

# A methodology for conducting RCTs on the strategic road network

An approach for using behavioural economics



**LE**  
**London  
Economics**

May 2017

## About London Economics

London Economics is one of Europe's leading specialist economics and policy consultancies. Based in London and with offices and associate offices in five other European capitals, we advise an international client base throughout Europe and beyond on economic and financial analysis, litigation support, policy development and evaluation, business strategy, and regulatory and competition policy.

Our consultants are highly-qualified economists who apply a wide range of analytical tools to tackle complex problems across the business and policy spheres. Our approach combines the use of economic theory and sophisticated quantitative methods, including the latest insights from behavioural economics, with practical know-how ranging from commonly used market research tools to advanced experimental methods at the frontier of applied social science.

We are committed to providing customer service to world-class standards and take pride in our clients' success. For more information, please visit [www.londoneconomics.co.uk](http://www.londoneconomics.co.uk).

**Head Office:** Somerset House, New Wing, Strand, London, WC2R 1LA, United Kingdom.

w: [londoneconomics.co.uk](http://londoneconomics.co.uk)    e: [info@londoneconomics.co.uk](mailto:info@londoneconomics.co.uk)    : @LondonEconomics  
t: +44 (0)20 3701 7700    f: +44 (0)20 3701 7701

## Acknowledgements

We would like to thank the Department for Transport, the Department for Communities and Local Government, the Department for Environment Food & Rural Affairs, and Highways England for the valuable input and information they have provided for this study. Responsibility for the contents of this report remains with London Economics.

## Authors

**Dr Annette Harms**

**James Suter**

**Alessandro Castagnetti**



Wherever possible London Economics uses paper sourced from sustainably managed forests using production processes that meet the EU Ecolabel requirements.

Copyright © 2017 London Economics. Except for the quotation of short passages for the purposes of criticism or review, no part of this document may be reproduced without permission.

London Economics Ltd is a Limited Company registered in England and Wales with registered number 04083204 and registered offices at Somerset House, New Wing, Strand, London WC2R 1LA. London Economics Ltd's registration number for Value Added Tax in the United Kingdom is GB769529863.

---

# Table of Contents

Page

1	Introduction	1
2	Step-by-step guide to designing a randomised controlled trial	2
2.1	Sampling characteristics	3
2.2	Defining treatment and control groups	5
2.3	Success criteria	8
2.4	Roll-out	9
2.5	Analysis	10
2.6	Piloting	11
3	Conclusions	12
	ANNEXES	13
	Annex 1 References	14
	Annex 2 Statistical analysis of treatment effects	15
	Index of Tables, Figures and Boxes	17



# 1 Introduction

This report presents a step-by-step guide for conducting randomised controlled trials (RCTs) for testing anti-littering interventions on the Strategic Road Network (SRN). Such interventions could comprise

- Messaging campaigns (e.g. on the SRN, at service areas, on packaging of products sold at service areas);
- Investments in ‘binrastructure’<sup>1</sup>; and
- Other types of education, or behavioural interventions.

RCTs could be implemented in order to provide robust evidence on the effectiveness of any such interventions to tackle roadside litter.

Though RCTs are not the only methodology available for measuring success of anti-littering interventions, they are one of the most robust approaches. In general, RCTs involve splitting a sample population into treatment and control groups. Only the treatment group would be subjected to the intervention. Success criteria would then be observed and measured for each group. Any measured differences describe the impact of the intervention.

When evaluating anti-littering interventions, we are interested in assessing the true effect of the intervention without obscuring it with any other factors else that could have affected littering. To do this, it is useful to establish a control group (also called a baseline) to which we compare the effectiveness of the intervention. The careful design of a control group through randomisation is what distinguishes RCTs from other types of evaluations.

