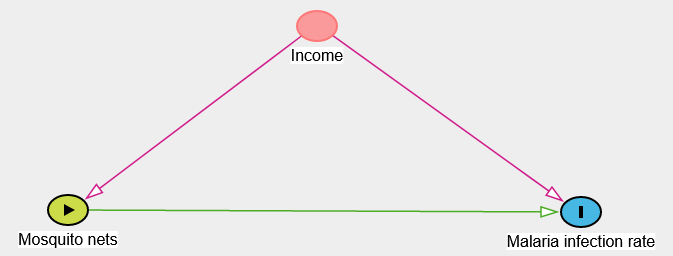
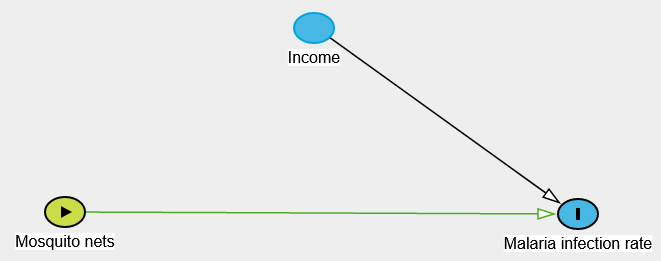
# 7: Randomisation and Matching

## The Magic of Randomisation

### Why randomise?

* If you don’t randomise and allow self-selection, selection bias can create different characteristic profiles in each group.

### RCTs and DAGs

* Consider the effect of possessing a mosquito net on malaria infection rates
  + Observational DAG could look something like this: 
  + Higher-income households use mosquito nets more because they can afford them and have a lower infection rate due to better access to healthcare etc.
  + A well-designed randomisation can eliminate the arrow from any confounders (Income) to the treatment node (Mosquito nets).
  + The only thing remaining is the effect of mosquito nets on the malaria infection rate i.e. the experimental DAG looks like this: 
  + Households can’t self-select into purchasing mosquito nets.
* Randomisation eliminates confounders so you don’t have to worry about front-doors, back-doors etc.

### How to randomise?

* 3 different steps.
  + Define the eligible population.
  + Create a representative sample.
  + Assign members of the representative sample to the treatment and control groups.

### Random assignment

* The issue with using a PC (Excel or a programming language) is that it’s never truly random and actually hackable i.e. you can predict what number will be generated.
* Atmospheric noise is used by states and governments to generate perfect randomness as it’s not hackable.

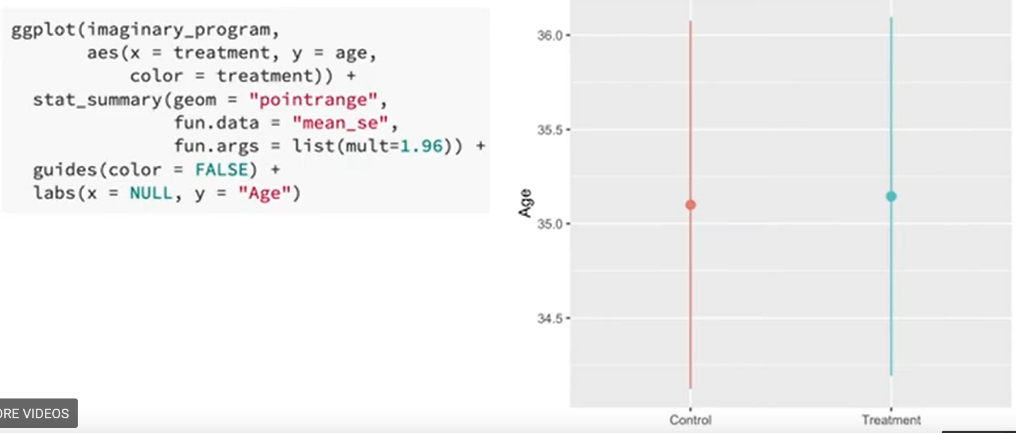
### How big a sample?

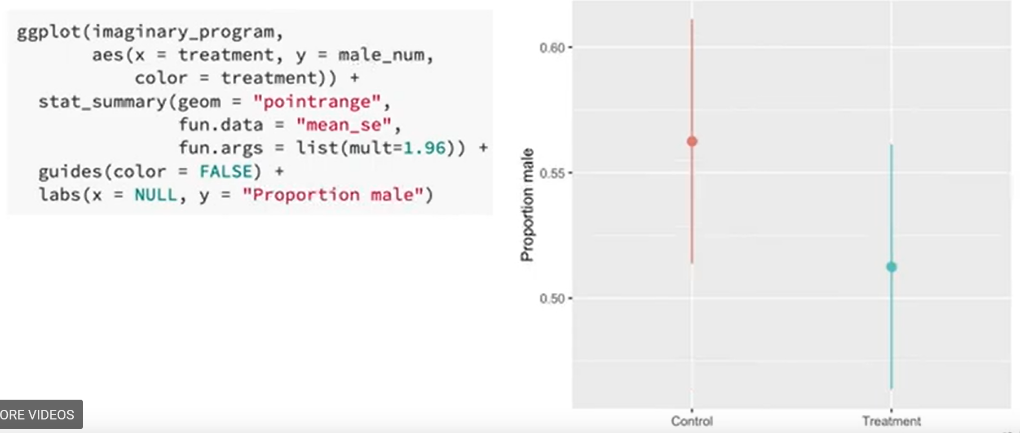
* Basically depends on how big an effect you want to be able to find.
* The size of the sample determines how easily you can detect an effect if there is none.

