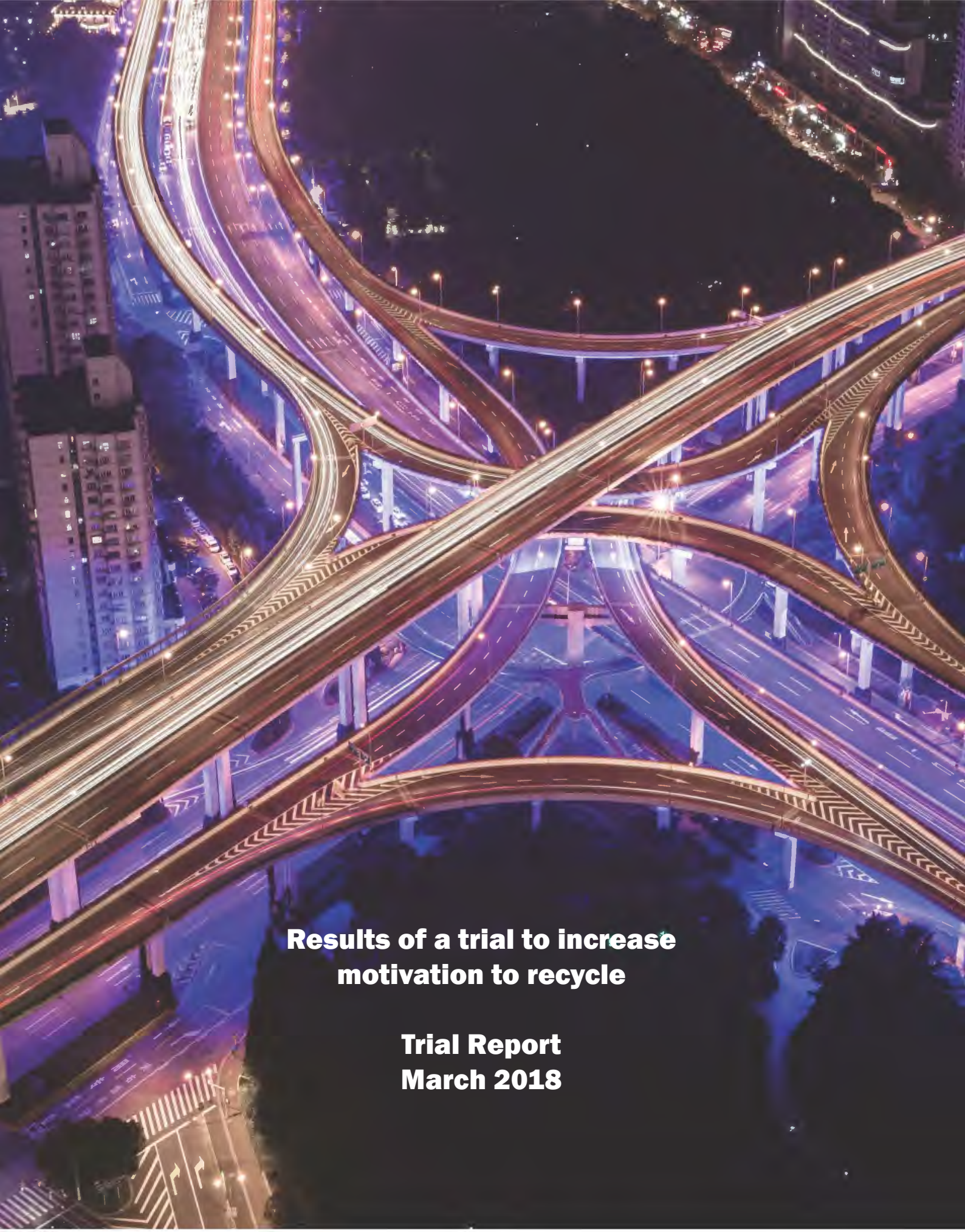


Results of a trial to increase motivation to recycle

Trial Report
March 2018

Department of Finance
An Roinn Airgeadais
www.finance-ni.gov.uk

Innovation Lab

Table of contents

Executive Summary

Background

During 2014/15 the household preparing for reuse, dry recycling and composting rate was 42%. This fell short of the Northern Ireland Programme for Government and Northern Ireland Waste Management Strategy targets of 45% by 2015 and 50% by 2020. There is a clear need for corrective action at a local council level in order to encourage more domestic recycling if Northern Ireland is to achieve future waste targets.

Method

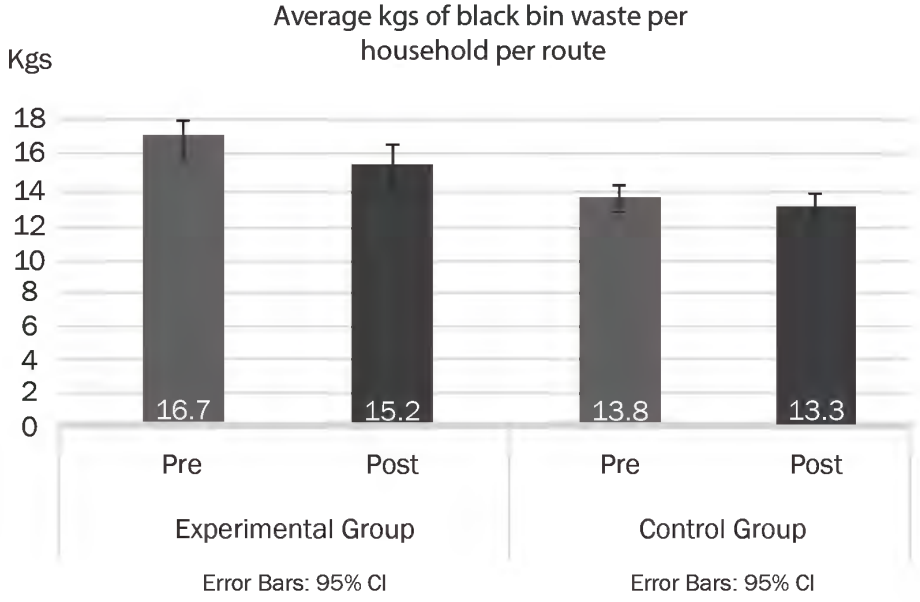
User research, idea generation workshops, desk-based research, and behavioural analyses were combined to design two different paper door hangers that would be hung on doors of individual houses along 18 routes in two council areas: Armagh, Banbridge, and Craigavon (ABC) and Antrim and Newtownabbey. The door hangers were intended to prompt recycling. A quasi-experimental design was used to test the efficacy of the door hangers: 9 routes received the door hangers and 9 routes did not receive anything. The weight of landfill waste, garden and food waste, and dry recycling were measured for 13 weeks before the first intervention to get a baseline. Door hangers were deployed twice with about 4 weeks between each deployment, and the weight of the waste was measured for about 9 weeks after the first deployment. We then ran statistical tests comparing these results to find out if there was a significant difference that we could attribute to the intervention.

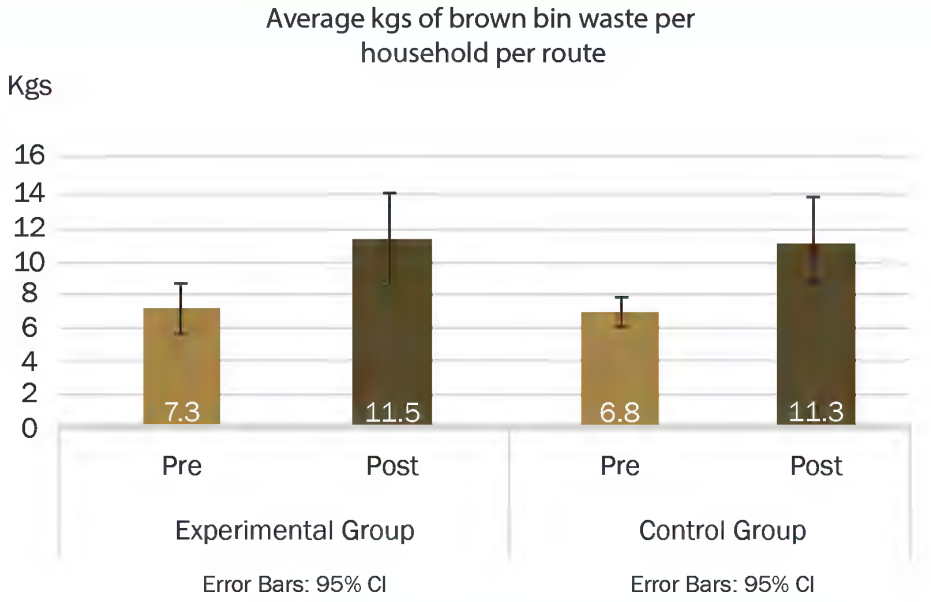
Door Hanger 1: Front and back

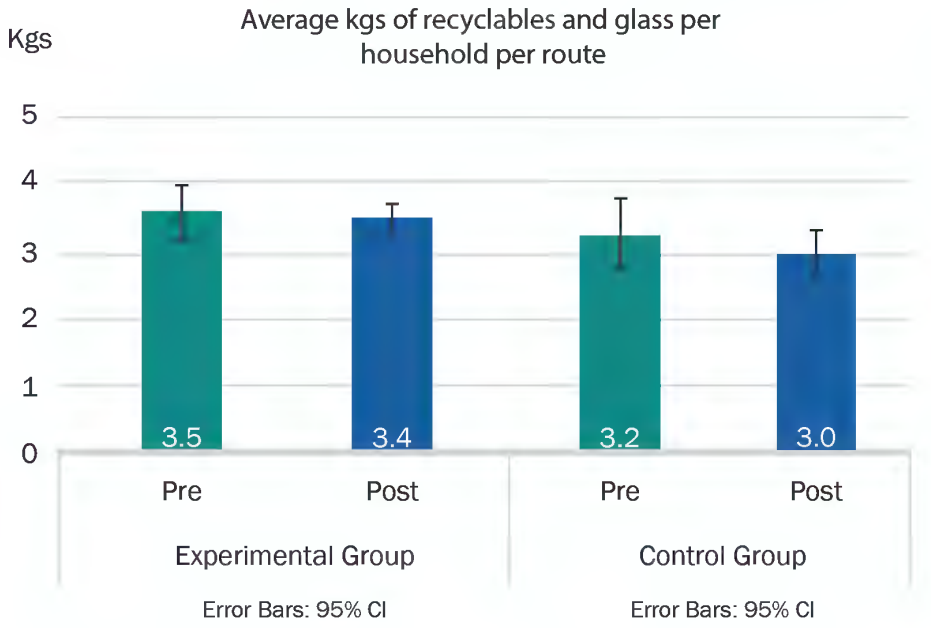
Results

There were reductions in landfill waste and dry recycling and an increase in food and garden waste collected over the trial period (see graph below).

However when we compared the control group to the intervention group, we found no statistically significant differences. This means that we are unable to tell if the differences are due to chance, if the intervention had no effect, or if the trial did not have enough power to detect this size of effect.







Conclusions

Importance of a strong comparison group

Landfill waste decreased and garden and food waste increased for the group that received the door hanger. However, the same pattern was observed in the control group over the same time period. If we had not had a comparison group, we might have made a recommendation to roll out this intervention across Northern Ireland which would have cost the councils time and money; and yet, because we included a control group, we can conclude that the intervention's effect was no better than chance, and we can experiment again to see what might work better next time.