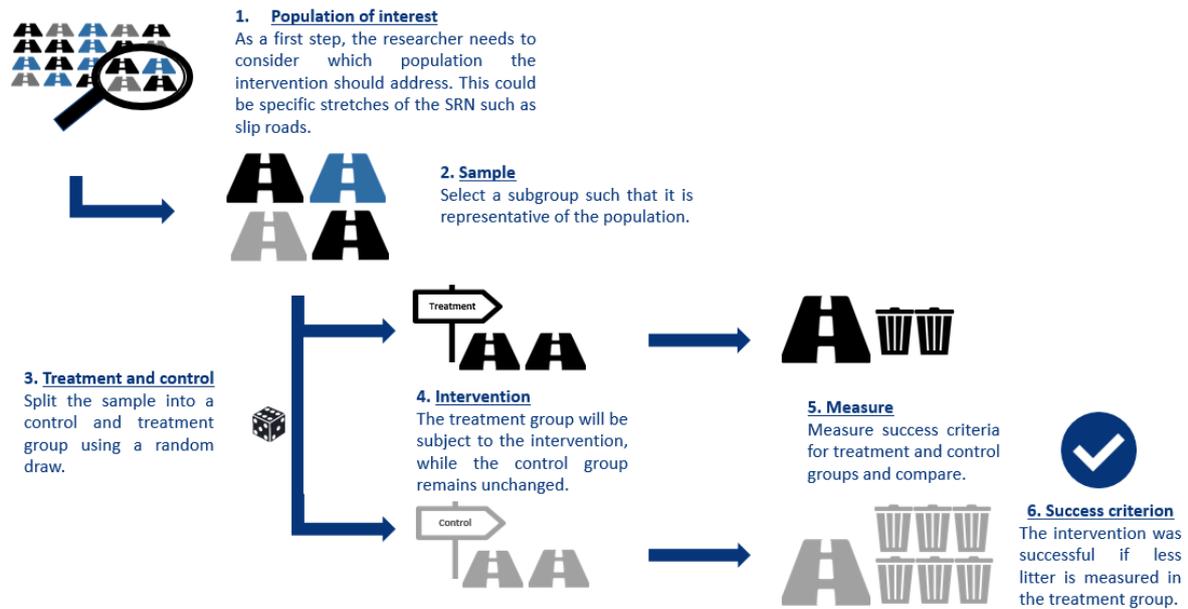
RCTs can be relatively simple, quick, and thus cost-effective to set up. Nonetheless, an experienced researcher who is familiar with the limitations and risks of the methodology should accompany the design and implementation.

Figure 1 shows an overview of the key steps of an RCT. The following sections give detailed step-by-step instructions.

---

<sup>1</sup> ‘Binrastructure’ refers to “*the design, number and location of public litter bins and other items of street furniture*”, see HM Government (2017) ‘Litter Strategy for England’.

Figure 1 The key steps of an RCT in brief



Source: London Economics

## 2 Step-by-step guide to designing a randomised controlled trial

The following steps assume that an intervention has been designed and is ready for being implemented for testing. While every step will depend on the specific context of the intervention, this guide illustrates the most important guiding principles that should be considered for running RCTs on the SRN. Case studies throughout the chapter provide illustrations of specific examples.

### Box 1 Possible interventions to test using an RCT

RCTs are useful for testing various types of anti-littering interventions. The following interventions are based on a report by Kolodko et al. (2016) written for Clean Up Britain and could all be tested using an RCT.

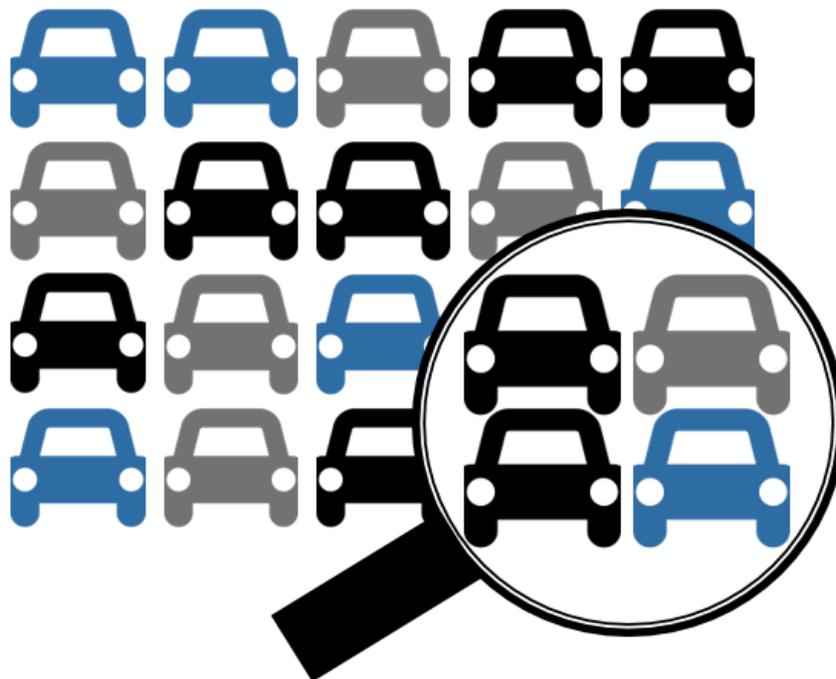
- **Binrastructure:** Since part of the litter that ends up on the SRN comes from Motorway Service Areas (MSAs), the interventions involving litter bins could vary the availability, accessibility and visibility of bins at service areas. For instance, the intervention design could include installing litter bins where there is a lack of them and/or increasing their accessibility and visibility by altering their colours and designs.
- **Messaging:** Messaging interventions could be designed to inform users on the detrimental consequences of littering, or could use behavioural insights to nudge people not to litter.
  - **Messages on the consequences of litter on the SRN:** If road users understand why littering is undesirable (e.g. unsightly, presents a safety hazard), it may motivate them to change their behaviour. Hence, informing people on the detrimental consequences of littering on health and the environment might prove to be successful.

