Introduction to Generalized Propensity Scores (GPS)

Anna Åkesson

EPI@Lund, Department of laboratory medicine, Lund University

&

Clinical Studies Sweden – Forum South

Unit for Medical Statistics and Epidemiology

Skåne University Hospital

Outline

- Background
- Propensity scores (PS)
- Generalized Propensity Scores (GPS)
- Things to keep in mind

Background - RTCs

- Randomized controlled trials (RCTs) are considered the gold standard for estimating treatment/exposure effects
- Random treatment allocation ensures that treatment/exposure status will not be confounded with either measured or unmeasured baseline characteristics

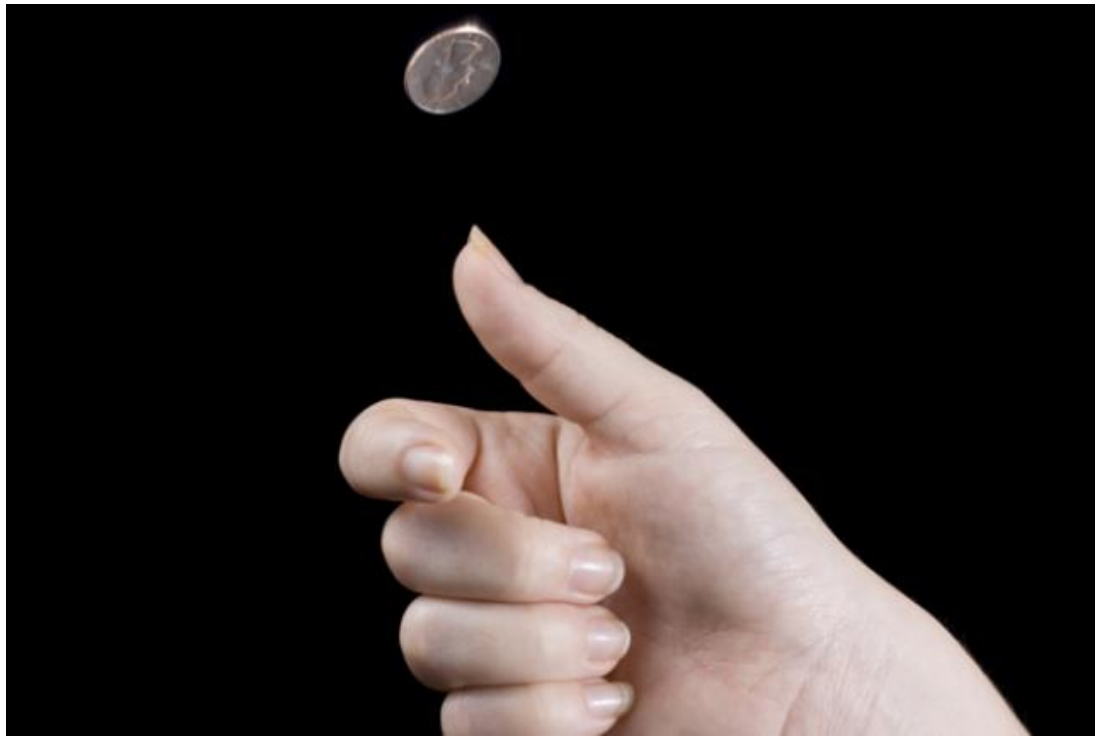


Image: Unknown; CC BY

Background - RTCs

- In RTCs the probability of exposure is 50/50 between the exposed and unexposed
- Therefore, we say that we have exchangeability between groups
- This exchangeability between groups is critical to the claim we want to make