The power of experimentation

Innovation is a process, not a one-off experiment, or an event. Our door hangers were not enough to prompt an increase in recycling or increase in landfill waste. But by experimenting, we can trial different messages or different behaviour change techniques until we find one that works. In addition, by using robust statisticial techniques, we can determine what sample size (number of routes/individuals) is needed to find a significant difference between the groups, if the intervention does have an impact.

The Innovation Lab and its method

...connecting, collaborating, listening, failing fast, learning, disrupting, inventing, and enabling.

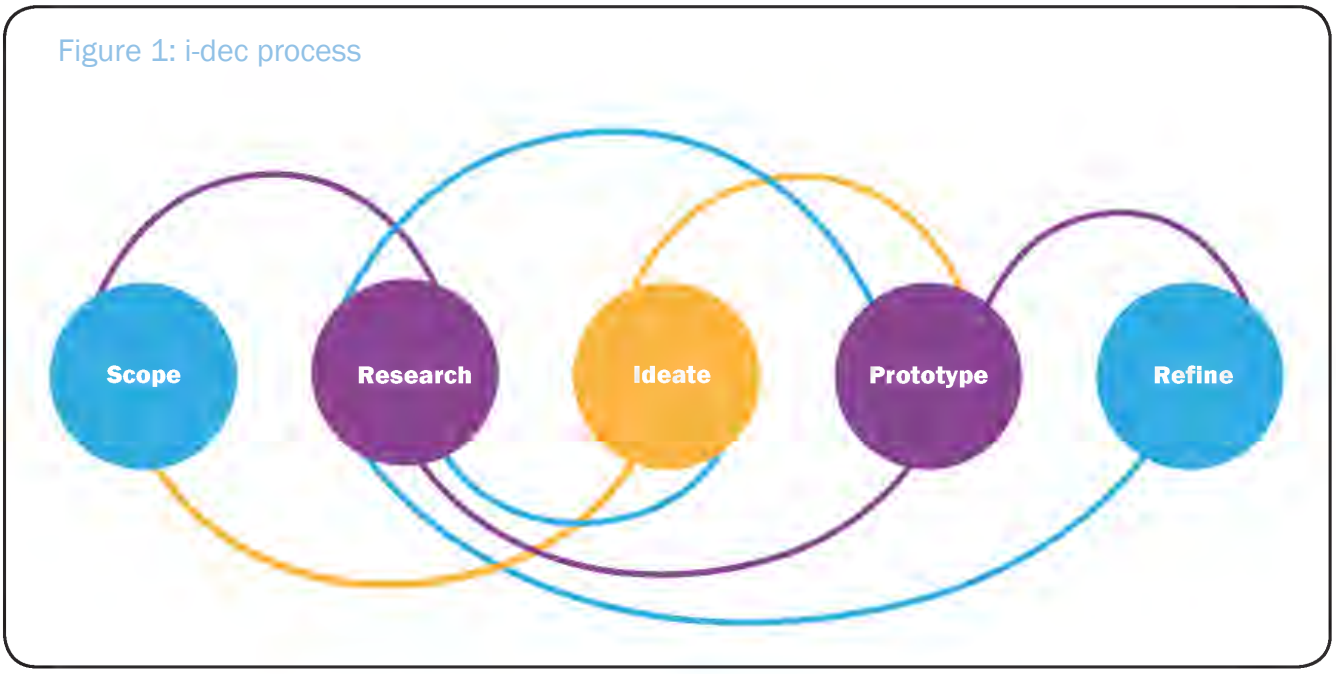
The Innovation Lab was established in 2014 and sits within the Department of Finance. The Lab has a role in Northern Ireland's Innovation Strategy for creating a culture of innovation by encouraging collaboration, openness to new ideas, innovation, and risk taking.

The Lab responds to challenges where effective service provision for the public has proved most difficult. It aims to improve public services by creating new and ground-breaking innovations through transformation and invention. We are committed to inspiring curiosity, empowering creativity, and bringing to life paradigm-shifting ideas. We believe in connecting, collaborating, listening, failing fast, learning, disrupting, inventing, and enabling. Our i-dec philosophy has been developed to address these challenges.

i-dec - innovation through design, experimentation and creativity

Our i-dec philosophy is built on design principles. Namely, putting users first, understanding relationships, developing prototypes, testing iteratively, and scaling up solutions which work. Our process is iterative and not stage-gate; projects will move backwards and forwards depending on what we learn and the ideas we can surface.

Figure 1: i-dec process



Testing interventions

In the Lab we know that not all ideas are good and not all good ideas work. Therefore, we do user testing and rigorous experimentation with every idea that we think might work. We work with our sponsors and stakeholders to select ideas and design methods to test ideas in lab settings or in live environments. Where we can, we propose randomised control trials to test interventions against a control group so that we can measure and validate the effect of the intervention.

Behavioural sciences and the Innovation Lab

The Lab has identified that behavioural science offers new ways of approaching problems and had delivered results in other contexts. It has been working to develop capacity and capability in behavioural science and has been developing behaviourally inspired solutions to business areas across the Northern Ireland public sector.

In essence, this stream of work applies psychological and social science insights to public sector problems with the specific aim of changing or influencing people's behaviour. This is a relatively young field with increasing applications across public policy problems. An important part of this work is using randomised control trials or experiments to test the effectiveness of interventions.
The Lab has been developing services using behavioural science which include the following:

- Desk-based research on comparable experiments
- Design based research on behavioural journeys and existing choice architecture
- Intervention Design
- Intervention Re-design
- Experiment design
- Experiment implementation and analysis

This project consisted of design based research, such as journey mapping and workshops with NIHE staff, and intervention design and re-design.

The Waste Project

The Department of Agriculture, Environment and Rural Affairs (DAERA) commissioned the Innovation Lab to take a fresh look at improving the quality and quantity of household kerbside recycling with a view to generating new ideas to meet the EU targets on recycling.

The target

The EU Waste Framework Directive has set out targets on how waste should be managed in the EU. In order to comply with the objectives of this Directive, Northern Ireland is to achieve the following target:

> by 2020, the preparing for re-use and the recycling of waste materials such as at least paper, metal, plastic and glass from households and possibly from other origins as far as these waste streams are similar to waste from households, shall be increased to a minimum of overall 50 % by weight[1]

Much work has been completed to date to maximise efforts to increase recycling rates and to continue the momentum the Innovation Lab prepared a workshop to facilitate the development of innovative ideas with DAERA and the Northern Ireland councils.

Idea generation workshop

The Lab conducted desk based research and citizen engagement, which identified three elements of the recycling system to shape discussion during a two day workshop.

> Infrastructure – the built environment, products and objects
> Service – the system providers and enablers that allow people to participate in a particular environmental practice
> Behaviour – relates to people, who we are and our disposition towards the environmental practice

We held the workshop over two days in March 2016. The participants included representatives from central government, Welsh government, all NI local councils, and experts from the NI and UK voluntary sector, industry and academia. The participants were taken through a number of exercises based on the insights gathered during the research and citizen engagement phase.

We asked participants to come up with ideas which addressed issues they had identified through their discussions. The workshop concluded with a broad consensus on 17 ideas and on their priority order for implementation.

One of the ideas which surfaced was to put stickers on bins. With its expertise in behavioural science, the Lab took this idea and developed it further. This report is an analysis of what we did and how it worked.

[1] You can read more about the model at www.behaviormodel.org (accessed 16th March 2018).

Research

Behavioural analysis and interventions

The theory

During the waste project we were applying BJ Fogg's Theory of Persuasive Design. (Since then we have been developing our expertise in other behaviour change models and are now primarily working with COM-B.[2]) Fogg theorises that behaviour can be explained by a combination of motivation to perform the behaviour, ability to perform the behaviour, and a trigger to perform the behaviour.

> Motivation: This refers to the impulses and desires which drive behaviours. Fogg argues that there are three core motivators: pleasure / pain; hope / fear; and social acceptance / social rejection.
> Ability: In this model, ability refers to how easy a behaviour is to perform. It has the following aspects: time, money, physical effort, brain cycles, social deviance, and non-routine.
> A Trigger: This is a prompt which initiates the behaviour. Fogg argues that this is essential and states that a behaviour will not occur without a trigger, even if motivation and ability are high.

Fogg proposes different types of triggers depending on the analysis of the drivers of behaviour:

- a spark is for people who have ability but lack motivation;
- a facilitator is for people who have motivation but lack ability; and
- a signal is for people who have both motivation and ability.

Therefore, the design of an intervention should include a trigger which is based on the analysis of the behaviour that you are trying to change.

The research on recycling

Separately, WRAP had identified seven behavioural categories of recyclers. We (roughly) mapped those categories onto Fogg's model (see Figure 2).

As you can see, there is a wide variety of explanations of behaviour. 52% of people probably only require signal triggers. 12% of people require complicated interventions to increase both motivation and ability. The remaining people are spread across behavioural categories.

Insight generation

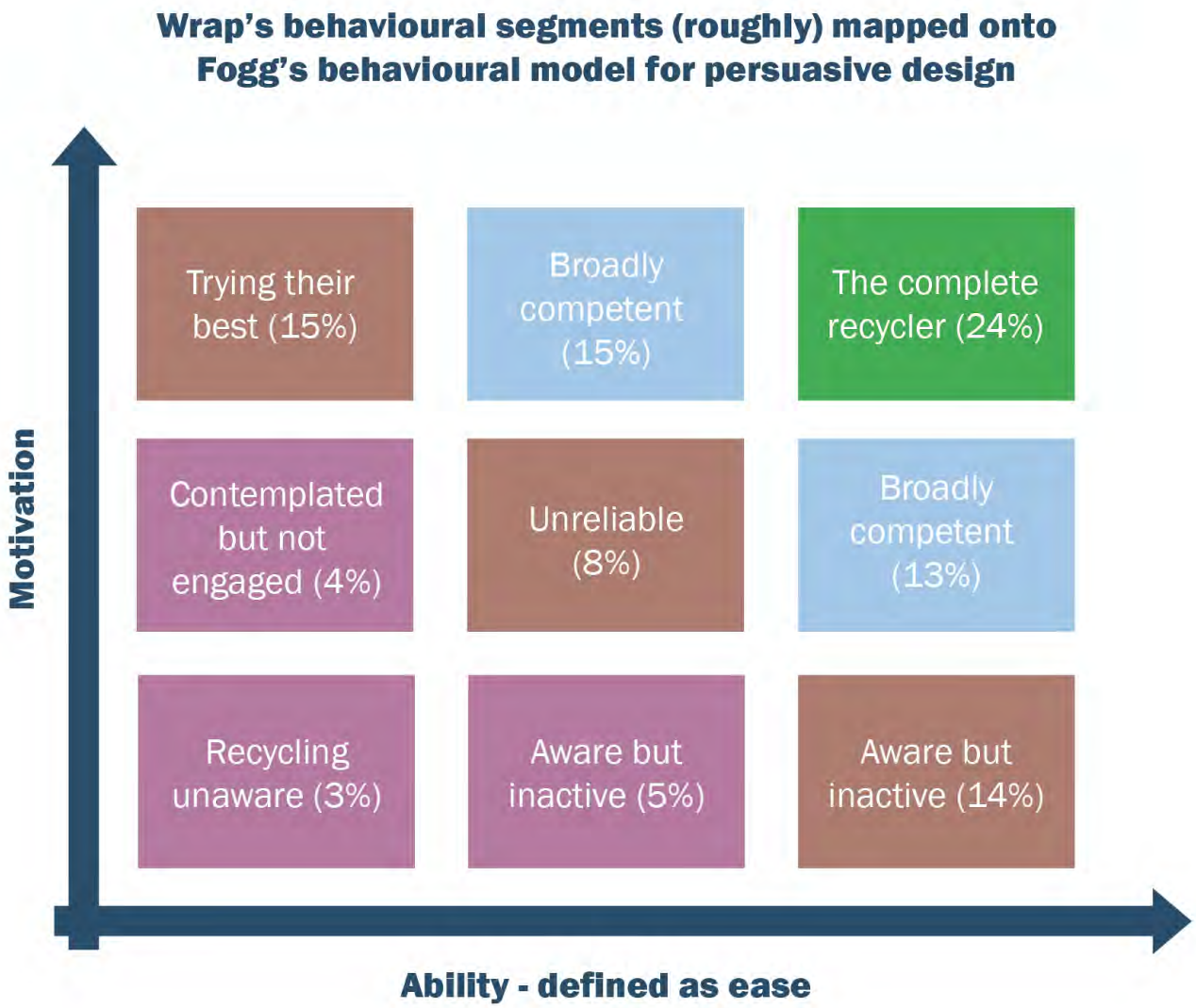
We also conducted some qualitative research to generate insight from citizens about their recycling attitudes and behaviours to help us with our analysis.