- **Behavioural messages:** Littering often happens unconsciously without putting much thought into the littering action. As a result simple anti-littering posters such as the “We’re Watching You Campaign” developed by Keep Britain Tidy may nudge people to refrain from littering.
- **Social norms:** Many RCTs have shown that people adhere social norms and feel pressure when deviating from them. Similar messages to the one used by Keep Britain Tidy saying “*It’s no secret what people think of you – Bin your rubbish or drive it home*” could be tested using an RCT.
- **Fines:** Direct interventions such as imposing fines on those who litter and to effectively communicate these littering-related fines to road users could reduce littering behaviour.

## 2.1 Sampling characteristics

As a first step, the researcher needs to consider which ‘*population*’ the intervention should address. This is required to select a sample for the testing. The sample, which is a subgroup of the population, should be representative of the population. This ensures that the RCT results can be extrapolated to the population as a whole. For example, if the intervention may (after testing) be rolled out across the whole SRN but the selected sample consisted only of litter hotspots, the results of an RCT would likely provide a misleading picture on the effectiveness of the intervention. This is because in a country-wide roll out, also areas with low litter incidence would be covered, and not only hotspots.

**Figure 2** Selecting a representative sample for testing



The sample should be representative of the underlying population. This means that it should have similar characteristics to the rest of the population.

**Source:** *London Economics*

The target population could comprise of any specific areas/elements of the SRN where littering should be targeted. For example the population could comprise slip-roads leading from motorway service areas (MSAs) to the SRN, motorways, A roads, etc. (or a combination of these), or specific SRN users such as, lorry drivers, or occasional users.

The sample for the testing should be a representative cross-section of the target population. That is, the sample should feature similar characteristics as the population as a whole, and display similar incidences of littering. These characteristics will depend on the specific context of the intervention. In particular, the researcher should seek to include among the sampling characteristics any characteristics that are observable. Moreover the characteristics should be linked to a) the level of littering on the SRN, and b) the expected effectiveness of the intervention.<sup>2</sup>

A number of potential sampling characteristics for researchers are described in Box 2.

### Box 2 Potential sample characteristics

The following provides guidance on the types of characteristics that could be used to select a representative sample for an RCT to test an anti-littering intervention. If, for example, slip roads leading to the SRN are targeted by the intervention, it would be beneficial to characterise the nearby MSA, the Motorway or A road it leads to, and the typical slip-road users.

The actual selection of sampling criteria will depend on the availability of data on those criteria for potential members of the sample. Nevertheless, it is useful to initially assess as many characteristics as possible before narrowing the criteria based on their feasibility.

#### ■ Road characteristics

- Motorway, A road, B road, etc.;
- Length of the road stretch of interest;
- Number of MSAs (or other types of service area) on the road stretch of interest;
- Location (e.g. UK region, rural, urban, leading to/from urban centre, distance to major intersection, distance to MSAs, or other stopping areas);
- Speed limits (and typical speed of users);
- Incidence of litter (e.g. high/medium/low presence of litter, speed of accumulation);
- Road signs (presence of signs, and what types);
- Surroundings (e.g. presence of ditches, hedges, trees etc., hard/grassed surfaces, above/below ground); and
- Other (e.g. provision of bins, messaging signs, road works, bridges etc.).

#### ■ Slip-road characteristics

- Various characteristics also listed among the road characteristics above (location, incidence of litter, road signs, surroundings); and
- Characteristics of the road to which the slip-road joins.

#### ■ Motorway Service Areas

---

<sup>2</sup> Having an idea about the expected effectiveness of an intervention is also helpful for determining the necessary sample size (see next section). This is because larger effects can be detected in smaller samples. Very small treatment effects instead require much larger samples.

