RESEARCH · RESEARCH DESIGN

Experimental Design: Randomisation, Control and Manipulation

The three elements that define a true experiment (manipulation, control and random assignment), why each is needed for a causal conclusion, and the practical steps that protect an experiment from bias.

ModQu research article · 1,150 words · 11 references (APA 7th edition)

An experiment is the most direct way to find out whether one thing causes another. Instead of observing the world as it is and looking for patterns, the researcher changes something deliberately and watches what happens. Three elements define a true experiment: **manipulation** of the factor being studied, a **control** condition for comparison, and **random assignment** of participants to conditions. Each solves a specific problem, and removing any one of them weakens the causal conclusion. This article explains what each element does, how they work together, and the practical safeguards that keep an experiment honest.

The problem experiments solve

Suppose a retailer notices that customers who use its mobile app spend more than those who do not. But app users may differ from non-users in many ways: they may be younger, wealthier or more loyal to begin with. Any of these differences could explain the gap in spending. This is the problem of **confounding**: a third factor related to both the supposed cause and the outcome.

Rubin (1974) and Holland (1986) framed the problem precisely. The causal effect of the app on a customer is the difference between what that customer spends with the app and what the same customer would have spent without it. Only one of these outcomes can ever be observed for any individual, because a customer either uses the app or does not. Holland called this the fundamental problem of causal inference. Experiments solve it not for individuals but for groups, by creating groups that are alike in every respect except the factor being studied.

Manipulation

In an experiment, the researcher decides who receives the treatment. Manipulation establishes the order of events, because the cause is introduced before the outcome is measured, and it ensures that the treatment is not chosen by participants for reasons that might also affect the outcome.

A manipulation must be strong enough to matter and specific enough to interpret. If a test of a new product description changes the wording, the layout and the photograph at once, a difference in sales cannot be attributed to any one of them. Campbell and Stanley (1963) treated this control as the defining feature of an experiment. Researchers often add a **manipulation check**, a measure confirming that the treatment was noticed or understood as intended. Hauser et al. (2018) note that such checks should be placed so that they do not themselves alter the outcome.

Control

A control condition provides the comparison against which the treatment is judged. Without it, any change in the outcome could be due to time, seasonality, a news event or simply being studied. The control group experiences everything the treatment group experiences except the treatment.

Choosing the right control matters. A **no-treatment control** shows whether the treatment beats doing nothing. A **placebo or attention control** gives participants something that looks similar but lacks the active ingredient, separating the treatment's effect from the effect of receiving attention. An **active control** compares the new treatment with the current standard, which is often the comparison a decision-maker actually needs.

Control also means holding other conditions constant. Orne (1962) showed that participants who know they are in a study often try to work out its purpose and behave accordingly. These **demand characteristics** can be reduced by keeping both groups equally informed, and by **blinding**, so that participants, and where possible the researchers who measure outcomes, do not know who received which treatment.

Random assignment

Random assignment makes the comparison fair. Fisher (1935) argued that allocating units to treatments by chance is the only way to guarantee that no systematic factor, known or unknown, determines which group a unit enters. With random assignment, the groups differ before treatment only by chance, and the laws of probability describe how large such chance differences are likely to be. This is what allows a statistical test to say whether an observed difference is larger than chance alone would produce.