Five themes emerged:

- lack of knowledge about what happens to waste sent for recycling;
- persistence of myths about what happens to waste sent for recycling;
- lack of knowledge about how to recycle certain types of waste;
- lack of knowledge about what types of waste could be recycled; and
- environmental motivations for recycling.

[2] Michie et al.: The behaviour change wheel: A new method for characterising and designing behaviour change interventions. Implementation Science 2011 6:42.

Figure 2: Fogg's Behavioural Model



Desk Based Research

We did a search through google scholar looking for other interventions which had been trialled to improve recycling behaviour. There were not many that were relevant and so we expanded the search to look for interventions designed to improve other types of pro-environmental behaviours. We were unable to find very many of these either. The following are brief synopsis of the research that we did find and which inspired our thinking.

OPOWER trials on energy conservation: perhaps the most famous experiment using feedback and other techniques to change energy usage behaviour. OPOWER was able to reduce energy consumption by 1.9%-2% by providing tailored feedback to households. They found that effects decreased over time after an intervention but providing subsequent interventions increased the effects again.

Feedback intervention on recycling in Norway: a team of researchers in Norway tested the effect that personalised feedback on recycling would have on behaviours. They found that households that received letters increased their share of waste that they recycled by 2%. Households that received a letter promoting waste reduction increased their recycling but it is not clear that they reduced the total quantity of waste produced.[3]

Lab and field experiments on using normative appeals to influence sustainable consumer behaviours (such as recycling): Experiments tested three types of appeal to change behaviours: injunctive appeals (highlighting what others think you should do), descriptive appeals (highlighting what others are doing), and benefit appeals (highlighting the benefits of the action). They found that the effectiveness of the appeal depended on whether the individual or collective identity is activated.[4]

Distortion bias and identity bias in recycling habits: researchers found that people are more likely to recycle items which hadn't been distorted (eg a drinks can which had not been crushed) or which were linked to their identity (eg a Starbucks cup with their name on it).[5]

Since we conducted our research, a systematic review on psychological strategies to promote household recycling has been published. It identified that the following strategies had been used: prompts and information, feedback, commitment devices, incentives, environmental alterations, and social modelling. They found that the top four determinants of behaviour that interventions were designed to target were motivation, information and knowledge, beliefs and perception about the consequences of recycling, and social influence. It found that the most effective interventions used social modelling (using demonstrations of behaviour) and environmental alterations (ie making recycling easy and accessible).[6]

[3] Milford, Anna Brigitte, Arnstein Øvrum and Hilde Helgesen, 'Nudges to increase recycling and reduce waste', Norwegian Agricultural Economics Research Institute, Discussion Paper No 2015-01.

[4] White, Katherine and Bonnie Simpson, 'When Do (and Don't) Normative Appeals Influence Sustainable Consumer Behaviours?', Journal of Marketing, Vol 77, March 2013, 78-95.

[5] Trudel, Remi, 'The behavioral economics of recycling', Harvard Business Review, 7th October 2016, https://hbr.org/2016/10/the-behavioral-economics-of-recycling

[6] Varotto, Alessandra and Anna Spagnolli, 'Pschological strategies to promote household recycling. A systematic review with met-analysis of validated field experiments', Journal of Environmental Psychology, 51, 2017, 168-88.

Intervention Design

On the basis of the research, the insights we gathered, the behavioural theory, and other information from people working in the area we generated four different concepts for interventions.

These interventions were 'simple interventions' which built on specific behavioural insights to develop a trigger which was either a spark or a facilitator.

An ability based intervention using cognitive ease: this was based on insight that recycling was difficult and complicated and many people were not clear about what goes in each bin and what they needed to do to their waste beforehand. We looked for ways to design an intervention applying the behavioural insight of 'cognitive ease' to address this issue.

A motivational based intervention using feedback on behaviour: this would tell people how much they were recycling, how much their community or local area was recycling, and how well it compared to others.

A motivational based intervention using feedback on outcomes: this would tell people what the outcome of their recycling was, i.e. how many tonnes of CO_2 had been saved or how much waste had been redirected from landfill. We were looking for creative ways for communicating this information.

A motivational based intervention using reframing: This involves suggesting the deliberate adoption of a perspective or new perspective on the behaviour in order to change thinking or emotions about the behaviour.

We held a workshop with waste management officials from councils to brainstorm ideas for how to turn these concepts into reality. We assessed the ideas on the basis of their practicality, acceptability, and likelihood of success and, on this basis, we selected the motivational intervention using feedback on behaviour.

Constraints

We were faced with a number of constraints which impacted on the final design and possibly on the impact of the intervention.

- Data is held at route level, which meant:
  - we could not provide household level feedback; and
  - we could not easily provide community level feedback as routes do not service identifiable geographic communities.
- We could not provide short term feedback based on the success (or otherwise) of an intervention because data could not be verified and utilised quickly enough.

The Delivery Mechanism

The original idea was generated during our workshop was to put stickers on bins. This was inspired by a successful intervention developed by North Down and Ards Council. We conducted further research on what stickers have previously been used (in Northern Ireland and elsewhere) and what the evidence was for their effectiveness.

Given that there was an evidence base for stickers we decided to investigate other delivery mechanisms. We identified:

- Posted leaflets: leaflets that are individually addressed and delivered with the regular mail.
- Maildrops: leaflets that are delivered as part of promotional material
- Bin tags: tags which are placed on a bin when it is emptied
- Door hangers: hotel style hangers which are placed on a front door of a house

In order to decide which type of delivery mechanism to use, we assessed each method against the following criteria:

- Reliability of delivery: Must be received and read by every household
- Fidelity: Must be reliable carrier of message or information
- Practicality: can the intervention be delivered
- Cost: how expensive is it to deliver the intervention

Table 1: Delivery mechanism assessment matrix

	Reliability	Fidelity	Practicality	Cost	Total score
Posted leaflets	3	5	3	3	14
Maildrops	1	5	5	5	16
Bin tags	3	5	2	2	12
Door hangers	5	5	4	3	17

All options were considered to ensure the fidelity of the message. While maildrops were practically the easiest option and the cheapest, they were considered the least reliable for transmission of the message because they could either get lost among other junk mail and, moreover, were actually likely to be considered as junk mail. Post leaflets had similar considerations and would be more expensive. Door hangers and bin tags each had novelty value but door hangers were considered as a more reliable mechanism because it was possible to miss a bin tag. We also had concerns about the practicalities of deploying bin tags. On the basis of this analysis, door hangers were the preferred option.

The final design

We worked with a graphic designer from the Department of Infrastructure to turn the concept into a design. We developed three designs.

Design One: The first was a simple thank you

Design features:

- Feedback message - Thank you message
- Reinforcing imagery – gold star
- Reinforcing message – 'let's keep it up'
- Information feedback – tonnage from last year
- Messaging effect – signed by local bin men
- Reinforcement message (on rear) – making a difference to your community

Behavioural change techniques used:
(Note: we were not using the Behavioural Change Technique Taxonomy at this stage so this analysis has been provided post facto to clearly identify the linkages between the design and theory. The reference numbers (eg 7.1) refer to the classification system of the BCT Taxonomy)

7.1: prompt / cue: an environmental or social stimulus with the purpose of prompting or cueing the behaviour. The prompt or cue would normally occur at the time or place of performance.
2.2: Feedback on behaviour: Monitor and provide informative or evaluative feedback on performance of the behaviour (eg form, frequency, duration, intensity).
2.7: Feedback on outcome of behaviour: Monitor and provide feedback on the outcome of performance of the behaviour.
5.3: Information about social and environmental consequences: Provide information (e.g. written, verbal, visual) about social and environmental consequences of performing the behviour.
9.1: Credible source: present verbal or visual communication from a credible source in favour of or against the behaviour.
15.3: Focus on past success: advise to think about or list previous successes in performing the behaviour (or parts of it)

Design Two: The second was a social norm feedback

Design features:

- Feedback message – We're a first class recycling community
- Social norm message – community focused
- Reinforcing imagery – rosette
- Messaging effect – signed by local bin men
- Feedback message (on rear) – thank you
- Reinforcement message (on rear) – let's keep it up
- Feedback message (on rear) – we recycled more than ever before

Behavioural change techniques used:

7.1: prompt / cue: an environmental or social stimulus with the purpose of prompting or cueing the behaviour. The prompt or cue would normally occur at the time or place of performance.
2.2: Feedback on behaviour: Monitor and provide informative or evaluative feedback on performance of the behaviour (eg form, frequency, duration, intensity).
2.7: Feedback on outcome of behaviour: Monitor and provide feedback on the outcome of performance of the behaviour.
6.3: Information about others' approval: provide information about what other people think about the behaviour. The information clarifies whether others will like, approve, or disapprove of what the person is doing or will do.
6.2: social comparison: draw attention to others' performance to allow comparison with the person's performance
9.1: Credible source: present verbal or visual communication from a credible source in favour of or against the behaviour.
10.4: Social reward: Arrange verbal or non-verbal reward if and only if there has been effort and/or progress in performing the behaviour (includes positive reinforcement)
15.3: Focus on past success: advise to think about or list previous successes in performing the behaviour (or parts of it)

Design Three: Thank you for recycling food waste

Design features:

- Feedback message - Thank you message
- Reinforcing imagery – man with thumbs up and pictures of food waste
- Messaging effect – signed by local bin men
- Reinforcement message with instructions (on rear) – keep putting peelings and leftovers in brown bin

Behavioural change techniques used:

7.1: prompt / cue: an environmental or social stimulus with the purpose of prompting or cueing the behaviour. The prompt or cue would normally occur at the time or place of performance.
2.2: Feedback on behaviour: Monitor and provide informative or evaluative feedback on performance of the behaviour (eg form, frequency, duration, intensity).
4.1: Instruction on how to perform the behaviour: Advise or agree on how to perform the behaviour.
9.1: Credible source: present verbal or visual communication from a credible source in favour of or against the behaviour.