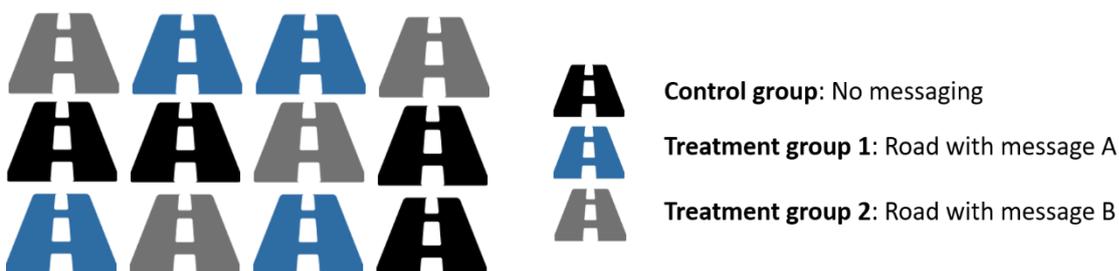
- Location (see above);
- Characteristics of the road on which the MSA is located.
- Company operating the MSA;
- Businesses operating at the MSA (e.g. coffee shops, supermarkets, fast-food restaurants etc.);
- Users (e.g. approx. volume of users, types of users, typical visit duration); and
- Typical user activities (e.g. shopping, parking, eating, fuel).
- **Road users**
  - Professional drivers (e.g. lorries, vans, coaches, taxis);
  - Regular private users (e.g. commuters); and
  - Occasional private users (e.g. holiday traffic).
- **Possible litter measurement characteristics**
  - Is litter monitored? And if so, how (e.g. measurement method, frequency)?
  - Who monitors litter?
  - What interventions exist (e.g. regular cleaning, messaging etc.)

## 2.2 Defining treatment and control groups

Once the test sample is defined, the researcher needs to split the sample into treatment and control groups. Only the treatment group(s) will receive the intervention(s). The conditions of the control group must remain unchanged because it uniquely serves for measurement purposes. This also means that the control groups should remain as isolated as possible from the treatment group(s) to avoid contamination effects from the intervention.

There will usually be one control group. At the same time there could be multiple treatment groups to test different interventions, or different variants of the same intervention, see Figure 3. To keep the later analysis of the RCT (see section 2.5) as simple as possible, the control and treatment group(s) should be of similar size.

**Figure 3 Treatment and control groups**



The sample will be split into groups some of which will receive the intervention(s) – the treatment group(s), other will not receive the intervention – the control group (or baseline).

*Source: London Economics*

The key guiding principle for assigning members of the sample to the control and treatment groups is that the split should be done using a random draw. However, there are alternative ways to make this randomisation, which mainly depend on the size of the sample and are described in turn in the following subsections.

### 2.2.1 Large sample RCTs

The most common form of RCTs are large sample RCTs. The key requirement for using this type of test approach is that there is measurement information on a large number of observations. This could be hundreds, or even thousands of observations in each test group.<sup>3</sup> To achieve this type of sample size, the unit of observation would typically be an individual, or a micro-level observation. This could be, for example, large numbers of individual road users, or specific miles on the SRN.

The researcher would randomly split the sample into two or more groups, and assign control or treatment status to each group. This could, for example, be done using a random number generator, tossing a coin, or rolling a dice. In the case study, see box 3, Lambeth council split 170 streets randomly into 5 groups.

Such randomisation of a large sample will ensure that the different groups resemble each other on observable characteristics (see Box 2 for examples). This is true because of the law of large numbers, which says that the outcome of a random draw will resemble the population average if repeated many times. As a result, we would expect the different test groups to be very similar, as long as they are sufficiently large. The researcher may do so-called randomisation checks to verify whether the samples are indeed similar. This could, for example, imply comparing the number of cars passing per day on given roads in both groups. Or comparing the frequency of litter-related complaints on service areas near test roads in the control and treatment group(s).

#### Box 3 Case study: A large sample RCT on street cleanliness in Lambeth Council<sup>4</sup>

The Institute for Fiscal Studies together with Lambeth Council collaborated to run an RCT. The objective was to engage citizens in street cleansing. Lambeth Council invited residents to become “Street Champions”. Street Champions were expected to be responsible for efforts to increase the cleanliness of their environment.

##### ■ The RCT design

- *The population:* streets in Lambeth London Borough Council.
- *The sample:* the study reached 170 streets out of 946 residential streets in Lambeth.
- *Splitting the sample:* using a random draw the sample was split into five groups of streets. Each treatment group tested different ways of encouraging greater levels of street cleanliness. Individuals in the control group received no communication from the Council about the Street Champions scheme.
- *The interventions:* the four treatment groups received different informational letters including mentions of specific incentives for participation.

##### ■ The RCT findings

- The simple offer of being a Street Champion motivated individuals to join the programme. Specific incentives encouraged enrolment further.

---

<sup>3</sup> The case study in the box below reports a sample of 170 streets as part of a large scale RCT.

<sup>4</sup> See IFS Briefing Note BN184 (2016), ‘An evaluation of different ways to incentivize citizens to co-produce public services in Lambeth’.

- No difference in street cleanliness was measured. Instead, the intervention increased beautification efforts in the treated streets and increased residents' satisfaction with their local area.

### 2.2.2 Small sample RCTs

In case a large sample is difficult or impossible to obtain, small samples can be arranged to mimic the large sample setup. A small sample RCT would be conducted if the measurement is not available at a fine-grained level. For example, when measurement information is not available for specific miles of the SRN, but is instead available on a more aggregate level such as complaints registers for local authorities.

