DATA TO INSIGHT: AN INTRODUCTION TO DATA ANALYSIS
THE UNIVERSITY OF AUCKLAND

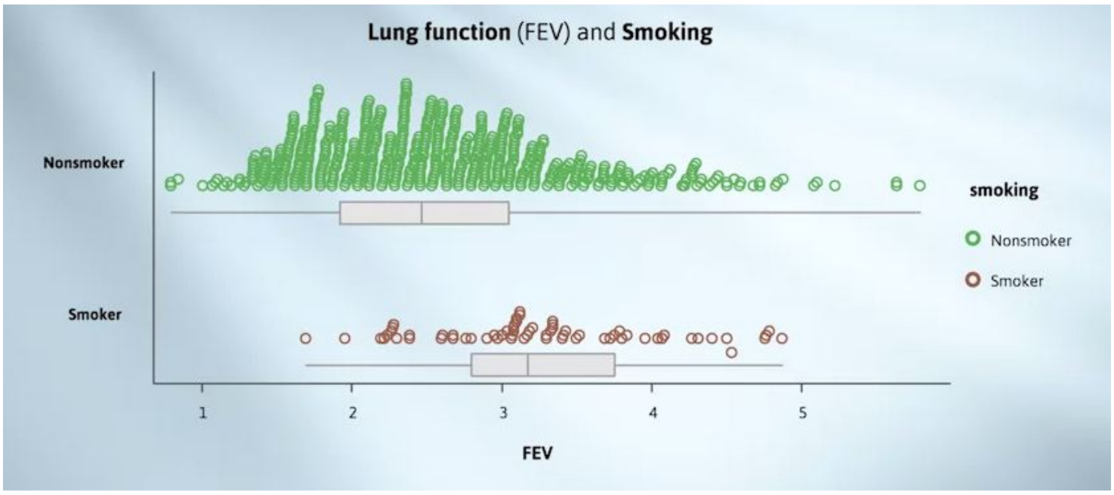
THE UNIVERSITY OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

WEEK 7
RANDOMISED EXPERIMENTS by Chris Wild

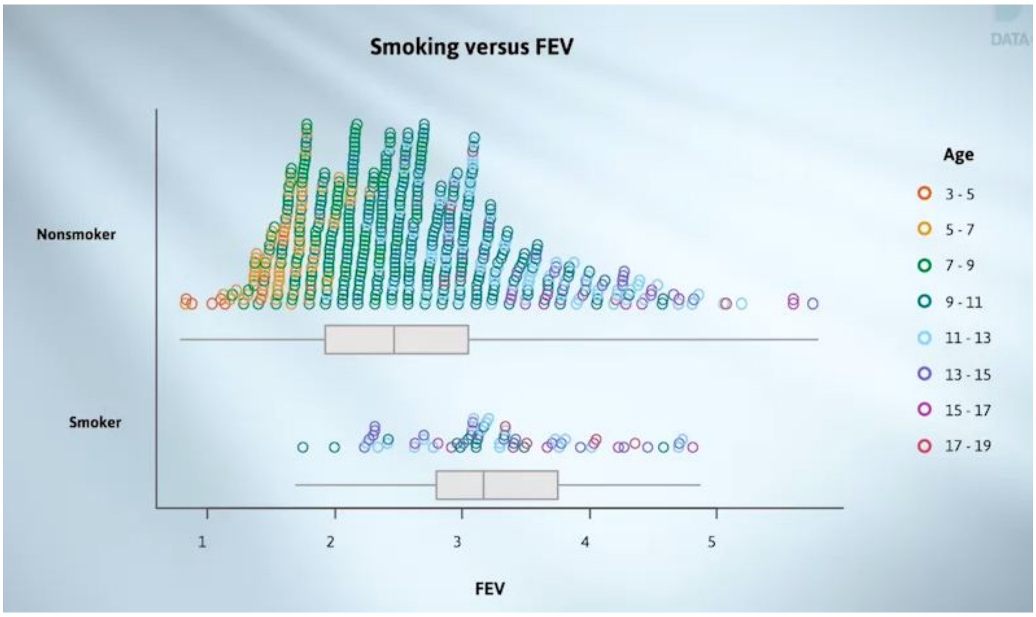
Welcome to the start of Week Seven.

This week we're going to talk about randomised experiments and how you deal with objections that say, "You've proved nothing. Your results could just be due to chance."