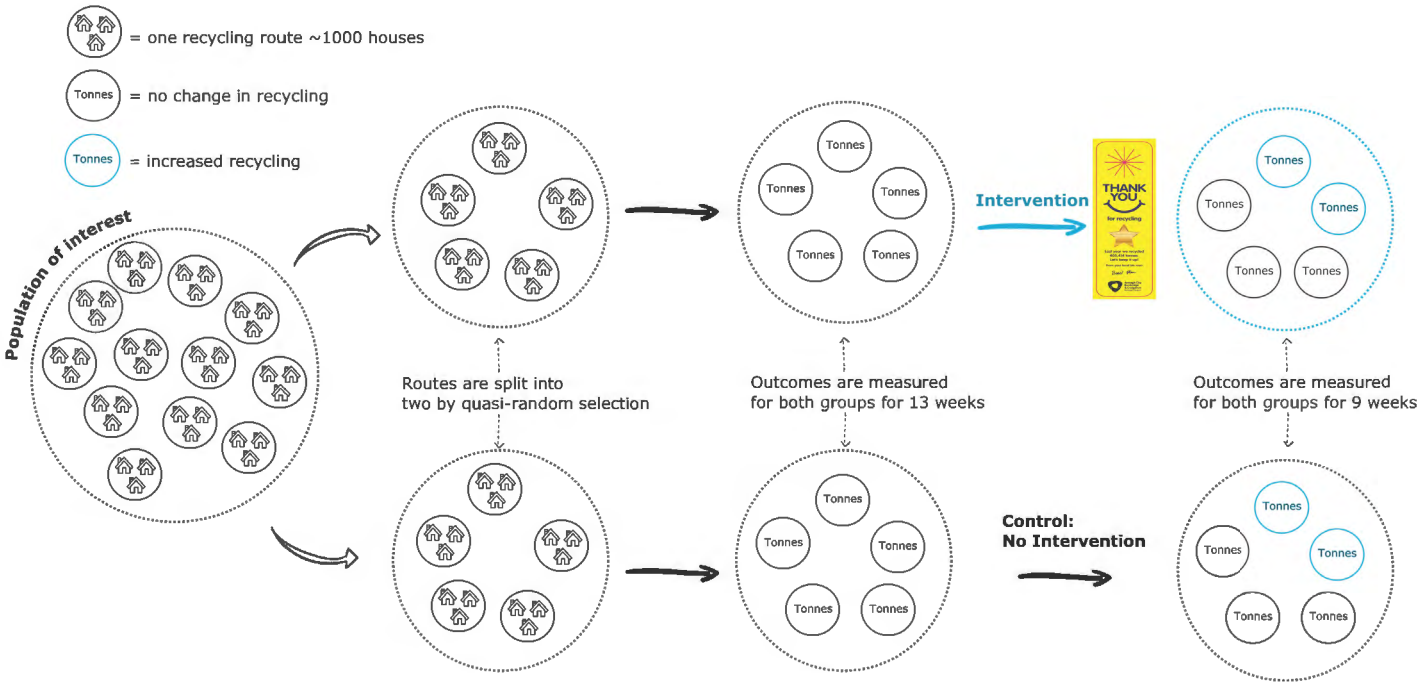
All three designs performed the function of a signal trigger by providing a prompt to recycle. In general, we also hypothesised that feedback provides a motivational function in this context. The social signalling in the second design was also a spark trigger and the third design had a facilitation trigger (performed by the instructions on the reverse of the hanger).

The translation of concept to design was inevitably a messy process where the ideation process and the constraints of the delivery mechanism inevitably meant that theory and concepts became blurred in some places. However, we were satisfied that the final design was theoretically informed and provided a real test of assumptions about how to change recycling behaviour.

The Trial Design and Analysis

The trial protocols are included in Annex 1 and 2. We used a quasi-experimental design (QED) to determine if the interventions were effective. We used a QED methodology primarily because a full randomised controlled trial (RCT) was not possible because only route level data was available and therefore there was insufficient power to run a full randomised controlled trial (RCT). However, the principles behind a QED are very similar to that of an RCT. In this case routes were assigned to either a control condition or an intervention condition. The results of each condition were measured and compared against each other (see Figure 3).

Figure 3: The quasi-experimental design used for the recycling trial



We worked with two council areas and matched 8 routes in one area and 9 routes in the other. That gave us around 4,500 households in each area which were assigned to the intervention group. The households which were assigned to the intervention group received two interventions: one at week 1 and one at week 6. We used Design One as the second intervention in both areas, while one area received Design Two as the first intervention, and the second area received Design Three.

We then measured the weight of the vehicles which collected the waste and recycling about 13 collections before the intervention and 9 collections afterwards. These measurements were made at a route level. From the data we calculated how much waste the average household collected that week on each route. We then used these figures to compare the two groups. We used a 'difference in differences' analysis which essentially compares whether the change over time in one group is statistically different from the other group. A statistically different result is one which passes a test to determine whether it is likely that the change can be accounted for by mere chance or whether it is the result of the intervention.

Further technical detail on the data collection, processing, and analysis can be found in the technical annex.

Results

The Results

We will show the results here for both council areas together. We will discuss the results for the first intervention and the second intervention separately. The results will be broken down by the type of waste: landfill; garden and food; and dry recycling. Full details of the analysis can be found in the Technical Annex.

Overall, the results show that there are significant changes in some categories of waste pre and post intervention but we were not able to detect a significant difference between the intervention and control group.

First intervention - Landfill waste

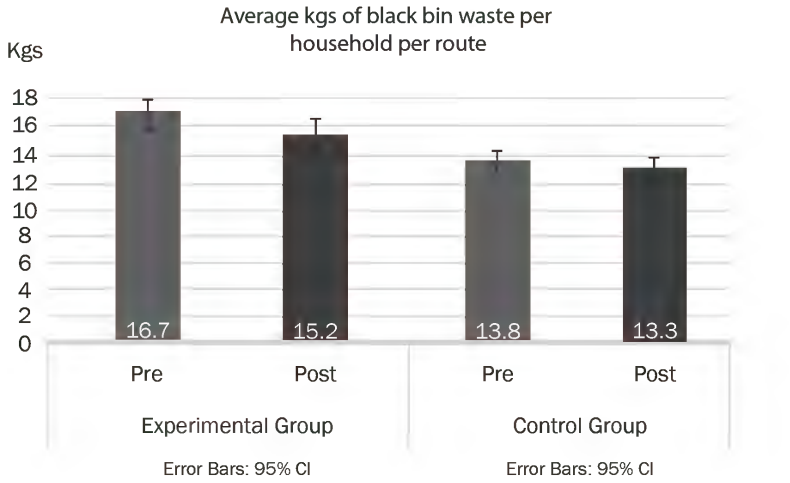
There was a decrease in the weight of the landfill bin in both groups (see Table 2).

Table 2: Change in weight of landfill waste

Group	Change in weight of waste (kg/household per route; pre/post intervention)	Change
Door hanger	-1.5kg	Statistically significant change
Control	-0.5kg	Non-significant change

We then compared the results for the two groups to see if the decrease in the door hanger group was statistically greater than the decrease in the control group. We do this statistical comparison to see if the difference in the decrease of weight was a 'real' difference or could be due to chance. When we made this comparison, we found that there was no statistically significant difference between the two groups. This means that although the door hanger group looks like it decreased more than the control group, we are unable to tell if the difference is due to chance, if the intervention had no effect, or because the trial did not have enough power to detect this size of effect.

Chart 1: Average black bin waste per household per route



First intervention - Garden and food waste

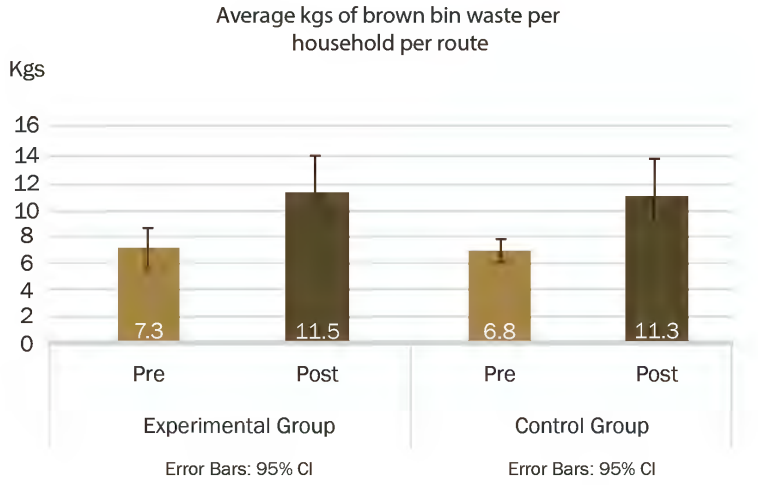
There was an increase in the weight of the brown bin in both the door hanger and the control group (see Table 3). This increase is likely due to the fact that the period of measurement was from January to May, and there is a significant increase in the production of garden waste over this period.

Table 3: Change in weight of garden and food waste

Group	Change in weight of waste (kg/household per route; pre/post intervention)	Change
Door hanger	4.2kg	Statistically significant change
Control	5.5kg	Statistically significant change

As in the black bin example, however, we did not find a statistically significant difference between the increases of both groups. In other words, though the control group appears to have a greater increase of garden and food waste over the period, we are unable to tell if the difference is due to chance, if the intervention had no effect, or because the trial did not have enough power to detect this size of effect.

Chart 2: Average brown bin waste per household per route



First intervention - Dry recycling

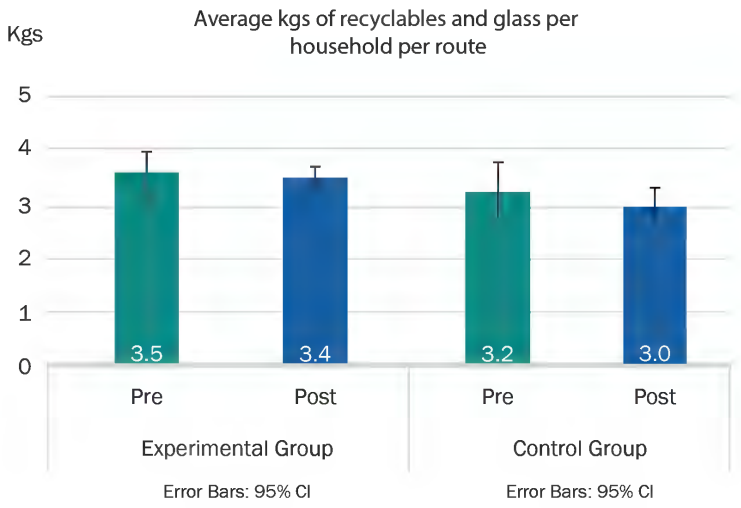
There was a decrease in the amount of dry recycling waste that was collected in both groups (see Table 4).

Table 4: Change in weight of dry recycling waste

Group	Change in weight of waste (kg/household per route; pre/post intervention)	Change
Door hanger	-0.1kg	Non-significant change
Control	-0.2kg	Non-significant change

As in the black bin and garden waste examples, however, there was no statistically significant difference between the decreases of both groups. In other words, though the control group appears to have decreased dry recycling slightly more over the period, we are unable to tell if the difference is due to chance, if the intervention had no effect, or because the trial did not have enough power to detect this size of effect.

Chart 3: Average recyclables and glass per household per route



Second Intervention

We also looked to see if the second intervention had an effect on recycling behaviour. We had hypothesised that a second intervention with a reinforcing message would, at a minimum, solidify an effect. We measured and compared the waste collected for 4 weeks after the first intervention and before the second intervention and for 6 weeks after the second intervention.

