A/B Testing https://docs.google.com/presentation/d/1sdFDud2Wrv8PmmY2_XDknkg1tgvpheNMD_9Old9IA8c/edit?usp=sharing

Bias & Sampling https://docs.google.com/presentation/d/133bBtOQHPaBYSu7RBmbGuu8PN0UUxt278qTqOWSuJIQ/edit?usp=sharing

Adjustment / blocking https://docs.google.com/presentation/d/1xDY7Qh_ip4HA01unq8mxLF7M0ti8AGA0xmKeOihkqjg/edit?usp=sharing

One way to combat confounding is at the stage of design. If we were to randomize the ad campaigns across sites, then (at least with high probability) the audiences would be similar. Of course, we might get unlucky, and imbalances of the audiences may still occur, but the chance of that gets smaller as we randomize across more sites. This is the premise of **A/B testing**. In A/B testing one formally designs an experiment with randomization to make the groups being compared (A versus B) as comparable as possible. You can read more about A/B testing here (where they not so kindly call it jargon): https://en.wikipedia.org/wiki/A/B_testing

What can we do when we don't have randomization? Also, if we know and and collect a variable that clearly will be a confounder, shouldn't we incorporate that into our design rather than leave its balance across treatment groups up to chance? These questions are addressed by blocking and adjustment. In a **blocked experiment**, we randomize within levels of a potential confounding variable. **Adjustment** is a strategy that is used after the data has been collected. In adjustment, we look at the relationship between the predictor and outcome with levels of the confounding variable held fixed. So, if audience demographics confounds the relationship between our ad campaign and purchases, we look at the relationship within demographics. Regression models do this sort of adjustment for us automatically, with some assumptions. When putting a variable into a regression model, it can have all sorts of effects on the relationship of interest. We go over several examples in the lecture. Read more about blocking here https://en.wikipedia.org/wiki/Blocking_(statistics) and regression adjustment here http://oem.bmj.com/content/62/7/500.full

The final aspect of experimental design is **sampling**. Ideally one can use random sampling to obtain a sample that, with high probability, is a good representation of the population that you'd like to describe. Often, however, it's impossible to have control over the sampling process of an experiment. **Sampling bias** occurs when the sample is not indicative of the target population resulting in inferences that are off. We discuss three strategies to work around issues with the sample. First, is random sampling. The second is weighting, the process of allowing certain observations to carry more influence in models. The final is modeling. That is, trying to model the process that are biasing the sample. Read more about sampling bias here https://en.wikipedia.org/wiki/Sampling_bias