PH 458/231
Evidence & Policy
Lecture 8

John Worrall
LAK 3.02
Office Hours:
Monday 13.30-14.30
Thursday 12.30 – 13.30

Testing Causal Hypotheses

- Ash tray ownership (A) is ‘merely associated with’ cancer (C) and is not a cause of it.
- How do we know?
- 1. because we have some background knowledge about causes
- 2. we can tell by conditionalising and taking further data
- Divide the population into S and not-S then the association between A and C ‘disappears’ (is ‘screened off’) :
- Although P(C/A) > P(C)
- When you ‘factor in’ smoking
- P(C/S &A) = P(C/S) (= P(C/S & not-A)

Testing Causal Hypotheses

- On the positive side, we do think that there is evidence that smoking causes cancer
- What did Bradford Hill and Doll do to provide that evidence?
- Controlled for lots of alternative explanations
- One such ‘air pollution’ (P)
- P(C/S & P)> P(C/P)
- (and P(C/S & not-P) > P(C/not-P)

Testing Causal Hypotheses

- Nature of science that no matter how strong, evidence is always defeasible
- Fisher's (and Eysenck's) 'genetic hypothesis'
- Why this remains unscientific/without evidence.

Randomization and Causes

- So back to the claim that only if you have randomized do you have evidence of a causal connection.
- Well, you can certainly have evidence for *lack* of a causal connection without randomising:
- A and B seem to be probabilistically dependent, but you conjecture the ‘real cause’ C, and observe that C ‘screens off’ A from B.

Randomization and Causes

- Positive evidence (in the apparently paradigm case) comes from failed attempts to 'screen off'
- But always the fear: remember the Fisher/Eysenck hypothesis
- ***Hope***: that randomization will deal with this?
- And any other possible screener off??
- So in the end these arguments are just re- expressions within (different) causal frameworks of the 'controls for all factors known and unknown'
- Response the same: it *might* do! And can't make things worse (ceteris paribus)
- BUT a mistake to think that it *will* or even, perhaps, that there is some sort of probabilistic guarantee that it will do so 'with high probability' *in this case*
- AND a big mistake to leave controlling for 'known' factors to chance

External Validity

- What's the problem?
- "Internal validity"
- Wikipedia: **Internal validity** is a property of scientific studies which reflects the extent to which a causal conclusion based on a study is warranted. Such warrant is constituted by the extent to which a study minimizes systematic error (or 'bias').

- Let's suppose – for the sake of this argument - that RCTs are most 'internally valid'
- But what *exactly* is the causal conclusion "warranted" by an RCT?

External validity

- Typical research reports:
- ‘Efficacy and safety of ustekinumab .. in patients with psoriasis..’
- ‘Active symptom control with or without chemotherapy in the treatment of patients with malignant pleural mesothelioma ..’
- Sounds very general: ‘ustekinumab *is effective* for psorias’ etc
- BUT IN FACT
- They report (usually randomized) trials on some often highly *selected group* of patients
- selected by quite complicated inclusion & exclusion criteria
- generally using some *very precise treatment regimen;* and where
- the treatment is given for some *relatively brief period*
- (Rawlins ‘ Most RCTs, even for interventions that are likely to be used by patients for many years, are of only six to 24 months duration.’)

External validity

- Result maybe stated – as noted - as ‘drug is ‘more effective’ than the comparator for condition C’
- What exact theory has been tested?
- Not really this (vague) claim but rather:
- D when administered in a very particular way to a very particular set of patients for a particular length of time is more effective than some comparator treatment (perhaps placebo).
- RCT provides – let’s say impeccable- evidence for this

External validity

- Result maybe stated as ‘drug D is more effective than the comparator for condition C’
- What exact theory has been tested?
- Not really this (vague) claim but rather:
- D (i) when administered in a very particular way to (ii) a very particular set of patients for (iii) a particular length of time is more effective than some comparator treatment (perhaps placebo).
- RCT provides – let’s say impeccable- evidence for this

External validity

- But this is not the claim that a practitioner is interested in.
- She wants evidence for:
- D is effective (in a wide sense) when prescribed to the sorts of patients she would like to prescribe it to: the ‘Target population’
- This will include the excluded (in particular those with comorbidities)
- For that population, dosage may be adjusted
- Patients in that population may not be so good at remembering to take the drug etc
- Often for chronic conditions and so interested in long-term effects

External validity

- So the problem of ‘external validity’ really starts from the fact that the ‘wrong’ theory is tested
- NB not a Humean ‘purely philosophical’ issue
- Specific grounds for thinking the target and study populations different
- And background knowledge tells us that these factors may well play a role in outcome.
- Bartlett et al looked at RCTs on NSAIDs (25) and Statins (27) and found older people, women and ethnic minorities all consistently underrepresented
- [NOTICE: randomization has no conceivable impact on *this* issue.]