### Power

* In a small sample, the variation in possible outcomes is huge, so it’s much more unlikely the effect observed will be statistically insignificant.
* It doesn’t mean there’s no effect, just that the effect is undetectable.

## How to Analyse RCTs





* Age is well-balanced.
* Proportion of males is less-so, but the confidence intervals overlap so probably fine here too (could do a t-test to check).
* The two outcome value confidence intervals don’t overlap, so can be fairly confident that there is a statistically significant difference in income between the two groups.

### Should you control for stuff?

* Essentially, should other variables be included in the regression equation?
  + No! as there’s theoretically no confounding and would bias the results.

## The “Gold” Standard

### Types of research

* Any study that doesn’t come from an experiment is an observational study.
* Observational studies work fine for identifying causal effects as long as they’re done well e.g. include specific strategies for isolating the pathway between treatment and outcome.
* Can be a lot of self-selection in observational studies.
* Can be more than just impractical – in some cases impossible or even unethical.

### “Gold standard”

* Not actually true that if you conduct an RCT correctly, you’ll get valid causal inference e.g. issues with external validity.
* RCT’s help with correct inference as all the confounding influences are eliminated unlike observational studies.

### RCTs and validity

* Any trend effects will occur to both the treatment and control groups in a RCT.
* Often subjects will get assigned to a group and no longer want to participate as a result of that assignment e.g. medicine.
* Results will be biased if there’s a systematic reason for subjects dropping out.

### Addressing attrition

* Try to get a diverse segment of the target population.
* If subjects don’t care about the experiment, they’re more likely to drop out.
* Collecting as much pre-study information as possible will enable assessment of whether attrition was correlated with a specific sub-group.

### RCTs and validity

* ITT (intention to treat effect) – effect of being assigned treatment vs the effect of actually being treated.

### Other limitations

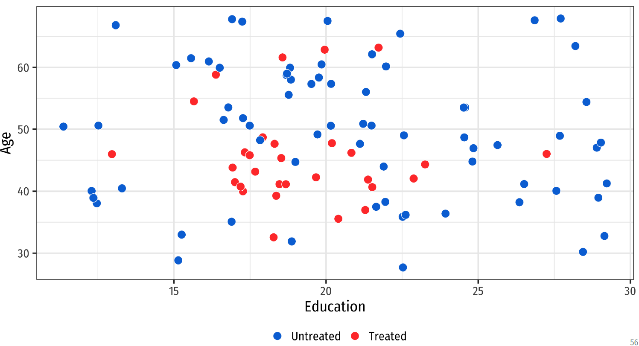
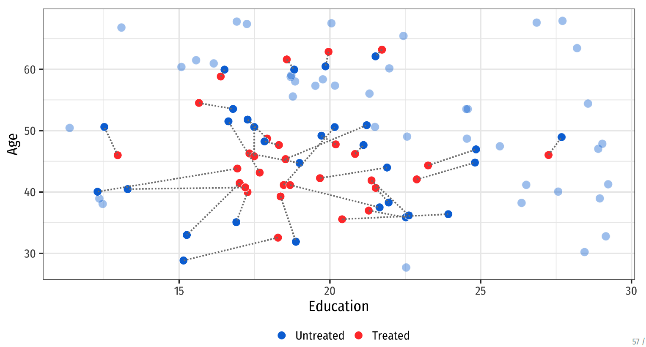
* Biggest criticism is that of external validity.
* RCTs great at finding small-scale, localised effects in specific situations but things start to fall apart when scaling up.

## Adjustment with Matching

### Why match?

* Can include lots of different transformations to make the model fit the data, which means there can be huge researcher influence on the final model structure and hence biased results.
* Removed all of the things that make the two groups different (except for treatment).
* The best general way to perform causal inference if you’re not doing any of the fancier techniques like RDD or DiD etc.
* There are techniques we can use to reduce dependence on model form for the size of the causal effect.
  + Only use subsets of the data – i.e. only regions with both treated and untreated observations.
  + Can remove the other regions with matching.
  + Essentially creating a synthetic control group.

### Nearest neighbour matching

* Consider a situation in which age and education are confounders.
  + Want to plot just the confounders and find subjects very similar in these characteristics (but choosing different treatments).
  + 
  + Want to find two dots of different colours very close to each other.
  + Use just those pairs to close the back-door in the DAG.
* Use a distance-metric to identify the similarity between pairs.
* 
  + Some odd matches, but this can be a result of the matching algorithm chosen e.g. here 1-1 matching without replacement was enforced.
* Now throw away all the unmatched observations.

### Potential problems with matching

* Matching may reduce statistical power too much when there’s small sample sizes.

### Propensity score matching

* Popular but not great at isolating the causal effect between treatment and outcome.

### Weighting

* Use scaling to make the sample more representative of the population.
* Occurs all the time in political polling.
* Most surveys will have weights that should be used to make the results more generalisable.

### Inverse probability weighting

* Really popular because you don’t have to throw away any non-matched data.
* IP weights are often trimmed to a maximum value (but no systematic rule-of-thumb).
* If there are any other confounders that might influence the relationship between treatment and outcome that haven’t been included in the propensity model, they’re included in the second regression stage.
* In the education and age example, these are included in the initial propensity score model.
* These weights are then used in a weighted regression model of treatment on outcome (no other covariates) to isolate the causal effect.
* Works better than plain propensity score models as it handles non-linearity.
* Just controlling for things assumes that there’s a linear relationship (for linear/logistic regression at the very least).
* IPW is a more robust approach to handling interactions between variables.