Second Intervention - Landfill waste

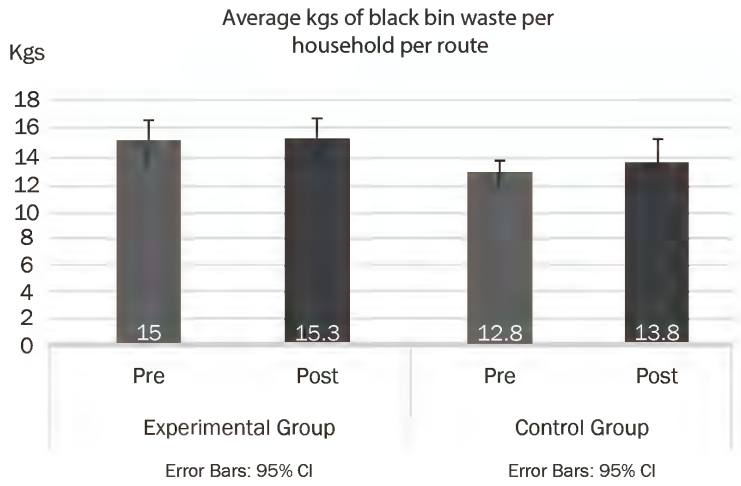
There was slight increase in the amount of waste collected in the landfill bin after the second intervention.

Table 5: Change in weight of landfill waste (2nd Intervention)

Group	Change in weight of waste (kg/household per route; pre/post intervention)	Change
Door hanger	0.193kg	Non-significant change
Control	0.997kg	Non-significant change

There was no statistically significant difference between the increases of both groups. In other words, though the control group appears to have increased landfill waste slightly more over the period, we are unable to tell if the difference is due to chance, if the intervention had no effect, or because the trial did not have enough power to detect this size of effect.

Chart 4: Average black bin waste per household per route (2nd Intervention)



Second Intervention - Garden and food waste

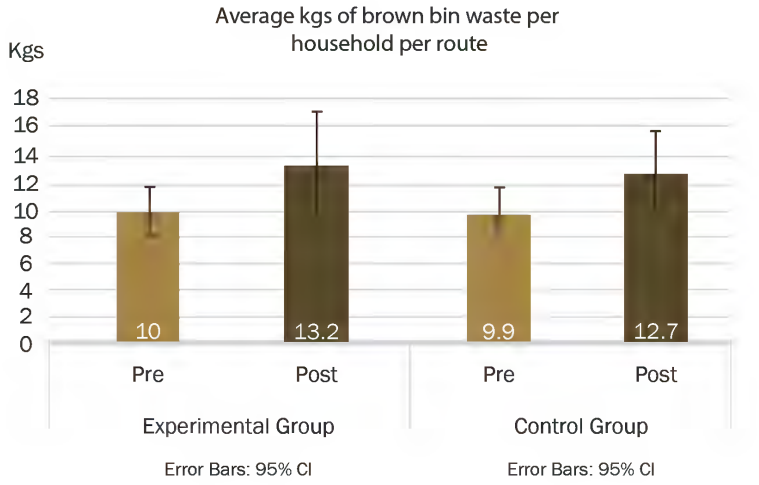
There was slight increase in the amount of waste collected in the landfill bin after the second intervention.

Table 6: Change in weight of garden and food waste (2nd Intervention)

Group	Change in weight of waste (kg/household per route; pre/post intervention)	Change
Door hanger	3.205kg	Statistically significant change
Control	2.748kg	Statistically significant change

While there were significant increases in the amount of garden and food waste collected, there was no statistically significant difference between the increases of both groups. In other words, though the intervention group appears to have a slightly greater increase of garden and food waste slightly more over the period, we are unable to tell if the difference is due to chance, if the intervention had no effect, or because the trial did not have enough power to detect this size of effect.

Chart 5: Average garden and food per household per route (2nd Intervention)



Second Intervention - Dry recycling waste

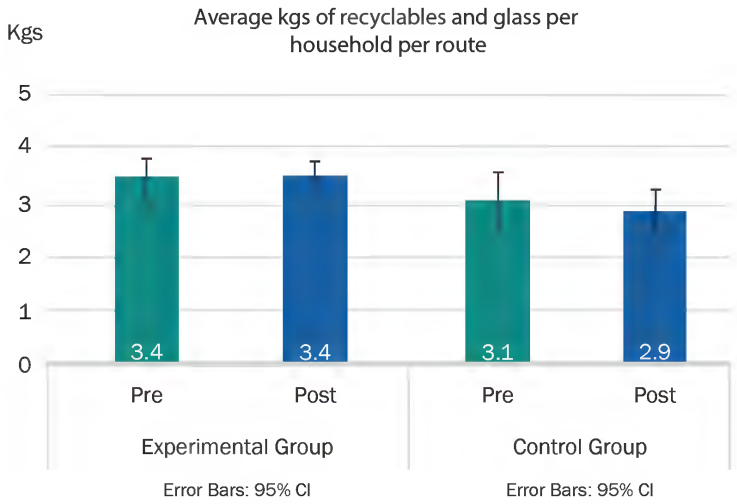
There was slight increase in the amount of waste collected in the landfill bin after the second intervention.

Table 7: Change in weight of dry recycling waste (2nd Intervention)

Group	Change in weight of waste (kg/household per route; pre/post intervention)	Change
Door hanger	-0.007kg	Statistically significant change
Control	-0.199kg	Statistically significant change

There was no statistically significant difference between the decreases of both groups. In other words, though the control group appears to have decreased dry recycling by slightly more over the period, we are unable to tell if the difference is due to chance, if the intervention had no effect, or because the trial did not have enough power to detect this size of effect.

Chart 6: Average recyclables and glass per household per route (2nd Intervention)



Photograph by Charles Deluvio on Unsplash

Conclusions and Recommendations

There are significant changes in recycling and waste collection in both the control and intervention conditions but these are not significantly different from each other. Why might this be the case?

- The intervention did not produce behaviour change
- The change which the intervention prompted was too small to be detected by our trial

[Technical note: Because of our small sample size (18 routes in total), the trials that we ran had low power. This means that the effect of the intervention would have had to have been very large for us to have detected it. Typically, we would expect a small to medium effect of an intervention like this one, so it is possible that an effect was there but our sample size wasn't large enough to detect it. See Technical Annex]

Either way, we cannot conclude that the intervention increased recycling behaviours.

In addition to these specific conclusions from the trial and the data, we can make some broader conclusions about the policy area and future development work.

Signal triggers not sufficient for this policy problem at this stage

One of the insights from waste managers was that it was easier to get already activated people to recycle more than to try and convert non-activated people to change behaviour. This accords with the behavioural theory – it is easier to get someone to do more of something that they are already doing than it is to get someone to do something completely new. The intervention that we designed was targeted at those people and, based on the analysis of other interventions that were primarily instructional, we used a motivational behavioural change component.

The fact that our intervention did not produce a significant enough change to be detected by our trial would suggest that these simple signals are not sufficient for the challenge at this stage. It may be more appropriate to focus on understand how to create changes in people who are not as activated.

Importance of comparison

While the case for using Random Control Trials (RCT) in government is rapidly advancing, this study is another example of why using comparison groups is important, even in circumstances where RCTs are not possible or not easily run. There were significant effects detected from the intervention when these results are looked at by themselves. However, these effects are not significantly different from what the control group experienced over the same time period. If we not had a comparison group, we could have made a recommendation to roll out this intervention across Northern Ireland which would have cost the councils time and money; and yet, because we included a control group, we can conclude that the intervention's effect was no better than chance, and we can experiment again to see what might work better next time.

Power of experiment and size of effect

Before we started, we had no clear idea what size of an effect we could expect to see from an intervention. We designed our experiment based on the practicalities of data collection, the budget available to us, and the practicalities of delivery. We also didn't know what the cost effectiveness calculations – essentially, how much is it worth to try and find a certain size of effect. We now know more about these issues. Using this methodology, we would need to increase the number of routes in each condition, or collect household data, in order to detect smaller differences between groups, if those differences actually exist. That obviously increases the cost of a trial and so it may become more cost effective to develop bespoke trial designs and data collection exercises.

We now know more about what effect sizes might be possible and future trials can start from this point and power trials appropriately.

Limitations on intervention design

A key constraint of this trial was the data. The type of data which we were able to obtain constrained the design of the intervention, the deployment of the intervention, and the analysis of the effectiveness of the intervention. While all of this is a natural consequence of working with applied solutions and administrative data, it does limit the ability to do small scale experiments which test new ideas and concepts. Other researchers have developed unusual ways of collecting data on recycling and this issue may also require innovative solutions if experimental methods are to be used on this challenge.

Process of experimentation

Innovation is a process, not a one-off experiment, or an event. Our motivational based signal trigger did not increase recycling or decrease landfill waste. What about an ability based trigger? Do we need to test whether the delivery mechanism is effective? Do we need to design more complex interventions? Are we sure about our behavioural analysis?

Here are some other types of experiments which could be run to test the behavioural theory outlined above:

- What is the most effective delivery mechanism of messages? We tested whether a theoretically informed door hanger was better than nothing but it may be useful to test different types of delivery mechanisms.
- What type of trigger is most effective at a population level? We tested whether a signal was better than nothing but it may be useful to test the effectiveness of a facilitator or a spark.
- What type of trigger is most effective for changing the behaviour of a specific behavioural segment (see Figure 2)? We tested an intervention at the population level and we were not able to tell if there were significant behavioural change in the individual segments that we identified in Figure 2. It may be useful to identify segments and trial tailored interventions with those segments.

Annex

Annex 1: Technical Annex

Data processing

Data could not be provided at a household level, so data was provided by the councils as an aggregate of the total weight (in tonnes) of each type of waste collected per route per week. Councils also provided data on the total number of houses per route. Each route consisted of around 900-1000 houses. We had no further detail on the number of occupants or type of housing of each route.

To standardize the data across routes, the data were converted to the average weight per household per route, which was calculated by multiplying the weight in tonnes by 1000 to get kilograms, and then dividing the total kilograms of waste by the number of houses for each route.

Example: On Jan 3, Route 1 collected 14.16 tonnes of black bin waste on week 1. Route 1 consisted of 990 houses. Average kilos per household for Jan 3 = (14.16*1000)/990 = 14.3.

Typically, landfill waste and recycling waste were collected on alternate weeks. These two weeks were collapsed into one data point. Therefore, data for each route and each type of waste were collected at a total of 7 timepoints (13 alternate weeks) For Antrim and Newtownabbey and 5 timepoints (10 alternate weeks) for Antrim, Banbridge, and Craigavon (ABC) before the intervention and 4-5 (9 alternate weeks) times after the intervention for both.

In Antrim and Newtownabbey, there were three types of waste collected: Black Bin: landfill waste, Brown Bin: garden/food recycling, and Triple Stack: a stack of three recycling bins where glass, paper and plastic products were collected together. In ABC, there were Black Bins: landfill waste, Brown bin: garden/food recycling, Glass collection, and Green Bin: paper/plastic/cardboard. For the analyses, we combined glass and green bin waste for ABC to create the 'Triple Stack' bins in Antrim and Newtonabbey so they could be combined and analysed.

In total, 9 routes received the intervention and 9 served as a control. There were 5 control/intervention routes each in Antrim and Newtownabbey and 4 control/intervention routes each in ABC.

Missing Data, Outliers, and Adjustments

One data point in April and one in May were missing in the black bin and Triple stack column in the intervention condition. These values were imputed into the database to be the average of the data points that came the week before and the week after.

After adjusting for missing values, any additional outliers in the difference between pre and post weights of waste/recycling were tested for by an inspection of two separate boxplots – one for the intervention group and one for the control group. One outlier was detected in the black bin outcome variable in the control group, three were detected in the Brown bin condition in the intervention group, and none were found in the Blue bin outcome. Outliers were not changed or removed.

Tests of normality of the data

The difference between pre and post weights of waste/recycling were normally distributed in all conditions ($p>0.05$) except for Control Group: Black Bin ($p=.024$) as assessed by the Shapiro-Wilk test.