With large sample RCTs, the law of large numbers guarantees that the different test groups are similar. For small sample RCTs, this is not the case. It is instead necessary to construct the control and treatment groups such that they are **as similar as possible** on specified characteristics (see Box 2 for possible characteristics).

For instance, if a test should consist of two slip roads close to MSAs, they should resemble each other in key characteristics such as:

- Type of MSA;
- Number of users per day or month;
- Litter measurement practices, etc.

Careful selection should minimise the difference between the control and treatment groups. For example, the number of road users should be similar because this may affect the number of litter-related complaints, or the quantity of litter disposed on the SRN. Only if the control and treatment groups are similar, the researcher can be sure that any measured impact of the intervention will be attributable to the tested intervention. This is because the intervention would be the only meaningful difference between the two groups during the intervention period.

It is good practice to assess whether the groups are indeed similar. The researcher should look at historical results of the success criteria (see section 2.3) in both groups. For example, by looking at the number of complaints received by both groups in past months. If the historical results are identical, or follow a similar trend, the researcher can be confident that the RCT results will be clean and robust.

In a final step, the researcher should randomly assign which group becomes the control group and which receives the treatment (i.e. intervention). Ideally, this should be unobservable to the participants in the test (e.g. single blind). It is important not to choose deliberately which group will be assigned to the treatment. This could affect the results and interpretation of the results. For instance, it should be avoided assigning the intervention to the group that currently receives more littering complaints just because the researcher expects the intervention to produce a larger uplift in this group.

**Box 4**      **A small sample RCT as an alternative to the “Corley Vehicle Litter Campaign”**

Keep Britain Tidy tested the impact of a poster intervention on littering behaviour on the SRN at the Corley MSA, in 2016.

The research consisted of two four-week phases: the first to establish a baseline, the second to measure the impact of the posters. The monitoring site was cleansed at the beginning of each monitoring phase and on completion of the trial.

The study findings suggest that the poster intervention reduced littering on the slip road. However, using as the baseline (i.e. control group) the same road site four-weeks prior to the intervention might provide a misleading picture of the effectiveness of the intervention. For instance, the reduction in littering might not be (entirely) caused by the poster intervention. It cannot be ruled out that other factors, such as a reduction in the number of road users during the trial period (e.g., due to seasonality) changed the amount of litter that accumulated.

Had the research instead been implemented as a small sample RCT, then the researcher could be more confident about the resulting evidence. In fact, the RCT methodology would allow the researcher to conclude that the measured effects are indeed caused by the poster and not by other factors related to the timing of the trials.

A small sample RCT could have consisted of:

- **Establish a baseline:** Monitor a small number (e.g. 4) of litter sites over a four-week period.
- **Split sample:** Create two groups (e.g. two pairs of sites) such that they look alike for the baseline measure. For example, each group could contain a high and a low litter site.
- **Assign control and treatment status:** Flip a coin to decide which group receives the poster intervention.
- **Implement:** Install the poster at the treatment sites, leave the control site without an intervention.
- **Measure:** Monitor all sites over the four-week trial period. Ensure that litter collection happens at the same intervals on control and treatment sites.
- **Analyse:** Compare the number of litter bags collected at treatment and control sites. If less litter accumulated at the sites with the poster intervention, it can be concluded that the poster was effective.

## 2.3 Success criteria

Once the RCT is implemented, the researcher needs to define success criteria so as to be able to assess whether the tested intervention(s) proved successful.

In an RCT, the success of an intervention is determined through a comparison between the success criteria found in the control group as compared to the treatment group. In the context of littering, this means, for example, that an intervention was successful if the roads of the treatment group accumulate less litter than those of the control group.

The following list shows success criteria that could be recorded for the control and treatment groups, and used to assess the effectiveness of the anti-littering intervention(s).

- **Measuring litter quantities:** this success criterion would require collecting litter at the control and intervention sites. The comparison could take place at different points in time, as long as litter is collected at the same time from both the control and treatment groups during, or at the end of, the RCT trial period. This measure could consist of tracking:
  - **Litter collected volume:** e.g. number of bags collected per 100yd;
  - **Litter collected weight:** e.g. weight of litter bags per 100yd;
  - **Litter accumulation speed:** measuring the speed at which litter accumulates in a given time frame;
  - **Binned litter at service areas:** if the anti-littering intervention(s) take place near service areas, it might be relevant to compare the number of bags collected from litter bins from service areas belonging to the control and treatment groups, as well as litter accumulated on adjacent roads. This is interesting in order to monitor the success of the intervention, and to verify for the presence of spillover effects. The intervention was successful if more litter is found in bins at the treated service areas and less litter on adjacent roads.
- **Measuring perceived litter:** these criteria do not require litter collection at the control and treatment sites. In fact, whether the intervention proved successful or not will depend on perceptions regarding the cleanliness of the roads.
  - **Perception surveys:** road users could be asked to rate the cleanliness of specific roads belonging to the control and treatment sites;
  - **Pictures:** pictures of the sample roads, taken just before the RCT and at different times during the trial, could be shown to people to let them judge on road cleanliness and accumulation of litter over time<sup>5</sup>;
  - **Received complaints:** the number of litter-related complaints could be counted and compared between the trial groups.
- **Litter spillovers:** road litter does not only cause detrimental consequences for the environment, but can also affect road behaviour. Hence, it is also possible to assess the success of the intervention(s) by analysing whether the control and intervention sites display differences regarding:
  - **Number of accidents:** road accidents reported to the police during the RCT;
  - **Road infringements:** road infringements reported during the RCT;
  - Other road safety statistics.