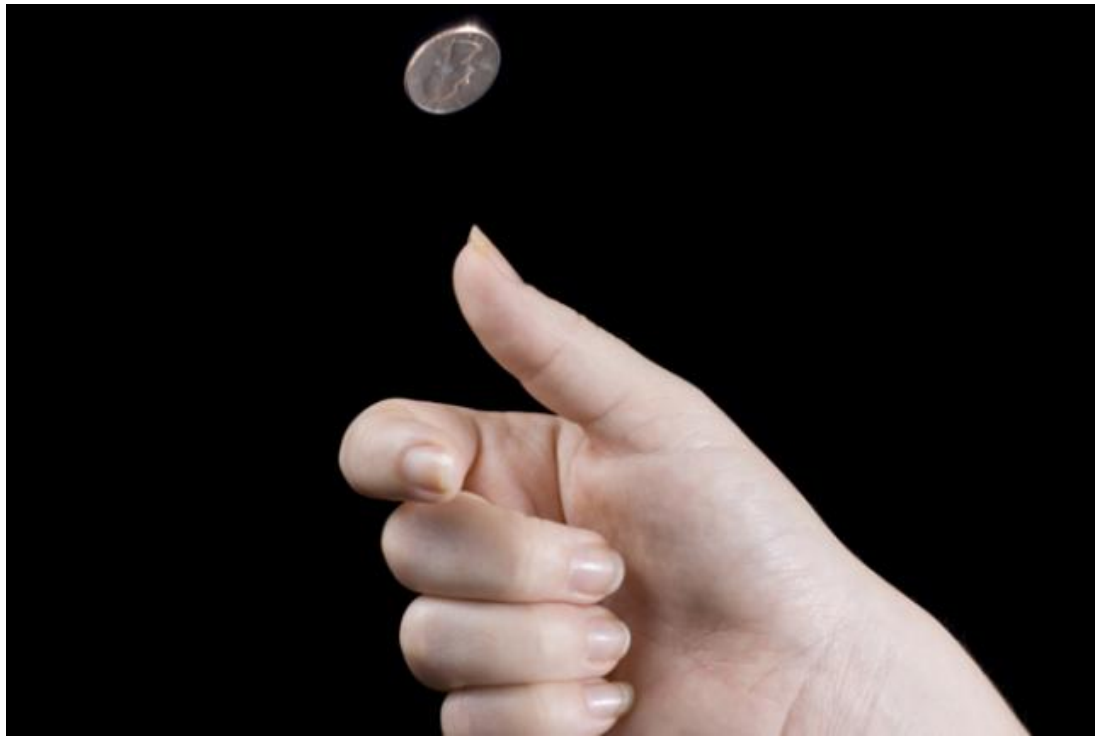


Image: Unknown; CC BY

Background - Observational studies

- How about in an observational study, what is the probability of exposure?
  - Is it random?