External validity

- Constructive proof: Benoxaprofen (Opren)
- NSAID (early ’80s) – for arthritis/musculo-skeletal pain (once a day)
- Trial (18-65 yrs) →
- Big positive result and cornered market
- ***Fact***: musculo-skeletal pain predominantly afflicts the elderly.
- Turned out that *among the elderly* treated with Opren there were a marked number of deaths from hepatorenal failure.
- Opren was withdrawn.
- (NB no one has ever suggested that within the sub-population ‘represented’ by the original trial group Opren has anything other than a marked positive effect.)
- Plenty of other examples: e.g. Cerivastatin (Rawlins lists well over 20 treatments endorsed by ‘high quality’ RCTs that had later to be withdrawn as useless (or worse))

External validity

- In GISSI study (statins for ischaemic heart disease) only 45% of the patients with heart disease who attended the coronary care units involved were randomized
- A back up study showed excluded group had almost twice the mortality of the included group
- (N.B.Absolute risk reduction allegedly found by GISSI was 1.3%)
- “Run in” periods
- Preliminary blood tests for markers deemed relevant
- Not as if the exclusion criteria are always precise
- ASSENT-2 trial: “any other disorder that the investigator judged would place the patient at increased risk”
- ISIS-2 trial: “any further reason for exclusion not specified by the protocol but by the responsible physician”

External Validity

- So what's to be said about external validity?
- 1. Study population is not a random sample of the 'target population' so no statistical argument that the result 'probably' holds generally.
- 2. Certainly cases where it has failed to hold
- 3. Guyatt holds that e.v. is the default assumption
- 4. Maybe something to this – but certainly know of categories where the default may break down
- 5. Background knowledge gradually supplies a list of factors we would at least want to check
- IN ANY EVENT – another question to be added to the list for evidence-savvy administrators: ' even if the result is valid for the study population, is there any reason to think it is valid for the 'target population'?'

Social Science examples

- Cartwright has written on ‘the long road from “it works (somewhere)” to ‘it works for us’
- Regards it as a different problem from external validity – but ????
- Examples:
- 1. World Bank wanted to address malnutrition in developing countries
- BNIP (Bangladesh Integrated Nutrition Project) modelled on
- TINP (Tamil Nadu Integrated Nutrition Project)
- Centrally involved education of pregnant mothers about child nutrition
- Cartwright a little unclear about whether TINP was an RCT but in any event
- TINP was a great success
- BNIP was not

Social Science examples

- 2. California Class Size Reduction Program
- California based its policy (costing $1bn p.a rising to $1.6 bn p.a.) on the success of STAR
- STAR was an RCT on the effect of class size reduction in Tennessee
- Result: students in smaller classes had better scores on all educational level tests and in particular minority & inner city students gained 2 or 3 times as much as white & nonurban peers.
- California policy a failure: ‘no conclusive link between class-size reduction & student achievement; and in particular no greater effect amongst disadvantaged students.’

Social Science examples

- Perhaps not quite the same as the medical cases?
- But raise similar questions – suppose the initial trial result ‘internally valid’ it is a separate issue whether the result ‘generalises’ to other situations.
- What went wrong?
- 1. Essentially different social structure in Bangladesh compared to Tamil Nadu – it was the mother-in-laws rather than the mothers who controlled the purchase, preparation and distribution of food.
- 2. a. In Tennessee project involved only schools with available extra space, in California this had to found and was largely substandard and taken away from other activities (special needs, music & arts, athletics..)
- b. In Tennessee no shortage of teachers to staff the extra classes, but Ca had to hire an additional 12000 teachers, many of whom were unqualified.

Social Science examples

- The US Education Department ‘ What Works Clearing House’ website has a list of ‘well-conducted’ RCTs on educational interventions
- And advises you that those interventions will work for you if the RCTs were performed in ‘school settings similar to yours’
- But how do you know?!
- Cartwright gives a very elaborate account of the ‘road from “it works (somewhere)” to “it works for me”’ involving ‘causal cakes’, ‘INUS conditions’, ‘stable capacities’..
- BUT seems to amount in the end to the view that if you want evidence that it will work for you/of external validity then you had better have evidence *outside the RCT* about causal mechanisms
- IF you don’t have that then you don’t really have any evidence that it will work in this other environment

External Validity

- You might say that WHILE in the medical cases there is background knowledge that biology usually generalises and so even in the absence of any specific knowledge about causal mechanisms, you can at least have some grounds for thinking the treatment might work in the target population
- This is NOT SO in the social science cases because societies are orders of magnitude more complex that biological organisms
- ??? But in the cases Cartwright cites, things are not so gloomy…
- In any event:
- Definite lesson for the evidence-savvy administrator – can I trace some sort of evidential path from the conditions holding in the system studied to the system to which I am thinking of applying the same intervention?