## 2.4 Roll-out

When performing RCTs, the researcher might find it useful to follow certain practical considerations for rolling out the control and treatment groups. These may comprise:

---

<sup>5</sup> When the success of an RCT is measured using public perception measures, the results will likely depend on type of litter at the test sites. For example, large and colourful litter items such as plastic bags and bottles harm the perception of (littered) areas more than smaller debris.

- Preparing the test sites. For example, if a cleaning is envisaged to take place prior to the intervention, the researcher should ensure that all sites are cleaned to a comparable standard.
- Monitoring the sites where the intervention is implemented. That is, the researcher should make sure that the intervention is successfully implemented at the scheduled time and place.
- It might be useful to collect information from road users to get some feedback regarding the design of the interventions while the RCT is ongoing.
- Ensuring that the success criteria are properly measured.
- Ensuring that the timing of the RCT takes into account timeframes and seasonality (e.g., it may be necessary to run the RCT in both summer and winter months, or during holiday and non-holiday seasons).

### 2.5 Analysis

This section explains how the researcher can effectively produce and present the results of the RCT.

The most important piece of analysis will be to measure and present differences between the control and treatment group(s) – i.e. the **treatment effect(s)**. This can be done by measuring the **success measures** (see section 2.3) separately for the control and treatment group(s).

The researcher will be able to conclude that the intervention was successful if the success measures have improved in the treatment group compared to the control group.<sup>6</sup> For example, a messaging campaign was successful if during the intervention period there was less litter accumulated at the site with newly installed litter bins compared to a site without new bins.

#### Box 5 Case study: Results of a hypothetical RCT on the SRN

A new messaging campaign has been tested between the two groups of A roads. Each group consisted of fifteen 2km stretches. Prior to the test, the two groups of 2km stretches resembled each other on numerous observable characteristics, such as litter incidence, location, and traffic intensity and followed similar seasonal littering patterns in historic measurements.

A coin flip has assigned one group as the control, and the other as the treatment group which was later exposed to the new messaging campaign.

Litter accumulation in both groups was measured for 4 weeks following a road cleansing to Grade A at all involved sites.<sup>7</sup> During this period 100 rubbish bags were collected in the control group, while in the treatment group only 60 bags were collected. This represents a 40% decrease of litter in the treatment group compared to the control group which is attributable to the new

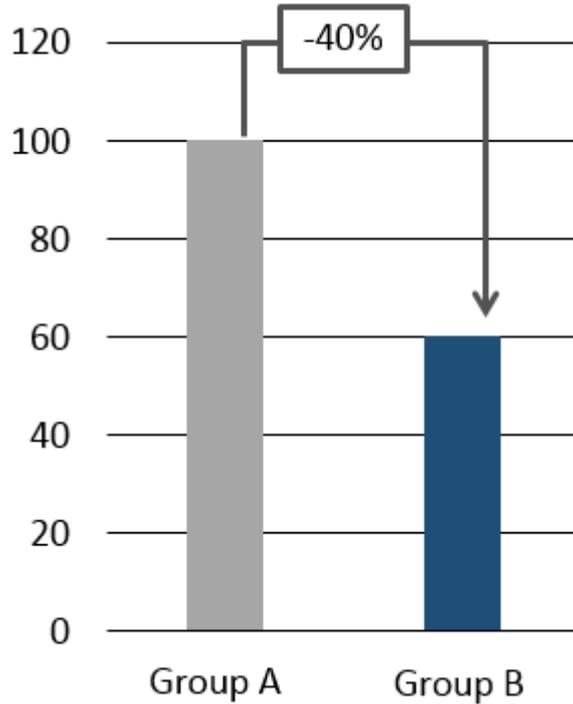
---

<sup>6</sup> See Annex 2 for guidance on statistical significance and hypothesis testing.