Analytic Procedure

Data for each route was averaged to get one value for each type of waste collected before and after the intervention. For example, for Route 1, black bin waste data was gathered 7 times before the intervention, and on alternate weeks, recycling data was gathered. These 7 values for black bin waste were averaged together for one 'Pre-Intervention' value for Route 1. The same was done to the values that occurred after the intervention to create one 'Post-intervention' value for that route. The values for each of the different types of recycling were also averaged together separately to get one 'Pre-intervention' and one 'Post-intervention' value for each type of waste for each route. There were 9 routes in the intervention condition and 9 in the control, which means there were 9 'Pre-intervention' data points and 9 'Post-intervention' data points for each type of waste in both the intervention and control groups. These pre and post values are small numbers, so the results of any analyses should be taken with caution.

These two data points were first compared with a Paired-Samples T-test for the control and intervention groups separately to see if there was a significant change across each condition after the intervention. A Difference in Differences regression model was run on the data to determine if the pre/post difference in the intervention group was statistically different than the pre/post difference of the control group.[7] For example, even though we might see a significant decrease in landfill waste in the intervention condition, this is not interesting to us unless that decrease is different from the decrease the control group also experienced.

The Regression model[8]:

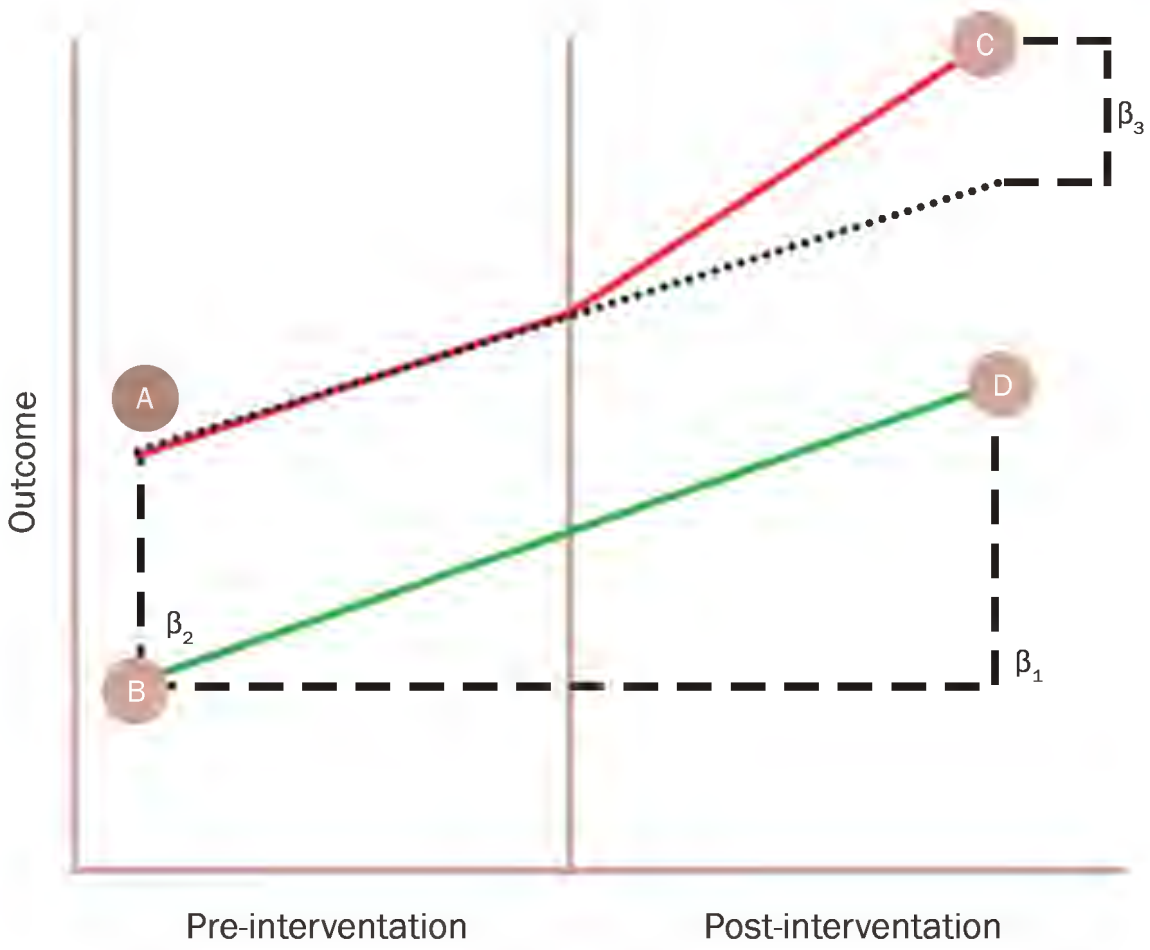
$Y = \beta 0 + \beta 1*[Time] + \beta 2*[Intervention] + \beta 3*[Time*Intervention] + \varepsilon$

Coefficient	Calculation	Interpretation
β_0	B	Baseline average
β_1	D-B	Time trend in control group
β_2	A-B	Difference between two groups pre-intervention
β_3	(C-A)-(D-B)	Difference in changes over time between the two groups

[7]Program Evaluation and the Difference in Difference Estimator. https://eml.berkeley.edu/~webfac/saez/e131_s04/diff.pdf

[8]Difference-in-Differences Estimation https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation

Figure 1: Model of difference in differences



Data were analysed in SPSS vs. 24.

Results

First Intervention: Waste collected between 10-13 weeks pre-intervention and 4-5 weeks post-intervention.

Black Bins: Landfill Waste

Both the intervention and control group showed a statistically significant decrease in the amount of landfill waste over the period. The decreases, however, were not statistically different from each other. The intervention did not decrease landfill waste any more than the control group decreased their landfill waste over the same period.

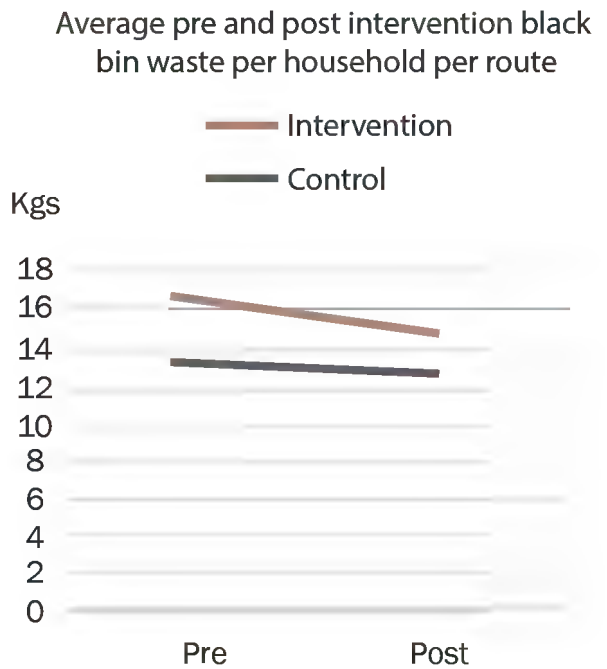
Table 1: Difference in differences estimator

Black Bin: Landfill Waste	Difference in the change between intervention and control group (β)	Standard Error	Significance Level (p value)	Confidence Interval – Lower Bound	Confidence Interval – Upper Bound
Difference in differences	-.854	1.255	.501	-3.411	1.703

Table 2: Paired T-test results

Black Bin: Landfill Waste	Pre-Intervention Time period: mean kilos per household per route	Standard Deviation	Post-Intervention Time period: mean kilos per household per route	Standard Deviation	Difference between pre and post-intervention mean	Significance Level (P value)
Intervention Group	16.674	1.829	15.199	2.323	-1.475	.007*
Control Group	13.882	1.779	13.263	1.508	-.620	-.620

*statistically significant p<0.05



Brown Bins: Garden and Food waste

Both the intervention and the control group showed a statistically significant increase in garden and food waste over the period. This is unsurprising as the period ran from January to May, coming into the spring when more garden waste is produced. The increases, however, were not significantly different from each other. The intervention group did not increase brown bin recycling any more than the control group did over the same period.

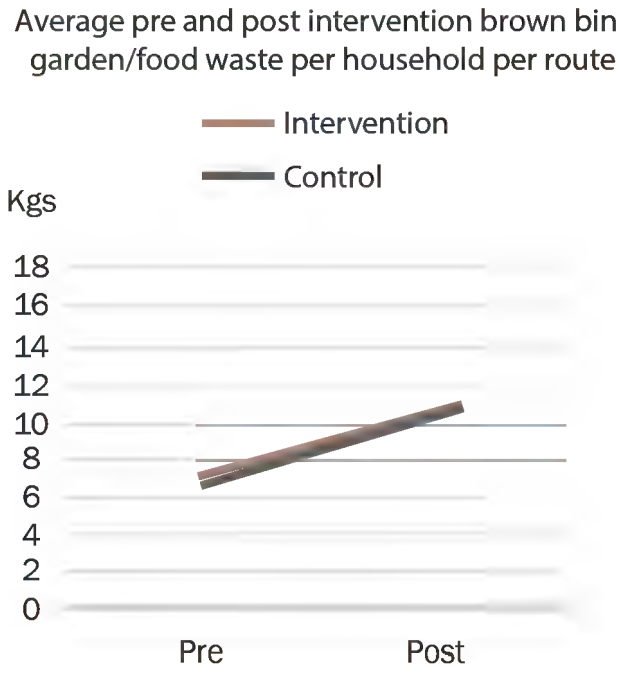
Table 3: Difference in differences estimator

Brown Bin: Garden and food waste	Difference in the change between intervention and control group (β)	Standard Error	Significance Level (p value)	Confidence Interval – Lower Bound	Confidence Interval – Upper Bound
Difference in differences	-.271	2.098	.898	-4.544	4.001

Table 4: Paired T-test results

Brown bin: Garden and food waste	Pre-Intervention Time period: mean kilos per household per route	Standard Deviation	Post-Intervention Time period: mean kilos per household per route	Standard Deviation	Difference between pre and post-intervention mean	Significance Level (P value)
Intervention Group	7.328	2.858	11.558	4.065	4.230	P<.0001*
Control Group	6.842	1.583	11.342	3.521	4.50	.002*

*statistically significant p<0.05



Glass/Plastic/Paper Recycling

Both the intervention and the control conditions showed reductions in the amounts of recycling in glass/plastic/paper recycling over the period, though the intervention group did not reduce as much as the control group. However, the reductions were, statistically, no different for the intervention than for the control group.

