PH458

Case Study: Medicine

Dr. Ioannis Votsis
LAK3.01 (Lakatos Building)
Office hours: Tuesdays 15:00-16:00
i.votsis@lse.ac.uk
www.votsis.org
Introduction
The movement and its aims

- Evidence-Based Medicine (EBM) is a relatively recent movement in the practice of medical research.

- The name seems to suggest that prior to the advent of this movement medicine was not evidence-based.

- This is obviously incorrect. Perhaps a better name would be: ‘Standardised-Evidential Practice in Medicine’.

- EBM emphasises the import of standards that rank sources of evidence in a unified manner across medicine.

**NB:** Medicine here refers to the science and practice.
Prior to the EBM movement, there were many different approaches to diagnoses, treatments and prognoses.

The history of medicine is replete with such cases where treatments were either ineffective or down-right harmful.

Example:

blood-letting (as a cure all)

vitriol (as a cure all)

NB: Each can of course be good for some things.
EBM in Detail
EBM-ers target the overreliance on:

(i) clinical experience and judgment
(ii) background theory

They make a number of recommendations, including:

* Regular updating of the knowledge of medics.
* Making results of clinical trials more widely known.
* Deciding what constitutes the best available evidence.

Our focus: The last of these, i.e. evidential weight.
Only evidence sourced from RCTs establishes anything in medical research.

**NB**: In short, no RCT → no real evidence.

That means traditional sources of evidence like clinical experience and judgment should be dismissed.

Sackett et al. (2000), for example, assert that “If the study was not randomized we’d suggest that you stop reading it and go on to the next article in your search” (p. 108).
But it is clear that there are many cases of therapies that do not enjoy RCT evidence and yet seem to work just fine.

**Examples:**
- penicillin in cases of (bacterial) pneumonia
- diuretics in cases of heart failure
- appendectomy in cases of acute appendicitis

In some of these cases, the lack of RCTs is down to ethical considerations; in others, there seems to be no need.
Other types evidence may be available and should be consulted but evidence sourced from RCTs trumps them.

To be precise, there is a hierarchy of evidence and RCTs sit at the top or very near it.

“EBM is not restricted to randomized trials and meta-analyses. . . . some questions about therapy do not require randomized trials (successful interventions for otherwise fatal conditions) or cannot wait for the trials to be conducted. And if no randomized trial has been carried out for our patient’s predicament, we must follow the trail to the next best external evidence and work from there” (Sackett et al. 2000, p. 72).
The R in RCT

• Why are RCTs such a respected source of evidence?

• Beyond the benefits you can get from CTs, e.g. placebo control and double-blinding, you also get randomisation.

• Recall that randomisation is meant to create *ceteris paribus* conditions between the control and treatment groups.

• It is meant to do that by evenly distributing known and unknown confounding factors between the two groups.
Internal Validity Problems
- **Internal validity**

  A study is internally valid when its results are valid for the study population.

  In more detail, it is internally valid when the treatment’s impact is accurately measured in the study population.
Questioning the epistemic power of RCTs

• Worrall (2002; 2007a; 2007b) argues that the epistemic power of RCTs is overstated.

• He doesn’t deny their usefulness but dismisses their claimed superiority over other kinds of evidence sources.

• One of the arguments he criticises is that randomisation guards against all (i.e. known and unknown) confounders.
Randomisation offers no guarantees

- Worrall uses a well-known ‘study’ to promote his view.

**Leibovici et al (2001):**
Records of 3393 inpatients at Rabin Medical Centre (Israel) admitted because of bloodstream infections 1990-1996.

In 2000, random number generator assigns patients into 2 groups: control group (nothing), ‘treatment’ group (prayer).

Hospital stay length and duration of fever much lower in the ‘treatment’ group. **Conclusion:** ‘Treatment’ is effective.

- **Upshot:** Clearly, the ‘treatment’ is *not* effective. Thus, Worrall claims, the R in RCT is no guarantee of effectiveness.
Indefinitely many confounders?

- The underlying reason why Worrall is doubtful of the epistemic power of RCTs is his belief that:

  “... the list of possible ‘unknown’ factors is indefinitely long...” (1007-8).

- Suppose, for the sake of the argument, that unknown confounders are indeed numerous in a given study.

- Then the probability that any randomisation will evenly distribute all of them goes down with the size of the study.