And when media reports say something like, taking statins significantly reduces the risk of heart attack, what does the word "significant" mean? It comes from the ways we deal with chance explanations. So we'll also pay attention to language and alert you to situations in which the words you read probably won't mean what you think they mean.



Let's go back to our example in Week Five where we compared the lung functions of smoking and non-smoking children.



It became clear this wasn't a fair comparison to tell us about the effect of smoking on lung function, because the non-smoking group contained large numbers of young children, while the smoking group was comprised mainly of teenagers. Young children have small lungs and can't blow much air.

The groups weren't just different on smoking. They were also different on other important factors, most noticeably age. We weren't just comparing the effect of smoking. We were comparing some mixture of the effect of smoking and the effect of age. Both effects were mixed up, or confounded.

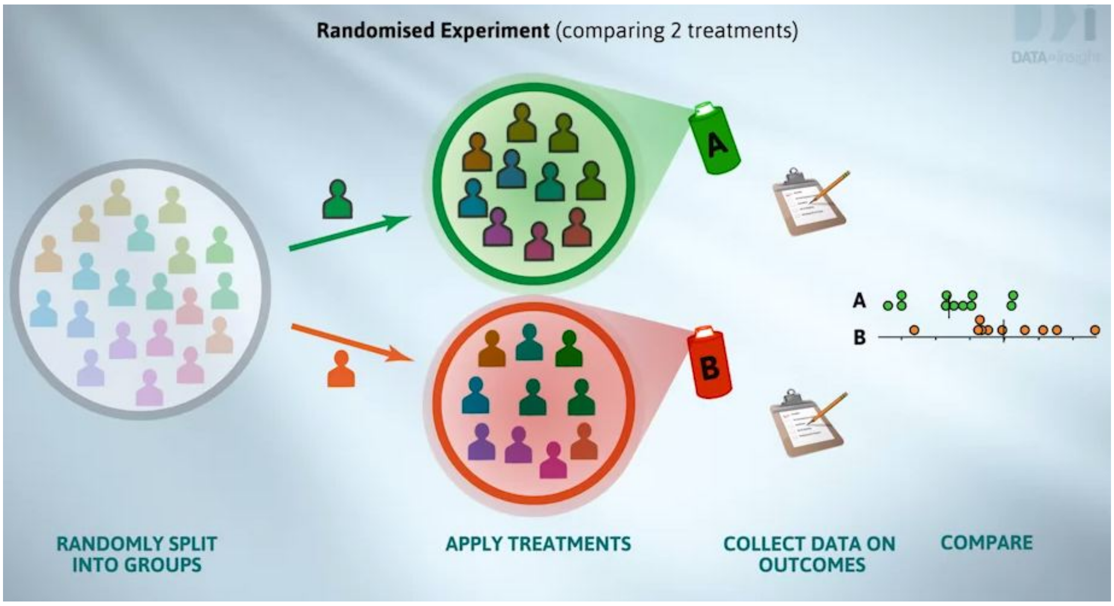
In Week Five we saw that a causal conclusion("This is what did it")-can never be justified on the basis of observational data alone. Any effect of the hazard is all mixed up with (or confounded by) the effects of any of the other factors that led to some people being exposed and others not.
The same considerations apply to assessing whether some medical treatments or some business strategies work better than others.

We can't make the obvious comparison and say, "Smoking caused that" because the true cause could have been any of the other ways in which the two groups were different.

Week Five gave some ideas about what we can do about the confounders that we know about and have data on.

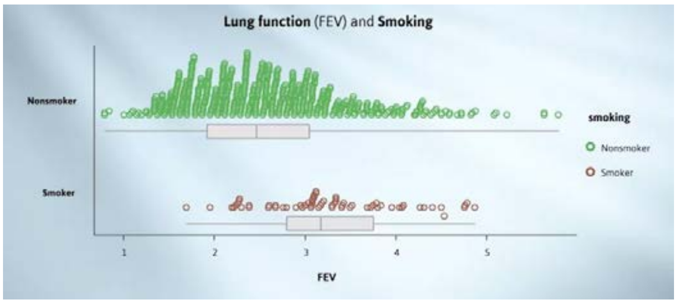
But what about the confounders we don't know about? How can we ensure that we're comparing like with like, that we're conducting a fair test? We can't just watch something that's unfolded in the world. We have to intervene, and we have to use a balancing strategy.

