PH 458/231
Evidence & Policy
Lecture 7

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

Argument 2: “controlling for all factors”

- Thinking back to our general consideration of the value of ‘controlling’
- We see the big attraction of randomizing – the idea that it might control for all possible ‘confounders’: ‘known’ and ‘unknown’
- Mike Clarke: “In a randomised trial, the *only* difference between the two groups being compared is that of most interest: the intervention under investigation.”
- Rawlins: “The greatest strength of an RCT is that the allocation of the treatments is random so that the groups being compared are similar for baseline factors.”
- “Other study designs … can detect associations between an intervention and an outcome. But they cannot rule out the possibility that the association was caused by a third factor linked to both intervention and outcome. Random allocation *ensures* no systematic [?] differences between intervention groups in factors, known and unknown, that may affect outcome.” (Sibbald and Roland, *BMJ,*1998,201)

The RCT as gold standard?

- Wikipedia:
- As their name suggests, RCTs involve the random allocation of different interventions (treatments or conditions) to subjects. As long as numbers of subjects are sufficient, this ensures that both known and unknown confounding factors are evenly distributed between treatment groups.

The RCT as gold standard?

- ‘Otherwise equivalent’ claim obviously false
- Clearly not ‘ensured’ in a *single* randomized division
- Quietly conceded by the injunction to look for ‘baseline imbalances’ (in ‘known’ confounders)
- But if possible in known then clearly also in ‘unknown’

The RCT as gold standard?

- Amusing example
- Leibovici et al “Effects of remote, retroactive, intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial” *BMJ* 2001

The RCT as gold standard?

- 3393 patients having a bloodstream infection while being an inpatient at the Rabin Medical Centre during 1990-6 were identified.
- In July 2000 a random number generator was used to divide these patients into two groups and which of these two became the treatment group was decided by a coin toss.
- 1691 were randomized to the intervention group and 1702 to the control.

The RCT as gold standard?

- Checked for baseline imbalance with regard to main risk factors for death and severity of illness.
- The names of those in the intervention group were given to a person ‘who said a short prayer for the well being and full recovery of the group as a whole.’
- Results: both length of stay in hospital and duration of fever were significantly shorter in the intervention group ($p = 0.01$ and $p = 0.04$)!

The RCT as gold standard?

- Conclusion: ‘Remote, retroactive intercessory prayer said for a group is associated with a shorter stay in hospital and shorter duration of fever in patients with bloodstream infection and should be considered for use in clinical practice.’
- Somewhat tongue-in-cheek of course (‘No patients were lost to follow up’!!)

The RCT as gold standard?

- Natural reaction shows we are all Bayesian (?):
- “If the pre-trial probability is infinitesimally low, the results of the trial will not really change it, and the trial should not be performed. This, to my mind, turns the article into a non-study, though the details provided (randomization done only once, statement of a prayer, analysis, *etc*) are correct.”
- Fisher a natural but egregious mistake

The RCT as gold standard?

- Kosher trial:
- Proper randomization
- No data mining
- Baseline imbalances
- But …

The RCT as gold standard?

The RCT as gold standard?

- Nobody believes surefire guarantee?
- Instead ‘probabilistic guarantee’
- But what does this amount to?
- ‘Either groups are equivalent or a chance event has occurred’?

The RCT as gold standard?

- Slip from what is true in the indefinite long run of many repeated random divisions, to what is (allegedly) true in the single random division
- Single case probabilities?
- Even if you are willing to accept that there is a plausible argument to the effect that for ***any single unknown confounder***, it will probably be evenly distributed between the two groups
- There are indefinitely many possible unknown confounders – you would want some "probabilistic guarantee" that they are all evenly distributed
- ??? (cp Howson and Urbach)

The RCT as gold standard?

- Bayesian position:
- Only legitimate role of randomization is elimination of selection bias
- Based on commonsense principle that ***you can't be in a better epistemic position than that of having no plausible reason to think that the control and experimentally groups are importantly different***
- And you may – laying aside selection bias – be in that position without having randomized
- (And if you 'simply randomized' you should ask yourself that question anyway.)

“last men standing”: (a) the argument from selection bias

- So, remember, there were 5 arguments for the claim that RCTs carry special evidential weight
- 1.Fisher: randomizing essential to underpin the logic of the significance test
- 2. Randomizing controls for all possible confounders known *and unknown*
- 3. Historical fact that non-randomized studies exaggerate positive effects
- 4. Randomizing controls for the specific possible confounder – ‘selection bias’
- 5. By randomizing, and only by randomizing, you get evidence that any observed positive effect in the trial is *caused* by the treatment rather than merely being associated with it.
- 1- 3 arguably don’t stand up to critical scrutiny
- We are left with 4 and 5

“last men standing”: (a) the argument from selection bias

- Undoubtedly a solid argument
- Though it is not the randomization as such that is doing the work
- But rather the fact that the division into ‘experimental’ and ‘control’ *is taken out of the hands of the experimenters*
- Can’t selection bias, or any rate any great amount of it, be effectively ruled out by other means?
- Cp ECMO again

“last men standing”: (b) evidence from RCTs underwrites *causes*

- This takes us to our next topic
- Is certainly widely believed
- “Other study designs … can detect associations between an intervention and an outcome. But they cannot rule out the possibility that the association was caused by a third factor linked to both intervention and outcome. Random allocation *ensures* no systematic [?] differences between intervention groups in factors, known and unknown, that may affect outcome.” (Sibbald and Roland, *BMJ,*1998,201)
- “[When, but only when] the treatment has been assigned at random – in the sense that all patients, whatever their other characteristics, have exactly the same objective probability of receiving the treatment T – we can now be **sure** that T is *not* itself objectively correlated with any other characteristic that influences R[ecovery].” (David Papineau, *BJPS*, 2000)

Randomization and causes

- Let’s first take a step back and think more carefully about the differences between a causal and a (“merely”) probabilistic/statistical hypothesis

Testing Causal Hypotheses

- What does ‘smoking causes lung cancer’ mean?
- Clearly *not* anything deterministic
- Something to do with P(lung cancer/ smoke) > P (lung cancer/ ¬ smoke)
- But that can’t be its full meaning
- 1. Probabilistic dependence is symmetric, cause is asymmetric
- 2. Back to *post hoc ergo propter hoc:*
- Smoking causes cancer and certainly P(C/S) > P(C)
- [P(C/S) > P(C) ≡ P(C/S) > P(C/not-S)]
- But let A = owns (say) more than 2 ashtrays,
- Then P(C/A) > P(C)
- BUT clearly ashtray ownership does NOT cause cancer

Testing Causal Hypotheses

- How do we know?
- A. because we have some background knowledge about causes
- B. 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’) :
- P(C/S &A) = P(C/S) (= P(C/S & not-A)
- On the positive side, we do think that smoking causes cancer
- And that we have strong evidence for believing so.
- 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’
- AND a big mistake to leave controlling for ‘known’ factors to chance
- Moreover, as the reference to Hill and Doll shows, there are other ways of obtaining telling evidence for causality that do not involve randomization.