

How to test

whether your behaviour change intervention works









When you design or change a policy, program or communication, testing is essential to understand whether it works.

Behavioural Insights tells us that even small details can influence participation in a service, completion of a task, and compliance with rules. Testing is important because many evidence-based interventions (over half of pre-registered studies) don't produce the same outcomes when translated into new contexts.¹ There are a few different ways to test, depending on your project and what you want to achieve.

Before you start:

1. Develop a hypothesis about how your intervention might have an effect on outcomes, and why. Write it out: 'If we change __, then the outcome will be __, because our background research suggests that __'.
2. Check that your success metrics are measuring the outcome you care about, and that you can access the data. See p.6 of BETA's guide for tips <http://bit.ly/BETA-Guide>
3. Choose the right testing method for your intervention. Use the table below to choose between a randomised controlled trial (RCT), an A/B test, or a before and after study. For other evaluation questions, refer to <http://bit.ly/NSW-evaluation>

Choosing a testing method

Ask yourself			
What is your time frame?	 Days or weeks	 Months	Choose the icons that best describe your project to see which method meets your needs. Also consider budget, ethics and data privacy when choosing.
How many participants do you have access to?	 1-1,000+	 1,000+	
How confident do you want to be that you have changed behaviour?	 Indication of a trend	 Certain as possible	
Can you compare more than one version at the same time?	 No	 Yes	

Experiments with randomisation (such as an RCT or an A/B test) are the gold standard method of testing. Randomisation allocates people into different groups using a 'lottery'. If the groups are large enough, the laws of statistics say that the average characteristics of each group will be roughly matched, e.g. education, income, personality, etc. You can then give each group a different intervention at the same time and be confident that any differences in their behaviour were caused by the intervention they received, rather than their characteristics or time factors.

¹ Warren, M. (2018). First analysis of 'pre-registered' studies shows sharp rise in null findings. *Nature*.

Randomised controlled trial (RCT)

RCTs involve *randomly* assigning people to different groups – typically a treatment group (which receives the intervention) and a control group (for whom nothing changes). The effect of the intervention is calculated by comparing the average behaviour of these groups at the end of the testing period.

A more detailed guide is here: <http://bit.ly/nesta-rct-guide>



How to run your RCT:

1. **Choose a success metric** for your intervention. This could be participation rate in a service, compliance rate with a regulation, or improvement in a particular social or environmental outcome that is linked to a behaviour.
2. **Define your comparison.** Consider testing your new intervention against business as usual (e.g. an existing service) or compare two completely different interventions (e.g. subsidy vs training) to determine which is more effective. What you test partly depends on the feasibility of treating customers differently – which is easier to do online (e.g. different email messages) than for in-person services (e.g. different case manager procedures).
3. **Identify your participants and check you can collect data** on your success metric for each participant in your RCT. Aim for at least 1,000 to 5,000 participants (individuals) per intervention (and the same number in the control group). A power calculator (<http://bit.ly/power-calc>) can help you calculate how many people you need in your trial for statistical significance, but the more people, the better. If your RCT is high-stakes, check with a statistician.
4. **Randomise your participants** to create two or more equal sized groups. The number of groups you need depends on the number of comparisons you want to test, e.g. you might have one control group (business as usual) and two different treatment groups. You can randomise either individuals (e.g. students) or clusters (e.g. schools) – see module 4.4 of this guide <http://bit.ly/rand-in-excel> – but individuals will give you more data points and therefore more confidence in your results.
5. **Deliver your intervention** over a set period of time, to the treatment group only. If you have multiple treatment groups, each will receive a different intervention at the same time.
6. **Calculate average success metrics for each group**, e.g. X% success in the control group and X% in the treatment group. Check whether the difference is statistically significant, ideally by conducting a regression analysis – see module 8.1 of this guide <http://bit.ly/rand-in-excel> – or by using this tool <http://bit.ly/analyse-tool> for a rough estimate. If the intervention was successful, scale it to everyone.

A/B testing

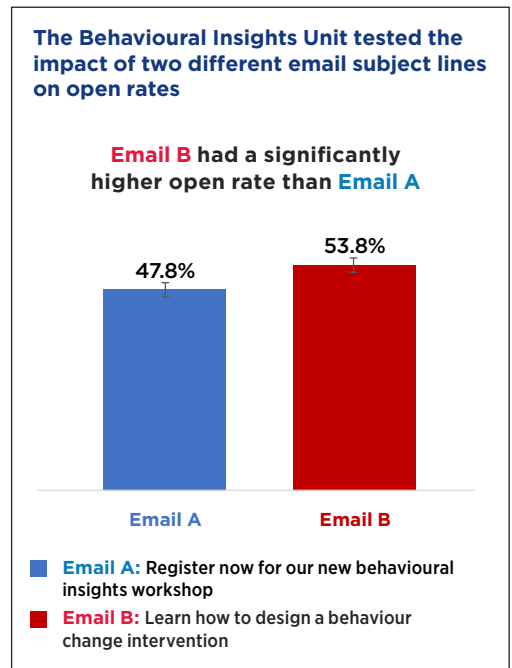
An A/B test is the same as an RCT but in an online environment. A/B testing (or split testing) can often be automated because most website, app, email and survey platforms have built-in randomisation and analysis tools or allow you to plug-in to third party tools.

Learn more about A/B tests here: <http://bit.ly/AB-Test-Guide>



How to run your A/B test:

1. **Choose a success metric** for your content, product or service, such as sign-up rate or completion rate.
2. **Create another version of your content, product or service** and change the elements you think will increase your success, such as text, layout and ease-of-use.
3. **Randomly send half your users version A and the other half version B.** Use the A/B testing function of your distribution platform, or this tool <http://bit.ly/randomise-tool>. Randomisation ensures you have two comparable groups of users.
4. **Use the analytics dashboard** of your A/B testing software or this tool <http://bit.ly/analyse-tool>, to compare the success metrics of version A and B, and determine whether the difference is statistically significant. If one is significantly more successful, implement it as business as usual – otherwise, keep testing!



Before and after study (pre-post)

Before and after studies can help you collect suggestive evidence about the impact of an intervention. Unlike an RCT, you won't know if your success metric would have changed over time anyway (without the intervention) but these studies are still useful when randomisation is not possible.

To increase confidence in your results, try to create a comparison group using quasi-experimental methods. Refer to BOCSAR's guide here: <http://bit.ly/BOCSAR>



How to run your before and after study:

1. **Choose a success metric** and check you can establish a baseline (before) and endline (after) over a relevant period of time, using the same source of data. Ensure your comparisons (e.g. spring 2018 vs. spring 2019) minimise the chance of time-related events (e.g. seasonal variation) affecting your results.
2. **Design the best possible version of your product, service or program** because during a before and after study you can usually only test one version at a time.
3. **Identify where you will run the study**, i.e. which service or location. To reduce the influence of external factors on your study results, avoid sites undergoing any other process improvements.
4. **Compare any changes in your success metric** before and after implementing your intervention. If you achieved meaningful improvement, scale up the new approach – otherwise, adapt your design and re-test!