment effect (PATT)

Trial nested within large clinical database
1. Effectiveness within RCT, use matching estimate SATT
2. Develop model for trial inclusion, use that model reweight the SATT
3. Placebo tests, contrast weighted outcomes RCT vs population
4. Use reweighting approach to report PATT
## Placebo tests - mortality

Population treated versus RCT treated after reweighting

Null hypothesis: the study design is invalid

<table>
<thead>
<tr>
<th></th>
<th>mortality difference</th>
<th>P Value</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>-3%</td>
<td>0.05</td>
<td>96%</td>
</tr>
<tr>
<td>Subgroup</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non teaching hospital</td>
<td>-3%</td>
<td>0.05</td>
<td>85%</td>
</tr>
<tr>
<td>Non surgical</td>
<td>-4%</td>
<td>0.06</td>
<td>83%</td>
</tr>
<tr>
<td>Elective Surgery</td>
<td>8%</td>
<td>0.46</td>
<td>8%</td>
</tr>
</tbody>
</table>
PATT versus SATT

Incremental net benefits PAC - No PAC

£20,000 per QALY
Interpretation and implications (I)

- Harnesses RCTs and large observational data
- Concern patients differed, adjusted for observed differences
- Tests assumptions for estimating population treatment effects
- In this example, placebo tests passed, overall, but not for subgroup
- High value of further research for elective surgical patients
- Leverage observational data, combined with RCT to personalise treatment (Kunzel, Sekhon et al, 2017)
- More generally, complex treatments or control a typical in RCT
Problem 2: treatment or controls a typical???

- Involved in 10 RCTs, total spend around £25m
- All published in eminent clinical journals
- 1 (acupuncture), effect on primary outcome (cost-effective)
- 1 effect on secondary outcome (cost-effective)
- 8/10 found no effect on primary and secondary outcomes..
- Common clinical response: ‘usual care really improved during the trial, that’s why we didn’t find a difference..’
- Another *did* find effect on main clinical outcome but major concerns about the external validity of the findings..
Example II: Telehealth

“the remote exchange of data between a patient and health care professionals as part of the diagnosis and management of health care conditions”

Telehealth devices enable items such as blood glucose level and weight to be measured by the patient and transmitted to health care professionals working remotely.
Whole System Demonstrator Trial (WSD)
Telehealth vs ‘usual care’ Steventon et al, 2012

• Huge Policy interest: is it effective and cost-effective?
• Adult Patients with diabetes, COPD, heart failure
• Cluster Randomised design
• Randomised 179 GP practices, 3230 adults
• Intervention: broad class of Telehealth devices
• Control: usual care at the trial sites
• Blinding: patients at consent, not for recruiters
• Outcomes: emergency admissions, outpatient visits
• Major concern was that usual care differed between settings
## Baseline Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Control group N=1584</th>
<th>Intervention group N=1570</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>70.8 (11.7)</td>
<td>69.7 (11.6)</td>
</tr>
<tr>
<td>Female no (%)</td>
<td>643 (40.6)</td>
<td>647 (41.2)</td>
</tr>
<tr>
<td>COPD no (%)</td>
<td>786 (49.6)</td>
<td>739 (47.1)</td>
</tr>
<tr>
<td>Diabetes no (%)</td>
<td>342 (21.6)</td>
<td>406 (25.9)</td>
</tr>
<tr>
<td>Heart failure no (%)</td>
<td>456 (28.8)</td>
<td>425 (27.1)</td>
</tr>
<tr>
<td>Mean (SD) no. chronic conditions</td>
<td>1.8 (1.80)</td>
<td>1.8 (1.78)</td>
</tr>
<tr>
<td>Mean (SD) Combined Model score</td>
<td>0.26 (0.20)</td>
<td>0.26 (0.20)</td>
</tr>
</tbody>
</table>
Telehealth arm: fewer emergency admissions

Percentage difference in emergency admissions: -20.6% (95% CI -33.8% to -7.4%)
• BMJ paper: “..results suggest Telehealth helped patients avoid need for emergency admissions..”

• Department of Health: "We funded a three-year randomised control trial..which clearly demonstrated that if implemented appropriately, telehealth can reduce emergency admissions by 20%...”

• David Cameron: "We've trialled it, it's been a huge success, and now we're on a drive to roll this out nationwide.”
Did RCT control arm have usual care?
Emergency admissions: before and after RCT
WSD
re-analysis to improve external validity

Step 1. Define target population from observational data, patients who met RCT inclusion criteria and received usual care (n=88,830)

Step 2. From target population find individuals (n=1,293), who match RCT controls (n=1,293), on 65 baseline covariates

Step 3. Placebo test: contrast outcomes from matched target population (n=1,293), vs RCT controls (n=1,293) (A)

Step 4. Sensitivity analysis: re-estimate Telehealth effectiveness
RCT Telehealth arm (n=1,229) vs matched target population (n=1,229) (B)
## WSD: placebo results

In incidence rate ratios.

<table>
<thead>
<tr>
<th>Placebo tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
</tr>
<tr>
<td>RCT (control) vs.</td>
</tr>
<tr>
<td>matched target population (usual care)</td>
</tr>
<tr>
<td>Emergency admissions per head</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Outpatient attendances per head</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

\(^a\): passes placebo test
WSD: re-analyses
Incidence rate ratios.