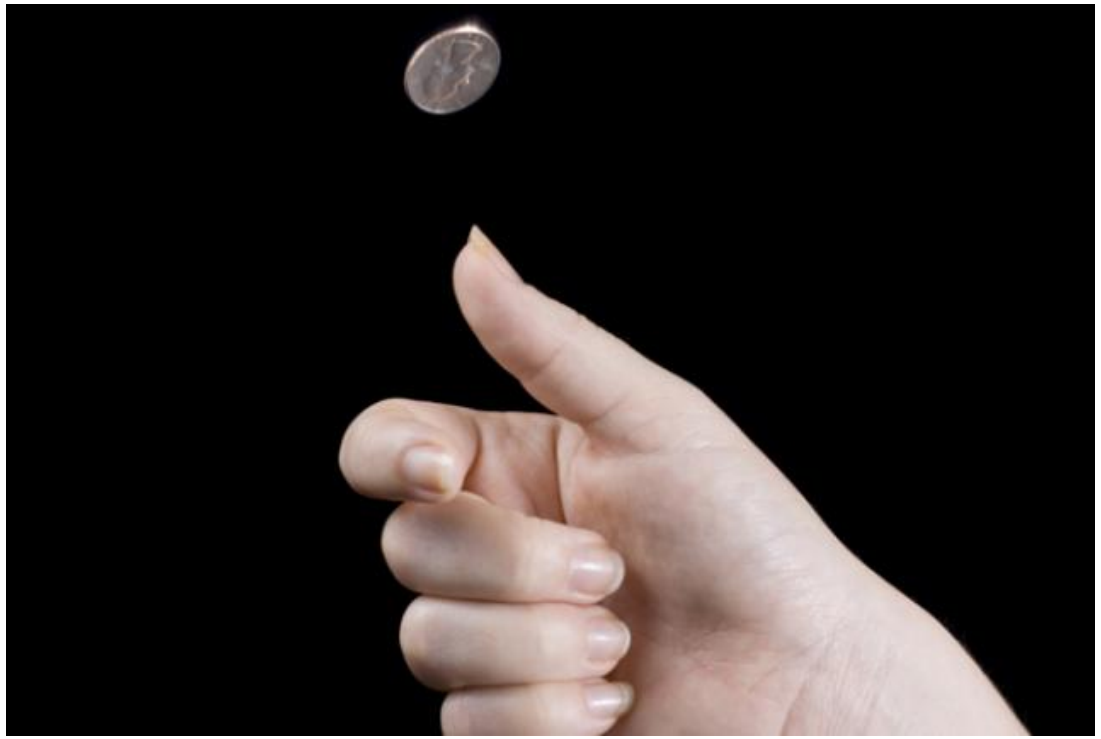
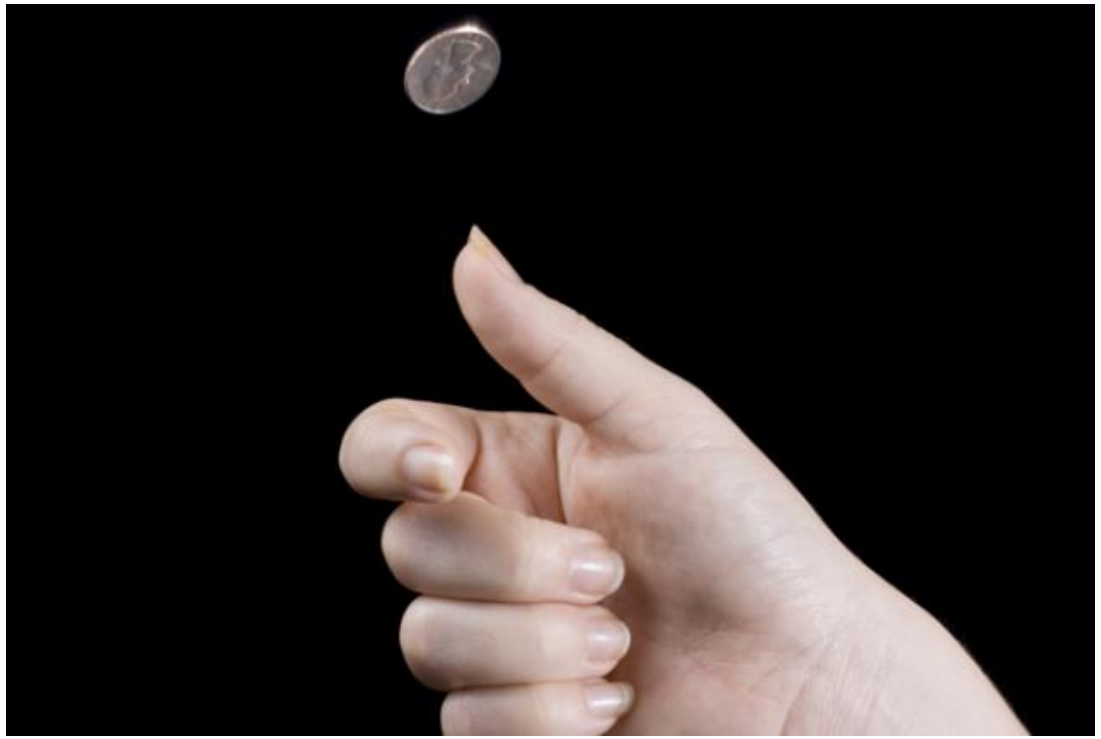


Image: Unknown; CC BY

Background - Observational studies

- In observational studies, treatment selection is often influenced by subject characteristics.
- As a result, baseline characteristics of treated subjects often differ systematically from those of untreated subjects.
- Therefore, one must account for systematic differences in baseline characteristics between treated and untreated subjects when estimating the effect of treatment on outcomes



Background - Observational studies

- For certain research questions it is not feasible or ethical to perform an RTC
  - E.g. effects of smoking
- Thus, we need observational studies and techniques to minimize treatment bias
  - Enter propensity scores

Propensity scores (PS)

- PS is a probability, to be exact it is the probability of treatment assignment conditional on observed baseline covariates
- The propensity score is a balancing score: conditional on the propensity score, the distribution of measured baseline covariates is similar between treated and untreated subjects.
- Thus, in a set of subjects all of whom have the same propensity score, the distribution of observed baseline covariates will be the same between the treated and untreated subjects