<sup>7</sup> Litter accumulation is often measured on a scale from A to D where A means “No litter or refuse” and Grade D means “Heavily affected by litter and/or refuse with significant accumulations”. The scale is defined in the Code of Practice on Litter and Refuse, Department for Environment, Food and Rural Affairs (2006).

messaging campaign. The RCT has thus shown that the intervention was successful because littering was reduced.<sup>8</sup>

**Figure 4** Illustration of a treatment effect measured in number of collected garbage bags during the measurement period



## 2.6 Piloting

The RCT could be piloted prior to rolling it out across the entire targeted sample. This implies running the RCT first on a smaller scale. Exposing, for example, a small number of MSAs to the RCT first, measuring the effectiveness of the piloted intervention, then exposing the full sample to the RCT.

The same methodological steps that apply to the roll-out also apply to the pilot. The only difference is that the stakes involved in a pilot are typically lower compared to a full roll-out, for example nationwide. A pilot furthermore allows the researcher to make adjustments to the intervention based on the pilot results prior to roll-out.

### Box 6 Case study: Piloting roadside funnel bins

A pilot could pre-test innovations to the 'binrastructure' such as the proposed funnel bins on the roadside targeted at lorry drivers. For this, prototype bins could be installed for a limited period of time. The effects of these bins would be assessed for the pilot sample. Using the pilot

<sup>8</sup> See Annex 2 for guidance on statistical significance and hypothesis testing.

results, the likely effects for a wider roll-out could be projected which could provide robust evidence to justify investments in large numbers of bins across the country.

#### 2.6.1 Pre-testing via interviews

A pilot could be accompanied – or, if a real pilot is impossible to conduct, replaced – by a pre-testing using interviews. Such interviews could be conducted with selected individuals who would be exposed to the intervention to enquire about their understanding of, and likely reactions to, the proposed intervention.

This can be useful to refine the intervention. However, such pre-tests via interviews are unlikely to deliver as robust results as a pilot that measures the actual littering behaviour. This is because self-reported behaviour is often less reliable compared to observed behaviour. This is especially the case in littering because interviewees may overestimate the effectiveness of the proposed intervention due to social-desirability (e.g. *“Of course I would use the roadside bin, this is a great invention”* or *“I would never throw litter onto the road”*), or due to misperceptions of their own behaviour.<sup>9</sup>

## 3 Conclusions

Researchers and policy makers have a number of methodologies available for testing the effectiveness an intervention. RCTs have some particularly desirable characteristics because they are relatively easy to set up, and produce reliable results. For these reasons RCTs are becoming increasingly popular not only among academics but also among practitioners and policy makers.

RCTs are a valuable tool for testing anti-littering interventions on the SRN. This report has put forward a step-by-step guide for designing and conducting RCTs on the SRN.

To conclude this guide, it should be noted that while the mechanics of RCTs are simple, their design and implementation might prove nonetheless challenging in the specificities of a particular intervention. An experienced researcher who is familiar with the limitations and risks involved in the methodology should thus accompany any RCT. The described methodology should only be applied if the outlined necessary conditions are likely fulfilled and all steps involved can be followed. This is because results from poorly designed RCTs may not be valid and thus may misguide policy recommendations.

---

<sup>9</sup> Evidence suggests that many people, who litter, do not see themselves as “litterers” and may therefore be unresponsive to anti-littering messages which are targeted specifically at “litterers”.

## **ANNEXES**

## Annex 1 References

Department for Environment, Food and Rural Affairs (2006) '*Code of Practice on Litter and Refuse*', available at:

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/221087/pb11577b-cop-litter.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/221087/pb11577b-cop-litter.pdf), accessed 18.05.2017.

HM Government (2017) '*Litter Strategy for England*', available at:

<https://www.gov.uk/government/publications/litter-strategy-for-england>, accessed 18.05.2017.

IFS Briefing Note BN184 (2016) 'An evaluation of different ways to incentivise citizens to co-produce public services in Lambeth', available at: [www.ifs.org.uk/uploads/publications/bns/BN184.pdf](http://www.ifs.org.uk/uploads/publications/bns/BN184.pdf), accessed 18.05.2017.

Keep Britain Tidy (2016) "Case Study: Corley Vehicle Litter Campaign", available at: <http://labs.keepbritaintidy.org/Documents/News/DownloadLinks/Corley%20Case%20Study.pdf>, accessed 19.05.2017.

Kolodko, J., Read, D., and Taj, U. (2016) '*Using Behavioural Insights to Reduce Littering in the UK*', for Clean Up Britain, available at: <http://cleanupbritain.org/WBS-Report-for-CLUB.pdf>, accessed 18.05.2017.

## Annex 2 Statistical analysis of treatment effects

In addition to analysing the effectiveness of the tested intervention using descriptive methods (e.g. counting litter volumes, or comparing the number of received complaints in the treatment and control groups), statistical methods can be applied. These methods would assess the robustness and credibility of the measured effects of the RCT.