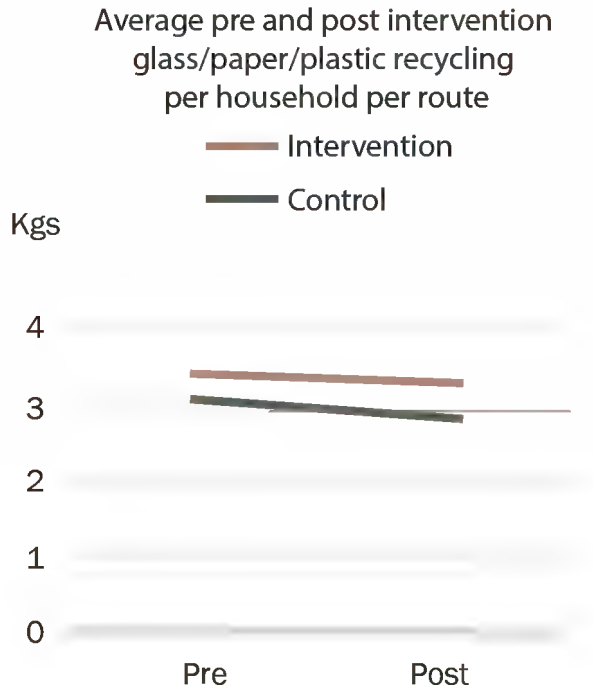
Table 5: Difference in differences estimator

Glass, paper, plastics	Difference in the change between intervention and control group (β)	Standard Error	Significance Level (p value)	Confidence Interval – Lower Bound	Confidence Interval – Upper Bound
Difference in differences	-.854	1.255	.501	-3.411	1.703

Table 6: Paired T-test results

Glass, paper, plastics	Pre-Intervention Time period: mean kilos per household per route	Standard Deviation	Post-Intervention Time period: mean kilos per household per route	Standard Deviation	Difference between pre and post-intervention mean	Significance Level (P value)
Intervention Group	3.502	.548	3.389	.403	-.113	.122
Control Group	3.172	.680	2.991	.637	-.182	.025*

*statistically significant p<0.05



Results

Second Intervention: Waste collected for 4 weeks after the first intervention and before the second intervention and again for 6 weeks after the second intervention.

Black Bins: Landfill Waste

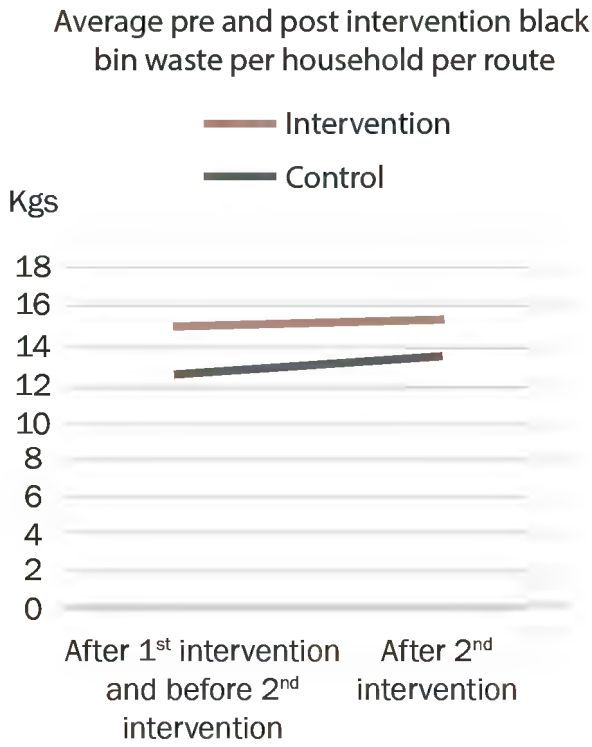
For both the Intervention and control Group, there was a slight increase in the amount of waste thrown away in the black bin (landfill bin) after the second intervention. The increase in the intervention group was not significantly different from the increase in the control group.

Table 7: Difference in differences estimator

Black Bin: Landfill Waste	Difference in the change between intervention and control group (β)	Standard Error	Significance Level (p value)	Confidence Interval – Lower Bound	Confidence Interval – Upper Bound
Difference in differences	-.804	1.502	.596	-3.864	2.256

Table 8: Paired T-test results

Black Bin: Landfill Waste	Pre-Intervention Time period: mean kilos per household per route	Standard Deviation	Post 2nd Intervention Time period: mean kilos per household per route	Standard Deviation	Difference between pre and post-intervention mean	Significance Level (P value)
Intervention Group	15.098	2.828	15.291	2.246	.193	.831
Control Group	12.764	1.754	13.761	2.047	.997	.236



Brown Bins: Garden and Food waste

Both the intervention and the control group showed a statistically significant increase in garden and food waste after the first intervention compared to after the second intervention. The increases, however, were not significantly different from each other. The intervention group did not increase brown bin recycling any more than the control group.

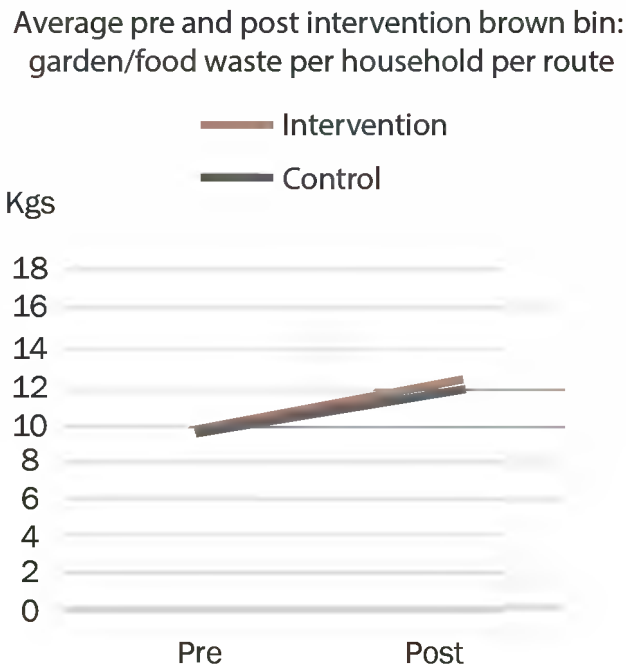
Table 9: Difference in differences estimator

Brown Bin: Garden and food waste	Difference in the change between intervention and control group (β)	Standard Error	Significance Level (p value)	Confidence Interval – Lower Bound	Confidence Interval – Upper Bound
Difference in differences	.458	2.916	.876	-5.481	6.397

Table 10: Paired T-test results

Brown bin: Garden and food waste	Post 1st Intervention, Pre 2nd Intervention: mean kilos per household per route	Standard Deviation	Post 2nd Intervention Time period: mean kilos per household per route	Standard Deviation	Difference between pre and post-intervention mean	Significance Level (P value)
Intervention Group	10.044	3.136	13.249	6.165	3.205	.050*
Control Group	9.984	2.950	12.732	4.467	2.748	.015*

*statistically significant p<0.05



Glass/Plastic/Paper Recycling

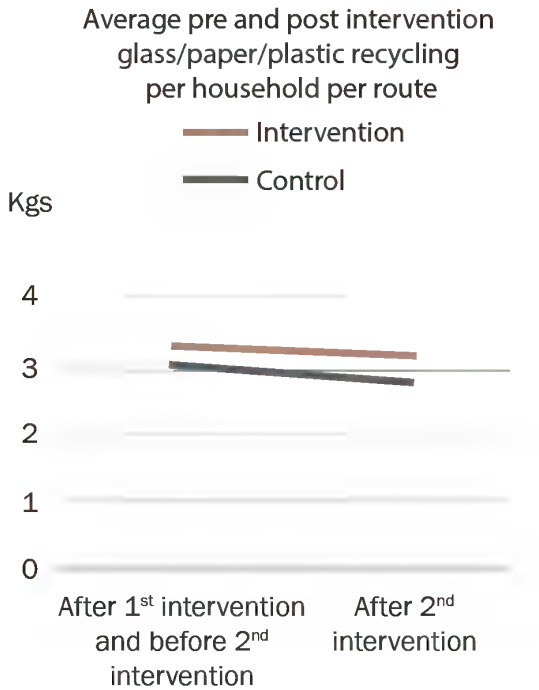
Both the intervention and the control conditions showed reductions in the amounts of recycling in glass/plastic/paper recycling over the period, though the intervention group did not reduce as much as the control group. However, the reductions were, statistically, no different for the intervention than for the control group.

Table 11: Difference in differences estimator

Glass, paper, plastics	Difference in the change between intervention and control group (β)	Standard Error	Significance Level (p value)	Confidence Interval – Lower Bound	Confidence Interval – Upper Bound
Difference in differences	.191	.362	.601	-.546	.929

Table 12: Paired T-test results

Glass, paper, plastics	Post 1st Intervention, Pre 2nd Intervention: mean kilos per household per route	Standard Deviation	Post 2nd Intervention Time period: mean kilos per household per route	Standard Deviation	Difference between pre and post-intervention mean	Significance Level (P value)
Intervention Group	3.396	.450	3.387	.389	-.007	.935
Control Group	3.172	.680	2.889	.656	-.199	.039



Power calculations

For black bin waste: From the regression:

Standardized Coefficient for the interaction term DiD (time x group): -.168

- This is considered a Medium effect size for the interaction.
- The Unstandarized Coefficient for this medium effect is -0.854 which means that, when you account for the control group's decrease after the intervention, the intervention resulted in 0.9 kilos less black bin waste per household averaged across 5 collection points (10 weeks)

Sample size: 18 (9 control, 9 experimental group)

GPower[9] was used to determine actual power achieved:

F tests - Linear multiple regression: Fixed model, R^2 increase

Analysis:	Post hoc: Compute achieved power		
Input:	Effect size f^2	=	.16
	α err prob	=	0.05
	Total sample size	=	18
	Number of tested predictors	=	1
	Total number of predictors	=	3
Output:	Noncentrality parameter λ	=	2.8800000
	Critical F	=	4.6001099
	Numerator df	=	1
	Denominator df	=	14
	Power (1-β err prob)	=	**0.3526022**

This indicates that there was only a 35% chance of detecting significant differences between the control and intervention group, even if they exist. There is, therefore, a 65% chance that the finding could have been significant, but we didn't have a large enough sample size to detect it.

To have an 80% chance of finding a medium effect, we would need to have at least 52 groups or individuals.

F tests - Linear multiple regression: Fixed model, R^2 increase

Analysis:	A priori: Compute required sample size		
Input:	Effect size f^2	=	.16
	α err prob	=	0.05
	Power (1-β err prob)	=	.80
	Number of tested predictors	=	1
	Total number of predictors	=	3
Output:	Noncentrality parameter λ	=	8.3200000
	Critical F	=	4.0426521
	Numerator df	=	1
	Denominator df	=	48
	Total sample size	=	**52**
	Actual power	=	0.8068454

[9] http://core.ecu.edu/psyc/wuenschk/MV/multReg/GPower-R2Change.pdf

For brown bin waste: From the regression:

Standardized Coefficient for the interaction term DiD (time x group): -0.032

- This is considered a small effect size for the interaction.
- The Unstandarized Coefficient for this small effect is -.271 which means that, when you account for the control group's decrease after the intervention, the intervention resulted in 0.3 kilos less brown bin waste per household averaged across 5 collection points (10 weeks)

Sample size: 18 (9 control, 9 experimental group)

GPower[10] was used to determine actual power achieved:

F tests - Linear multiple regression: Fixed model, R^2 increase

Analysis:	Post hoc: Compute achieved power		
Input:	Effect size f^2	=	.032
	α err prob	=	0.05
	Total sample size	=	18
	Number of tested predictors	=	1
	Total number of predictors	=	3
Output:	Noncentrality parameter λ	=	0.5760000
	Critical F	=	4.6001099
	Numerator df	=	1
	Denominator df	=	14
	Power (1-β err prob)	=	**0.1091162**

This indicates that there was only an 11% chance of detecting significant differences between the control and intervention group, even if they exist. There is, therefore, an 89% chance that the finding could have been significant, but, because our sample size wasn't large enough, we didn't have a large enough sample size to detect it.

To have an 80% chance of finding a small effect, we would need to have at least 248 groups or individuals.