The gold standard of making fair comparisons (so that we can infer cause and effect) is the randomised experiment, which we'll now begin to describe. Following standard practise, we'll call the conditions we want to compare "**treatments**." Our treatments could be smoking and not smoking; new drug, standard drug, and no drug; or four different web page designs being trialled on consumers to see which one gets the best rate of clicks to advertisements.



To have a randomised experiment, we have to be able to say who will get what treatment. The people or entities that receive a particular treatment are called a treatment group.

Then we choose randomly who will be in each treatment group. Apply the treatments, observe the outcomes, and compare the groups. This random assignment tends to balance the groups on everything except the treatment they're receiving, so that when we compare treatment groups, we are comparing like with like so that we are conducting a fair test.

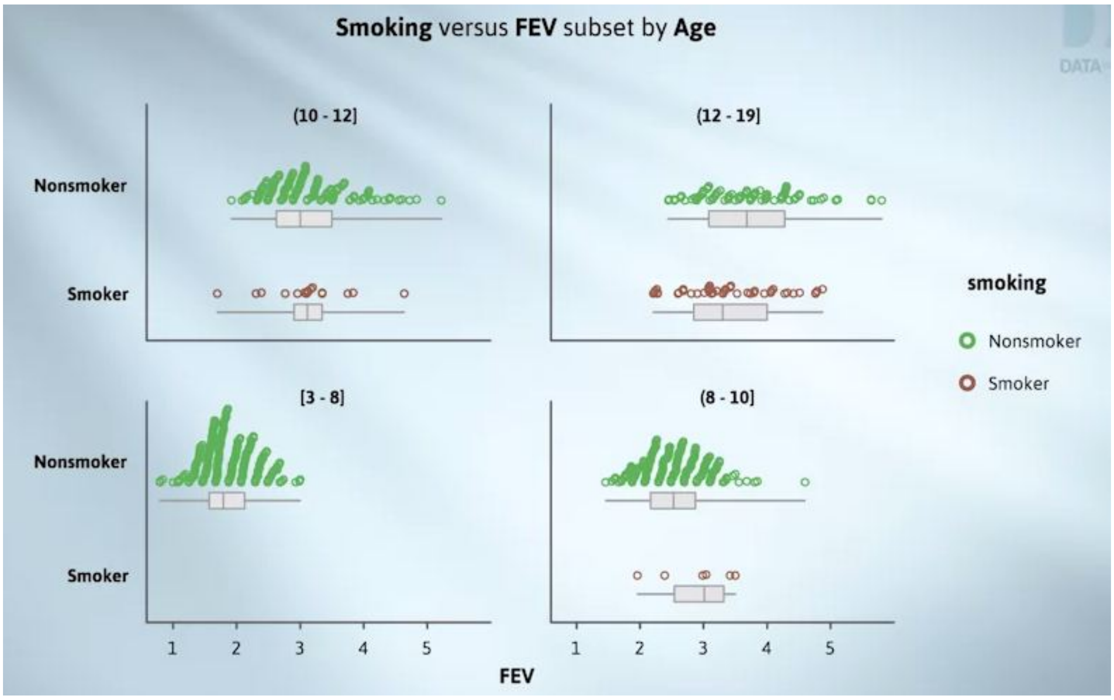


Our motivating example (lung function and smoking)

If we were able to do a randomised experiment for our motivating example, we would randomly choose which children would smoke and which children would not. This random assignment would tend to balance the smoking and non-smoking groups on age and everything else, so that we had a fair comparison of the effect of smoking.

In this case, however, we couldn't ethically conduct such an experiment because we can't purposefully expose people to known hazards. Setting the ethical argument aside momentarily, an experienced statistician wouldn't conduct an experiment on smoking and lung function here by simply randomising people to smoke or not smoke.

We know that lung function is strongly dependent on age. A better design would be first to divide subjects up into age groups. Statisticians would call this "blocking on age". Second, they'd randomise "smoke" or "don't smoke" within an age group (or block) and then make the comparisons between smokers and non-smokers within the same block (groups of children of similar age).



It's a bit like this graph, except that in each age group we would want to have similar numbers of smokers and non-smokers. The basic principle is that we block on factors we know are really important (like age here) and then randomise within blocks to take care of all other factors we haven't allowed for. But that's taking us beyond this course.

Random assignment to treatment groups is the most reliable way people know for balancing groups and making fair comparisons, but it's by no means perfect. In the next video, we'll show you shortcomings of randomisation. But those demonstrations of the problem also hold the seeds of the solution.