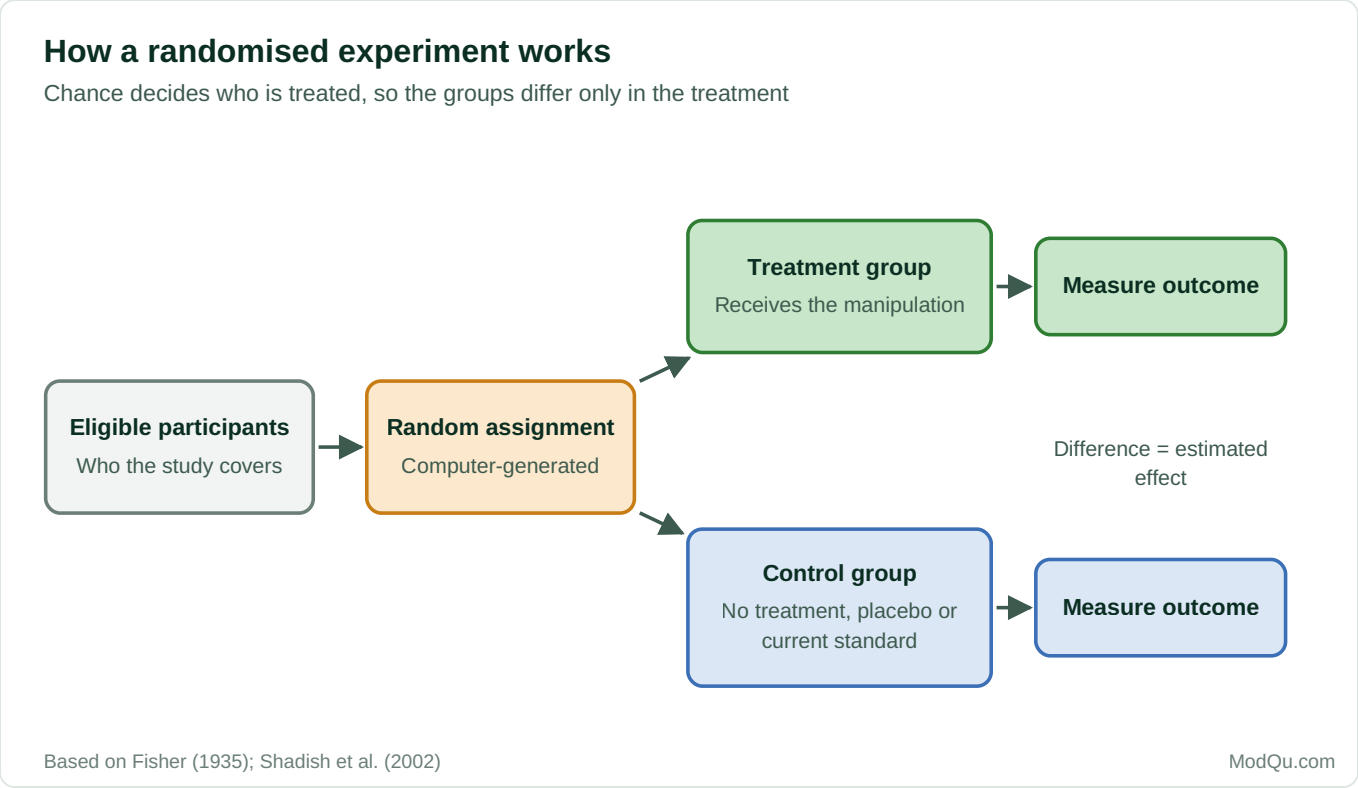


Figure 1. The logic of a randomised experiment. Random assignment makes the groups alike before treatment, so the difference in outcomes estimates the effect of the manipulation. Based on Fisher (1935) and Shadish et al. (2002); diagram by ModQu.

Random assignment should not be confused with random sampling. Random sampling determines who is in the study and supports generalisation to a population. Random assignment determines which condition each participant receives and supports causal conclusions within the study (Shadish et al., 2002). An experiment on volunteers can have strong internal validity and still apply only to people like those volunteers.

Randomisation must be genuine: assigning early visitors to one version and later visitors to another, or letting staff choose who gets a new service, is not random. Schulz et al. (2010) list concealment of the allocation sequence among the essential features to report in a trial, because when the people enrolling participants can predict or influence assignment, bias creeps back in. Assignment should be generated by computer and applied automatically.

When the sample is small, chance imbalances can still be large. Two refinements help. **Blocking** or **stratified randomisation** groups similar units together, for example by region or customer size, and randomises within each group. **Covariate adjustment** in the analysis accounts for baseline differences that remain (Imbens & Rubin, 2015).

An illustrative example

A software company wants to know whether a shorter sign-up form increases completed registrations. Each new visitor is randomly assigned to see either the current eight-field form (control) or a new four-field form (treatment). Suppose 4,000 visitors see each version, and 9.0% complete the current form compared with 10.5% for the shorter one

(illustrative figures). Because assignment was random, the groups should be alike in every other respect, so the difference of 1.5 percentage points can be attributed to the form, subject to a test of whether a difference that large could easily arise by chance. Kohavi et al. (2020) describe the safeguards needed to run such online experiments reliably.

Limits of experiments

Experiments show the effect of a specific treatment, in a specific setting, over a specific period. Gerber and Green (2012) emphasise that effects found in artificial conditions may not hold in natural ones, which is why field experiments are valuable. Some causes, such as age, gender or a recession, cannot be manipulated at all, and some manipulations would be unethical. In those cases, researchers turn to quasi-experimental and observational designs, which try to approximate the logic of an experiment without random assignment.

Conclusion

Manipulation ensures that the cause comes first and is not chosen by participants. Control provides a comparison that captures everything except the treatment. Random assignment makes the groups comparable in ways the researcher cannot see or measure. Together, they allow a difference in outcomes to be read as the effect of the treatment. When any element is missing or compromised, the conclusion becomes an association that needs further evidence before it can guide a decision.

References

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.

Fisher, R. A. (1935). *The design of experiments*. Oliver and Boyd.

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. W. W. Norton.

Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. (2018). Are manipulation checks necessary? *Frontiers in Psychology, 9*, Article 998. https://doi.org/10.3389/fpsyg.2018.00998

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945–960. https://doi.org/10.1080/01621459.1986.10478354

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy online controlled experiments: A practical guide to A/B testing*. Cambridge University Press. https://doi.org/10.1017/9781108653985

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*(11), 776–783. https://doi.org/10.1037/h0043424

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688–701. https://doi.org/10.1037/h0037350

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ, 340*, Article c332. https://doi.org/10.1136/bmj.c332

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.