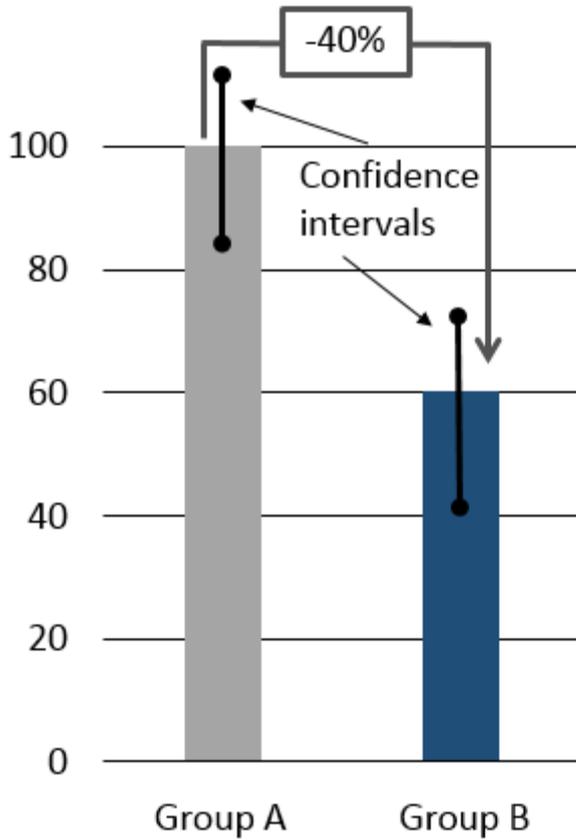
In a first step, this can be done by measuring the **mean of the success measure**, for example, the average litter volume accumulated over a specified period, in each of the test groups using descriptive statistics.

A further step then establishes whether any measured treatment effect is also statistically significant. This can be done using confidence intervals and t-tests.

- **Confidence intervals:** Confidence intervals illustrate the margin of error around reported results and are calculated for a chosen confidence level (typically 95%). The confidence level represents how 'sure' we can be that the true result lies within the confidence interval. Generally, the 95% confidence interval is calculated using the formula: Lower bound =  $M - Z_{95} * \sigma_M$ ; Upper bound =  $M + Z_{95} * \sigma_M$ . Where  $M$  is the sample mean,  $Z_{95}$  is the number of standard deviations extending from the mean of a normal distribution needed to include 95% of the area, and  $\sigma_M$  is the standard error of the mean. If the sample means of two groups do not have overlapping confidence intervals, it can be concluded that the difference in means is statistically significant.
- **T-tests:** This test will establish whether a resulting treatment effect is statistically significant. The null hypothesis would be  $L_1 = L_2$ , i.e. that litter volumes are the same in both groups, and the test would establish whether the null hypothesis can be rejected in favour of the alternative hypothesis  $L_1 \neq L_2$ , i.e. the average litter volumes are not the same. The result of the test would depend on whether the t-statistic calculated from the data exceeds the relevant critical value.

In RCTs with sufficiently large samples, the analysis can be augmented by assessing treatment effects for different sub-groups within the test samples. For example, it could be assessed whether the effectiveness of the treatment was larger in magnitude in littering hotspots, compared to sites with usually slower litter accumulation. Such analysis could comprise statistical tools such as linear regression analyses.

Figure 5 Assessing statistical significance of treatment effects



Notes: If the confidence intervals on the averages of the outcome measure in each of the test groups, such as the number of litter bags collected in Groups A and B, are not overlapping, it can be concluded that the measured treatment effect is statistically significant. This means that the difference in measured littering is too large to be measured by pure coincidence. Instead it was effectively caused by the tested intervention.

Source: London Economics

## Index of Tables, Figures and Boxes

### Figures

Figure 1	The key steps of an RCT in brief	2
Figure 2	Selecting a representative sample for testing	3
Figure 3	Treatment and control groups	5
Figure 4	Illustration of a treatment effect measured in number of collected garbage bags during the measurement period	11
Figure 5	Assessing statistical significance of treatment effects	16

### Boxes

Box 1	Possible interventions to test using an RCT	2
Box 2	Potential sample characteristics	4
Box 3	Case study: A large sample RCT on street cleanliness in Lambeth Council	6
Box 4	A small sample RCT as an alternative to the “Corley Vehicle Litter Campaign”	8
Box 5	Case study: Results of a hypothetical RCT on the SRN	10
Box 6	Case study: Piloting roadside funnel bins	11





Somerset House, New Wing, Strand,  
London, WC2R 1LA, United Kingdom  
[info@londoneconomics.co.uk](mailto:info@londoneconomics.co.uk)  
[londoneconomics.co.uk](http://londoneconomics.co.uk)  
[🐦 @LondonEconomics](https://twitter.com/LondonEconomics)  
+44 (0)20 3701 7700