<table>
<thead>
<tr>
<th>Placebo tests</th>
<th>Estimated effect of Telehealth</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) RCT control arm vs target population (usual care)</td>
<td>(B) RCT Telehealth arm vs target population (usual care)</td>
</tr>
<tr>
<td>Emergency admissions per head</td>
<td></td>
</tr>
<tr>
<td>1.22 (1.05, 1.43)</td>
<td>1.12 (0.95, 1.31)</td>
</tr>
<tr>
<td>Outpatient attendances per head</td>
<td></td>
</tr>
<tr>
<td>1.03 (0.94, 1.13)</td>
<td>1.04 (0.95, 1.14)</td>
</tr>
</tbody>
</table>

$^b$ we use data from 1229 in the Telehealth arm and 1229 in the control arm to ensure consistency across the re-analyses. See Steventon et al (2015) for details.
Interpretation and implications (II)

- Effective from RCT, overall and for subgroups
- But external validity assumptions fail
- Could be differences in unobservables vs target population
- RCT control arm did not receive usual care (placebo effect)
- More generally, better outcomes RCT control arm vs usual care
- If RCT controls do not receive usual care, other designs required
  - Expose structural uncertainty
  - RCT assumes: control therapy = usual care and no unobserved confounders
  - Matched analysis assumes no unobserved confounding
Are we paying attention to what matters?
What’s the gorilla in room?

• RCTs may be poor at predicting treatment effects for populations
• Especially for interventions with strong behavioural component
• Preferences may be vastly different
June 14, 2017: world blood donor day

World Blood Donor Day 2017


June 2017 -- This year's campaign will focus on blood donation in emergencies. In crisis or emergency situation, the natural human response is "What can I do? How can I help?" Therefore, the slogan for the 2017 campaign is: What can you do?, with the secondary message: Give blood. Give now. Give often.

Posters
Campaign materials
Graphs for social media

112.5 million
112.5 million blood donations are collected globally, half of these are in high-income countries.

1%
Blood donation by 1% of the population can meet a nation's most basic requirements for blood.

57 countries
57 countries collect 100% of their blood supply from voluntary, unpaid blood donors.
Problem 3: preferences may differ

Example economic evaluation of alternative blood collection services

- Concerns supply whole blood insufficient to meet demand
- Trial considers strategy: reduced intervals between donation
- Mismatch with policy-maker, interested in wider range of strategies.
- Health report, evening opening for the centres etc
- Also concern trial participants a typical preferences.
- Health Economics Modelling of Blood donation (HEMO) study
  - Wide range of strategies
  - Stated Preference (SP) Surveys
  - Population vs trial
### SP survey for blood collection: attributes and levels

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opening times</td>
<td>- 9am - 5pm&lt;br&gt;- 9am - 8pm&lt;br&gt;- 2pm - 8pm</td>
</tr>
<tr>
<td>Health report</td>
<td>- Provided&lt;br&gt;- Not provided</td>
</tr>
<tr>
<td>Maximum donations per year</td>
<td><strong>FEMALE</strong>&lt;br&gt;- 3&lt;br&gt;- 4&lt;br&gt;- 6&lt;br&gt;<strong>MALE</strong>&lt;br&gt;- 4&lt;br&gt;- 5&lt;br&gt;- 6</td>
</tr>
</tbody>
</table>

**Outcome:** frequency of donation
Method steps

- Stated and revealed preferences similar (De Corte et al, 2016)
- Hypothesis: stated frequency higher RCT versus not
- Responses survey ex-RCT participants (9,000) vs population (24,000)
- Estimate propensity being in RCT, by gender
- Reweight trial data, by gender
- Contrast reweighted preferences trial vs population (placebo test)
- Cost-effectiveness results: RCT sample vs population
Incremental effect (95% CI) of service change on: average donations

Females: RCT unweighted ▲

Last donation experience (baseline)

Health report

Venue Opening times (currently 9am-5pm)
- 9am - 8pm
- 2pm - 8pm

Max number of donations (currently 3)

Incremental effect of change to blood service on average donations
Incremental effect (95% CI) of service change on: average donations

**Females:** RCT unweighted ▲ vs weighted♦ vs Population ●

- Last donation experience (baseline)
- Health report
- Venue Opening times (currently 9am-5pm)
  - 9am - 8pm
  - 2pm - 8pm
- Max number of donations (currently 3)

Incremental effect of change to blood service on average donations per year
Incremental effect (95% CI) of service change on: average donations

**Males:** RCT unweighted ▲ vs weighted♦ vs Population ●

- Last donation experience (baseline)
- Health report provided
- Opening times (currently 9-5)
  - 9am - 8pm
  - 2pm - 8pm
- Max number of donations (currently 4)
  - 5
  - 6

**Incremental effect change to blood service on: average number of blood donations per year**
Interpretation and implications

• Small heterogeneity in preference RCT versus population
• Overall framework using observational data for external validity
• Avoids assume homogeneity of preferences in and out of trial
• Large scale observation data potential to test RCT assumptions
• Complementary vehicle for estimating treatment effects
• Insights on why RCT might not apply widely e.g. non-adherence
• Implications target (personalise) treatments in decision-making
Implications - health economist

- Imaginative use of large observational data
- Design focuses on causal effect for decisions matter
- Insights about populations/treatments/behaviours apply widely
- So far behaviour, CEA, and policy worked in silos
- Complex interventions - behaviour is key
  - patient-surveys,
  - observational data
  - and trials
- Provide useful CEA
- For populations and subpopulations relevant for policy
wider conclusion

“a guidebook is no substitute for skill, experience, judgement and lots of tension.” Charlie Fowler

“I've climbed with some of the best climbers in the world, more importantly, to me, they are some of the best people in the world. That's another reason why I climb.” Jim Wickwire
Acknowledgements

- Jas Sekhon
- Adam Steventon
- Kaat De Corte
- Noemi Kreif
- Stephen O’Neill
- John Cairns
- Anirban Basu
- The HEMO study
- Nick Black
- Roland Ramsahai
- NIHR for funding
The mystery guest is..
References