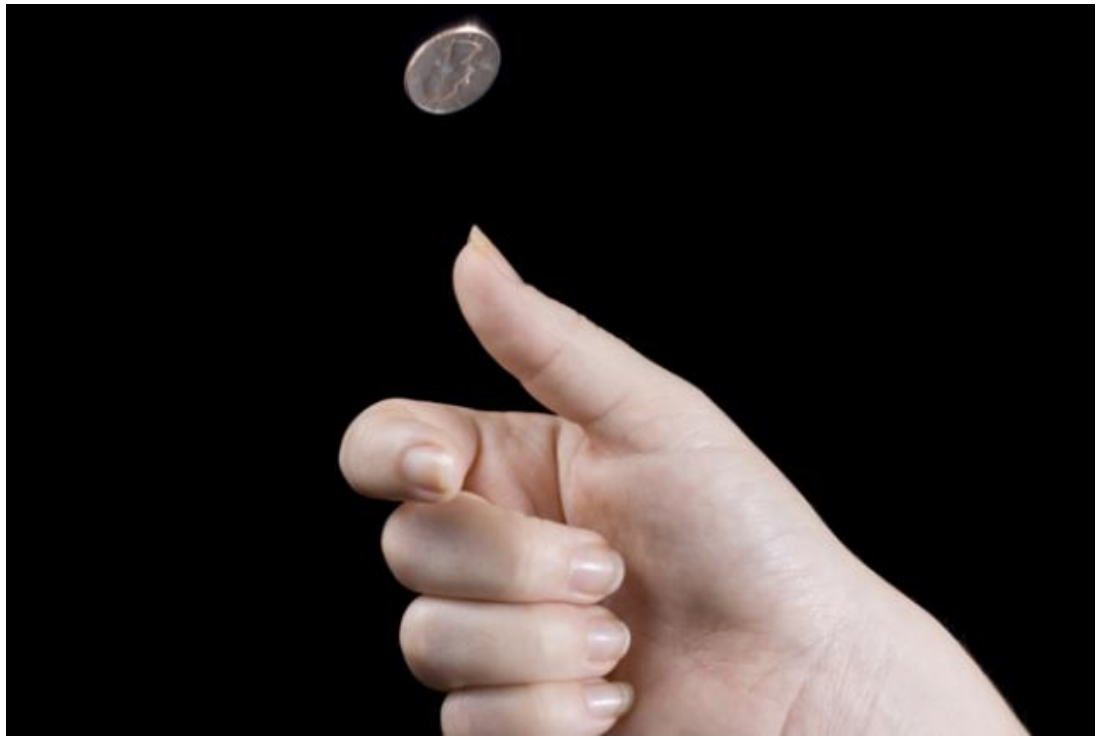


Image: Unknown; CC BY

Propensity scores (PS)

- Sounds familiar?
- the probability of being exposed is the same as the probability of being unexposed. The exposure is “random”
- We are mimicking the randomization from the RTC setting

Propensity scores (PS) – an example

- Bias in which patients receive surgery
- Get scores from logistic regression
- Matched patients who underwent surgery with those that were medical managed
- Regression analysis of the matched cohort revealed that surgery was associated with decreased mortality (HR 0.27; 95% CI, 0.13–0.55)

Early Surgery in Patients with Infective Endocarditis: A Propensity Score Analysis

FREE

Olcay Aksoy, Daniel J. Sexton, Andrew Wang, Paul A. Pappas, Wissam Kourany, Vivian Chu, Vance G. Fowler, Jr., Christopher W. Woods, John J. Engemann, G. Ralph Corey ... Show more

Clinical Infectious Diseases, Volume 44, Issue 3, 1 February 2007, Pages 364–372, https://doi.org/10.1086/510583

Published: 01 February 2007 **Article history** ▾

PDF Split View Cite Permissions Share ▾

Abstract

Background. An accurate assessment of the predictors of long-term mortality in patients with infective endocarditis is not possible using retrospective data because of inherent treatment biases and predictable imbalances in the distribution of prognostic factors. Largely because of these limitations, the role of surgery in long-term survival has not been adequately studied.

