Subgroups and personalisation:

Richard Grieve
richard.grieve@lshtm.ac.uk

content

• Overview of the problem
• The limitation of standard approaches to generating evidence, resting on RCTs
• Outline of the 3 talks
• Future challenges
Towards personalization, why

• Drive to better target treatment for individual patients
• This is described as stratified, precision, or personalised medicine
• Right treatment for right patient at the right time
• Clinical decision-makers intend to practice personalisation
• Interest in estimating heterogeneity of treatment effects (THE)
• Thrombolysis for acute stroke can improve or harm
• Want to find ‘true’ positives
• If ‘miss’ subgroups waste resources, minimise ‘false’ negatives
• See Espinoza et al, 2014

The problem - multiplicity

• Concerns about overfitting and multiple-testing
• 20 tests, will judge 1 as ‘statistically significant’ just by chance
• Star sign ‘found’ modify effect aspirin after MI (see Horton 2001)
• If we do not take account of the number subgroups tested when selecting and estimating subgroup effects
  – Estimates of differences between subgroups over-estimated...
  – uncertainty in these estimates will be under-estimated
  – Concerns about falsely identifying subgroups (minimize false positive)
The concern about false positives

Summary of problem

<table>
<thead>
<tr>
<th>Claim</th>
<th>Truth</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hetero</td>
<td>Hetero</td>
<td>FALSLEY IDENTIFY SUBGROUPS</td>
</tr>
<tr>
<td>No Hetero</td>
<td>MISS TRUE SUBGROUPS</td>
<td></td>
</tr>
</tbody>
</table>
Approach might differ by context

- Consider decision-making context
  - Inform regulatory decisions
  - Inform adoption and re-imbursement decisions
  - Inform clinical decision-making
  - Inform decisions regarding the conduct and design of future clinical studies

There are tradeoffs: where do you stand?
Are you a Karl, a Robert, or a Peter?

Karl Claxton: NHS will tend to get more net health benefit by considering finer definitions of subgroups.

Robert Hemmings: ...routinely dismissing results of subgroup analysis, is no scientific solution. It is important to realize that both action and inaction represent decisions.

Peter Sleight: Undue emphasis on a particular subgroup may result in inappropriate treatment.

The poll
https://myispor.cnf.io/sessions/gjbs/#!/dashboard
Approach to subgroups.
Which one would you prefer?

1. 5% false negatives, 95% false positives [Karl]

2. 50% false negatives, 50% false positives [Robert]

3. 95% false negatives, 5% false positives [Peter]
Harnessing causal inference..
Predicting individual-level treatment effects
(Rubin 1977, Holland 1986)

- $T_i$ is treatment indicator: 1 treatment group, 0 control
- Interested in causal relationship between $T_i$ and $Y_i$
- Individual, $i$ potential outcomes $Y_{i0}$ and $Y_{i1}$ under control and treated states
- Ideally observe treatment effect for each individual $\tau_i = Y_{i1} - Y_{i0}$
- BUT cannot observe both outcomes
- **Objective of methods:** impute missing potential outcome
- **Recognize each individual may have different response to treatment.**

RCTs inadequate

- Bradford Hill, 1960: ‘do not answer the concern about the most likely outcome when this particular drug is given to a particular patient’
- Must recognize and estimate heterogeneous treatment effects

Problem 1: lack of power
- RCTs powered for main treatment effects, to give similar power for interaction effects, four-fold increase in sample size
- Subgroup interactions, 20 binary characteristics, 1 million subgroups
- Low proportion true positives, low power, high rate false discoveries

Problem 2: ignores effect modifiers not unobserved
FDA Commentary

“Evidence efficacy in US population equivocal. ... PLATO not designed specifically to show evidence efficacy compared to clopidogrel in the US only.

Several potential explanatory factors explored including: compliance, statin exposure, low ticagrelor exposure, chance finding.... None satisfactorily explained the observed benefit of clopidogrel over ticagrelor in the USA.

“Although I consider the likelihood that the US/OUS was chance. .. I believe evidence aspirin dose explains difference powerful further basis for approval...”

By a 7 to 1 vote FDA recommended approval
Warren Stephens

• Questioning current methods of evidence generation
• Highlighting the trade-offs that are made implicitly with the limitations of current data
• How to better align incentives for regulators/payers/clinicians to make more efficient (equitable) decisions

Neil Hawkins

• Vital to incorporate prior beliefs about anticipated subgroup effects
• Incorporate that into fully Bayesian approaches.
• Bayesian approach relies on ‘valid’ expert elicitation
Anirban Basu

- Heterogeneity according to unobserved and observed factors
- Instrumental variable approach can fully explore heterogeneity
- Harnesses this with large-scale RWE
- Rests on valid, continuous instruments, large data, and some parametric assumptions

Future research agendas

- Want to push out the ROC curve..
- Require adaption of appropriate methods from causal inference
- Machine learning (e.g. LASSO), try and avoid false positives.
- See also Imbens and Athey ‘honest confidence intervals’

- RWE alone will not save the day
- Careful testing and evaluation of methods also essential
References


Horton R. From star signs to trial guidelines (2000). Lancet 355:1033-34


Athey S; Imbens G (2016). Recursive partitioning for heterogeneous causal effects. Proc Natl Acad Sci USA 113(27):7353-60 (ISSN: 1091-6490)