- **Warning:** We should not assume their numerosity a-priori! Invariant results over multiple RCTs count against this.
Beyond not being able to guarantee the even distribution of confounders there is also a problem relating to strength.

Suppose:
* a study tracks 100 patients
* 20 of them possess the same unknown confounder
* no other known or unknown confounders
* randomisation distributes these evenly

Does this mean that any positive result can safely be attributed to the intervention (i.e. treatment)?

Alas, no! The same confounder may manifest itself in different strengths, e.g. variable immune response.
• Note that the problems of the number and strength of confounders affect results across the board.

• That is, they can undo positive, negative & neutral results.

<table>
<thead>
<tr>
<th>Real Intervention Effect</th>
<th>Observed Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
</tr>
<tr>
<td>Neutral</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
</tr>
<tr>
<td>Negative</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
</tr>
</tbody>
</table>
External Validity Problems
• External validity

A study is externally valid when its results are valid for the general population.

In more detail, when the treatment’s impact accurately reflects what would happen also in the general population.
• Cartwright and Hardie (2012) are also critical of RCTs. But they don’t focus on issues of *internal validity*.

• On the assumption that RCTs can (and in many cases are) done well, they question whether their results *generalise*.

• That is, they worry whether they possess *external validity*.

• Instead, they take them to be mere starting points in figuring out whether a policy will work here as it did there.
• RCTs can be likened to black-boxes:

“You put the drug in, and out comes a cure. But you get no idea from the RCT how that happened” (p. 123).

• In Cartwright and Hardie’s language, RCTs provide no information about causal roles and support factors.

Support factors:
* whether there are any?
* if so, what they are?
* what their effect is?
• Cartwright and Hardie insist that answering the ‘how does it work?’ question is of paramount importance.

• On their account, this involves knowing something about the causal role and support facts in the particular locale.

• To continue with the black-box metaphor, this would involve opening the box and looking inside.
To emphasise the difficulty of this task in the current context, Cartwright and Hardie compare it to two cases.

- In the **case of machines**, e.g. cars, the manufacturer knows how they work even if most consumers don’t.

- Plus, the cars are continually tested in various locations – we know where they work and where they don’t.

- In the **case of physics**, scientists routinely construct shields to prevent confounding factors from interfering.

*Example*: Detecting neutrinos.
The human context

• In the context of dealing with human test subjects, by contrast, no such options are available.

Examples:
We are unable to study the effectiveness of a non-invasive treatment in complete isolation from ‘human vehicles’.

We are unable to test a great number of people continually.

• As a consequence, result exportability – that is, external validity – is not guaranteed.

• That presumably makes background knowledge about causal roles and support facts all the more pressing.
Teasing out relevant information

- Cartwright and Hardie discuss an anecdotal case found in Gray Klein’s *Sources of Power* (1999: 39-44).

- An experienced nurse can detect a potentially fatal condition in prematurely born babies in time for treatment.

  NB: She is much better than standard diagnostic methods.

- The nurse doesn’t know exactly how she does it but via reflection manages to come up with this list of symptoms: change in colour; heightened activity; poor appetite.

- **Point**: Exportability only when we answer how-questions.
Putative solution: More RCTs

- One of the reasons that vetting and policy warehouse agencies urge RCTs is that they are designed to be impartial.

- That’s why they are considered to be generally more trustworthy than background information.

- Thus, in answering the question ‘what’s better than an RCT?’, the answer is ‘two or more RCTs’.

- **Rationale:**

  The more we test different sub-pops the less likely it is that a treatment’s effectiveness is due to local confounders.
Cartwright and Hardie agree that more RCTs help but still think that their evidential impact is indirect and conditional.

Why conditional? Because they rely on the assumption that the different locales are relevantly similar.

“Lots of positive RCT results are a good indicator that the policy plays the same causal role widely enough to reach to you. But you are always betting on a hidden premise: that the studies vary across enough different kinds of circumstances to generalize” (p. 127).
Isn’t the whole point of multiple RCTs (whose judgments converge) that they lend support to that assumption?

In other words, isn’t such convergence across RCTs a good reason to think that results are generalisable?

The inference is, of course, inductive and hence defeasible but the point remains that some support seems to accrue.

Turning the tables on Cartwright and Hardie, there is no need to deny that causal information can help.

The question is whether multiple RCTs are frequently sufficient to establish external validity?
The End