LUND UNIVERSITY

Clinical Studies Sweden Forum South

Skåne University Hospital

Generalized Propensity scores (GPS)

- The PS is limited to binary exposures/treatment
- Or exposures that we have defined in binary terms
  - E.g. we have smoking habits (current smoker/former smoker/never smoker) as categorical variable, but smoking (Yes/No) is used
  - Or we have BMI, but obese (BMI ≥30) is used
- When we categorize a variable, we lose information
- And for some research questions the level of exposure might be more interesting to look at

Generalized Propensity scores (GPS)

- That is where the GPS comes in, where we extend the PS framework to include more than just binary exposures
- The GPS is defined as the conditional probability of receiving a particular **level of exposure/treatment** given observed baseline covariates
- GPS is a balancing score: conditional on the propensity **function**, the distribution of measured baseline covariates is similar between treated and untreated subjects

Generalized Propensity scores (GPS)

- An example, say we have the case of three categorical treatments. How would I estimate the generalized propensity score?
  - We could use multinomial logistic regression
  - What model you use depends on the exposure variable

- What should I include in the GPS-model?
  - In practice, in many settings, most subject-level baseline covariates likely affect both treatment/exposure and the outcome. Therefore, in many settings, it is likely that one can safely include all measured baseline characteristics in the model

Generalized Propensity scores (GPS)

- How do I know if the GPS strategy worked?
  - key step is to examine how the covariates are balanced after conditioning on the propensity score is accomplished
  - Let's say we have matched subjects on the GPS
    - With matching, assessing whether the propensity score model has been adequately specified involves comparing treated and untreated subjects within the matched sample.
    - comparison of the means or medians of continuous covariates and the distribution of their categorical counterparts between treated and untreated subjects

Things to keep in mind

- GPS relies on the same theoretical framework as PS, so same strength and limitations

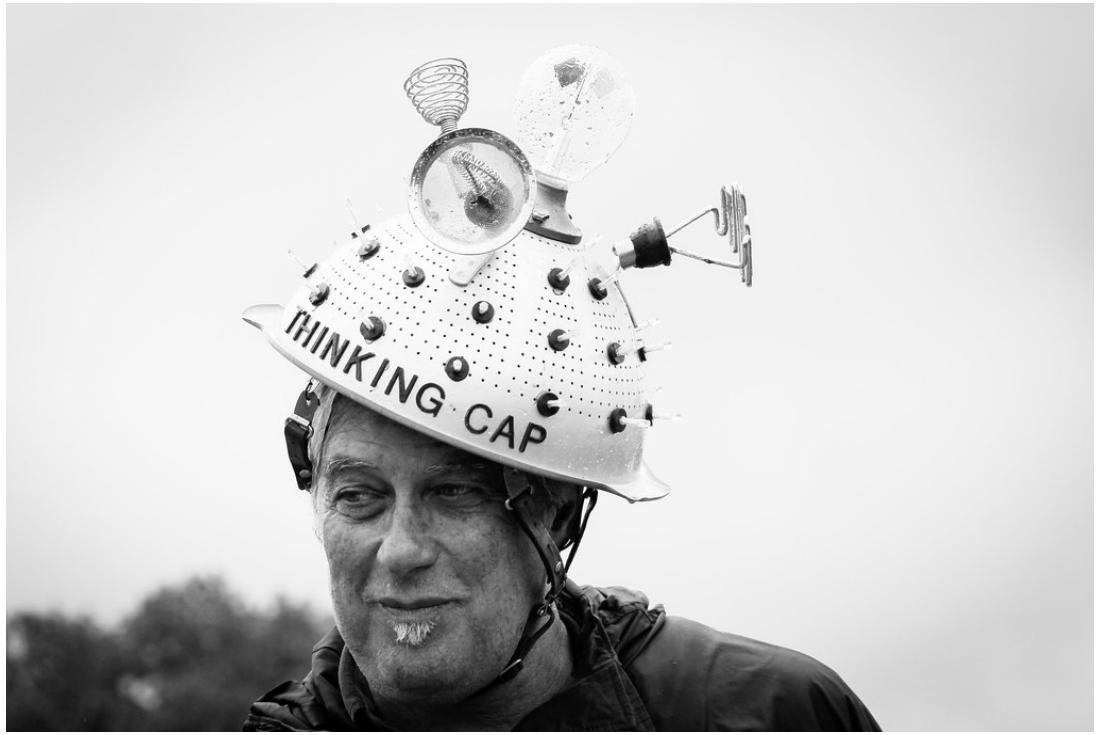
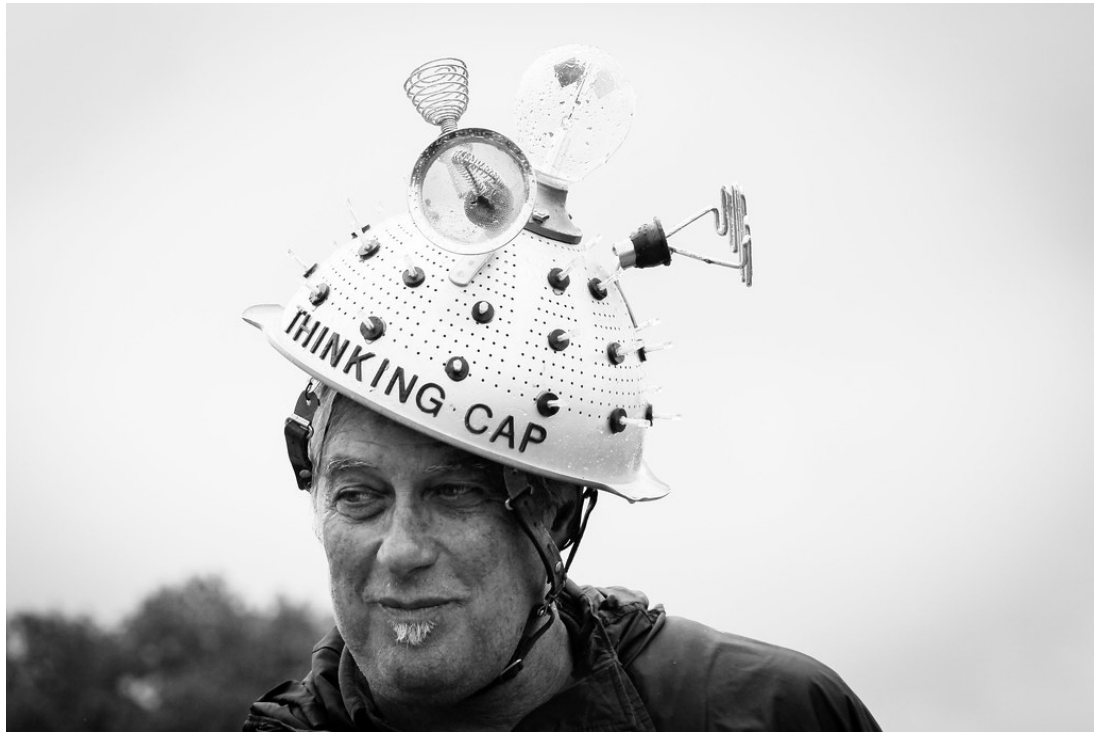


Image: Unknown; CC BY-NC-ND

Things to keep in mind

- Some strengths
  - uses one score(s) instead of multiple covariates in estimating the effect. This allows an investigator to use dozens of covariates, which is not usually possible in traditional multivariable models because of limited degrees of freedom and zero count cells arising from stratifications of multiple covariates
  - don’t use any information on the outcome when calculating the GPS, no analysis based on the GPS will bias effect estimation



Things to keep in mind

- Some limitations
  - only controls for measured covariates
  - If we have missing data, we get a missing score
  - works best in large samples to obtain a good balance of covariates

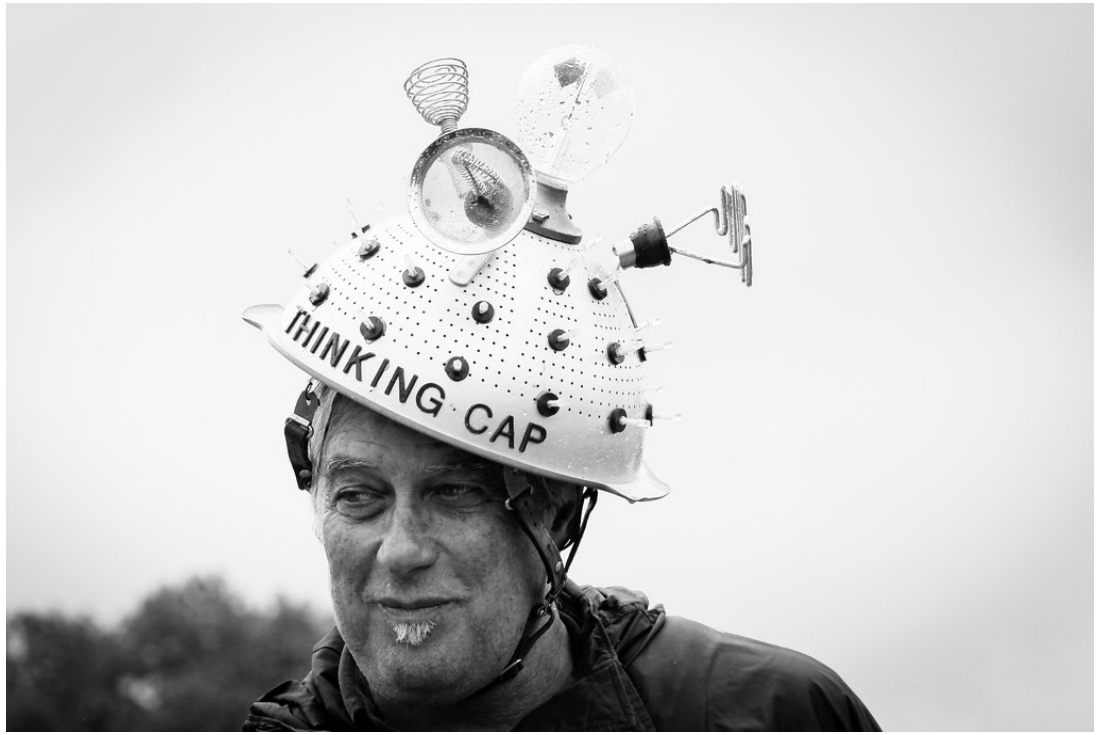
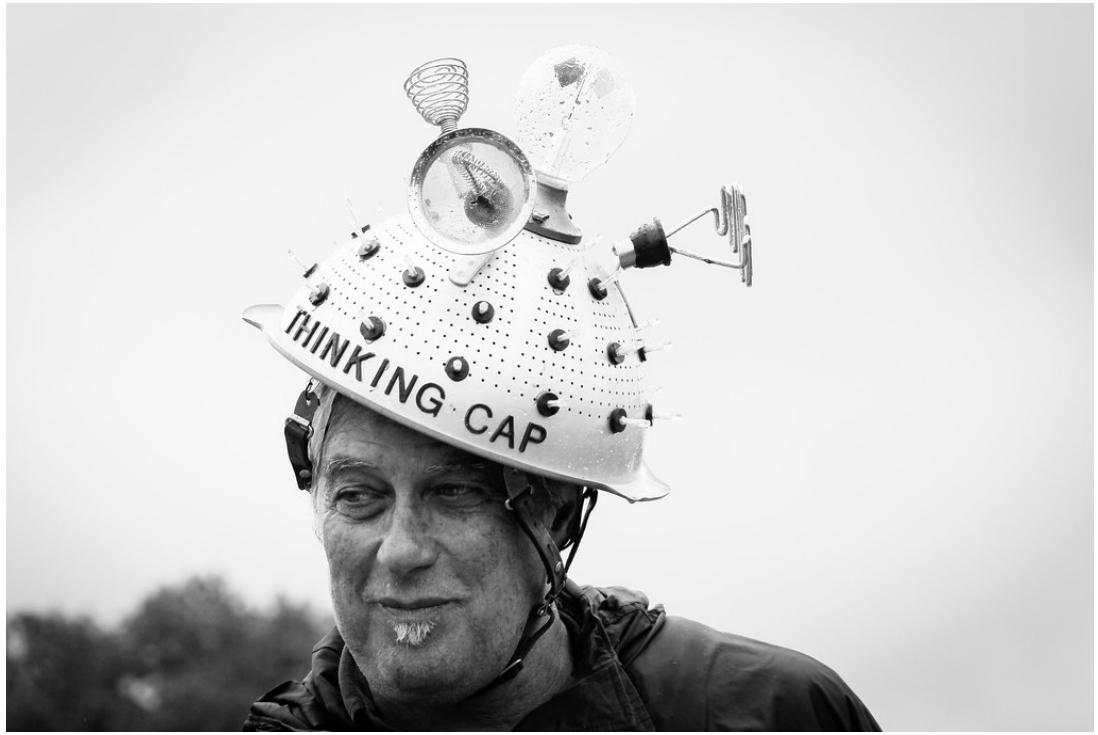


Image: Unknown; CC BY-NC-ND

Things to keep in mind

- When should you us these kinds of methods?
  - Bias in who is receiving treatment/exposure
  - Most importantly, when is suits the research question
  - Does not hurt to have quite a large data set



Thanks for listening!