F tests - Linear multiple regression: Fixed model, R^2 increase

Analysis:	A priori: Compute required sample size		
Input:	Effect size f^2	=	.032
	α err prob	=	0.05
	Power (1-β err prob)	=	.80
	Number of tested predictors	=	1
	Total number of predictors	=	3
Output:	Noncentrality parameter λ	=	7.9360000
	Critical F	=	3.8798520
	Numerator df	=	1
	Denominator df	=	244
	Total sample size	=	**248**
	Actual power	=	0.8012308

[10] http://core.ecu.edu/psyc/wuenschk/MV/multReg/GPower-R2Change.pdf

For Recyclables: From the regression:

Standardized Coefficient for the interaction term DiD (time x group): 0.051

- This is considered a small effect size for the interaction.
- The Unstandarized Coefficient for this small effect is 0.069 which means that, when you account for the control group's decrease after the intervention, the intervention resulted in 0.07 kilos more recyclables per household averaged across 5 collection points (10 weeks)

Sample size: 18 (9 control, 9 experimental group)

GPower[11] was used to determine actual power achieved:

F tests - Linear multiple regression: Fixed model, R^2 increase

Analysis:	Post hoc: Compute achieved power		
Input:	Effect size f^2	=	.051
	α err prob	=	0.05
	Total sample size	=	18
	Number of tested predictors	=	1
	Total number of predictors	=	3
Output:	Noncentrality parameter λ	=	0.9180000
	Critical F	=	4.6001099
	Numerator df	=	1
	Denominator df	=	14
	Power (1-β err prob)	**=**	**0.1453017**

This indicates that there was only a 15% chance of detecting significant differences between the control and intervention group, even if they exist. There is, therefore, an 85% chance that the finding could have been significant, but, because our sample size wasn't large enough, we didn't have a large enough sample size to detect it.

To have an 80% chance of finding a small effect, we would need to have at least 156 groups or individuals.

F tests - Linear multiple regression: Fixed model, R^2 increase

Analysis:	A priori: Compute required sample size		
Input:	Effect size f^2	=	.051
	α err prob	=	0.05
	Power (1-β err prob)	=	.80
	Number of tested predictors	=	1
	Total number of predictors	=	3
Output:	Noncentrality parameter λ	=	7.9560000
	Critical F	=	3.9033665
	Numerator df	=	1
	Denominator df	=	152
	Total sample size	**=**	**156**
	Actual power	=	0.8003337

[11] http://core.ecu.edu/psyc/wuenschk/MV/multReg/GPower-R2Change.pdf

Annex 2: Armagh, Banbridge and Craigavon Trial Protocol

ARMAGH, BANBRIDGE AND CRAIGAVON BOROUGH COUNCIL & INNOVATION LAB
INCREASE DOMESTIC RECYCLING TRIAL PROTOCOL

A field experiment to determine if there is a positive effect on levels of domestic kerbside recycling when householders receive a 'nudge' intervention to thank them for their recycling efforts compared to those that do not receive an intervention.

Version 1.1, 31 January 2017

SUMMARY

Key messages

- A quasi experimental design (QED) trial will be used to determine if there is an increase in the recycling rates of households that receive an intervention compared to those that don't.
- Participants will be allocated to either a treatment group (two nudges - initial nudge followed by a reminder nudge) or a control group (no nudge received).
- Outcomes will be determined from route level data obtained from council bin collection services.
- Statistical software will then be used to analyse the data in order to establish if the nudge interventions have had a positive impact on household recycling rates.

Strengths and limitations of this trial

- A QED is the research methodology that we have decided to use for this trial. A QED is a robust experimental design. It has been chosen primarily because a full randomised controlled trial is not possible. This is due to the fact that only route level data is available and therefore due to insufficient sample sizes only non-random allocation of participants can be used.
- Findings will determine whether the intervention has provided a statistically significant increase in recycling rate or no increase has been detected.
- Main limitations of the study are the cost of engaging an intervention group and the availability and integrity of data obtained from local councils household kerbside collection services during the field experiment.

1. BACKGROUND

During 2014/15 the household preparing for reuse, dry recycling and composting rate was 42%. This fell short of the Northern Ireland Programme for Government and Northern Ireland Waste Management Strategy targets of 45% by 2015 and 50% by 2020. There is a clear need for corrective action at a local council level in order to encourage more domestic recycling if Northern Ireland is to achieve future waste targets.

With this in mind the Innovation Lab held a workshop in March 2016 to bring together a range of people to work with Councils to hear about user insights, experience in other parts of the UK and academic research. After clearly defining the problem areas, the workshop participants generated ideas for improvement. These ideas were prioritised and the Priority One ideas were explored further in a subsequent Behavioural Insights workshop.

During the course of the Behavioural Insights workshop it was decided that some of the new ideas generated could be tested by the Innovation Lab team by way of an experimental trial to determine which were most effective. Those ideas that were proven to be effective could then be rolled out on a larger scale across NI and thus make a positive contribution towards achieving Northern Ireland's waste targets.

2. TRIAL OUTCOMES

2.1 Primary outcomes

- To determine whether a statistically significant increase can be detected in the recycling rates of households that receive an intervention compared to those households that do not receive an intervention.

2.2 Secondary outcomes

- Establish which of the 'nudge' interventions, if any, is the most effective in increasing household recycling rate.
- Use the data produced by the trial and the cost of each intervention to conclude which intervention represents the greatest return on investment.

2.3 Outcome measures

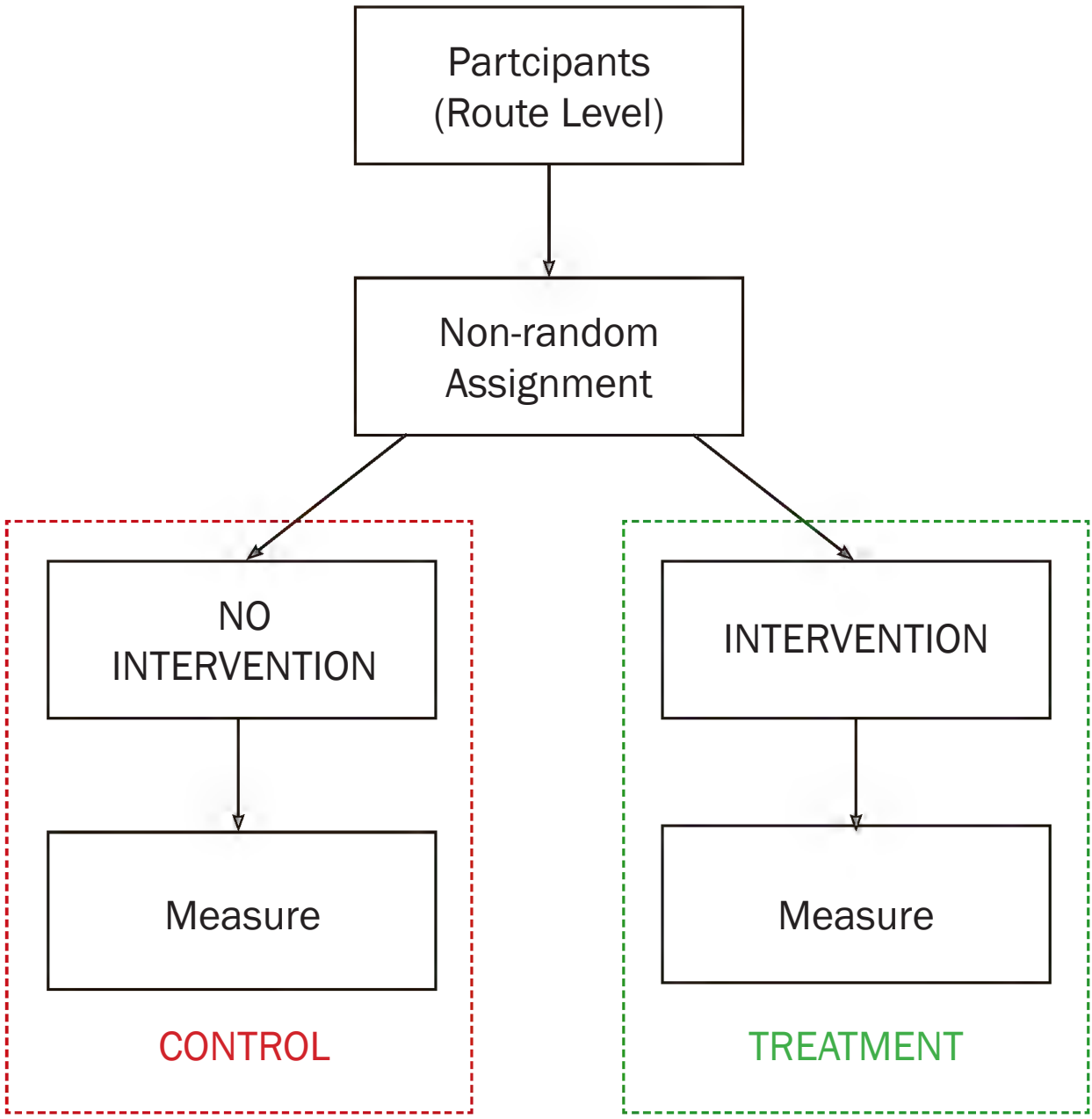
- A change in the average weight in kilograms per household of the total amount of residual waste/ dry recycling/food and compost waste for households will be the outcome measure. This will allow us to determine if there has been an increase in overall recycling and any transference between the different types of collection.

3. METHODOLOGY AND ANALYSIS

3.1 Design

The trial will be a 'Post Test Only' quasi experimental design with two trial arms that will use non-random assignment of participants. This means that one group (i.e. treatment group) will receive an intervention during the same time period that the other group (i.e. the control group) will receive no intervention.

Figure 1: General outline of the 'Post Test Only' QED



3.2 Allocation ratio

Eligible participants in each arm will be allocated as close as possible to a 1:1 ratio to the intervention groups or control group to receive one of the following:

- Two 'Thank you' nudges over an 8 week period (treatment group)
- No communication (control group)

3.3 Setting

Data will be collected in the field at a route level by the council. All types of bins will be collected and the collection vehicle weighed after the completion of each route to determine if there is an increase in recycling and/or transference from one type of collection to another.

3.4 Participant recruitment and eligibility

Active recruitment of participants will not be necessary as the data will be captured from the local council bin collection service as part of the normal collection process and does not require any personal input and/or interaction. For this reason written consent of participants will not be required as no personally identifiable information will be used. The only eligibility criteria for inclusion in the trial will be that the household is on the route identified by the local council as participating in the trial.

3.5 Intervention

There will be two 'nudges'. The first nudge will be deployed in the week prior to the beginning of the trial period. It will take the form of a door hanger which will contain a reaffirming message that we are doing well, as a community, at recycling and we want to maintain/improve our recycling rate even further.

The second nudge will follow in the fourth week of the trial period and will consist of a thank you for your efforts message to encourage households still further.

Hanger designs are included in Appendix B and there are some useful deployment 'Dos and Donts' available in Appendix C.

3.6 Data collection

Data collection method to produce the baseline data will involve obtaining historical household recycling data from local councils prior to commencement of the trial. This data will be sourced from the same population and cover a period of time that matches the trial period as closely as possible.

Outcome data will be collected from the bin collection service on a weekly basis during the trial period. The data will then be collated and quality checked then used for the analysis phase. It should be noted that the availability, reliability and validity of the council collection services data will be verified before any trial can be conducted.

3.7 Blinding

Both the Innovation Lab and council staff will know which households have been assigned to which particular treatment arm as they will not be blinded to the allocation process. The allocation procedures will be carried out in such a way that the control and treatment groups will be as similar as possible.

It will however, be difficult to determine to what extent participants will be blinded to the intervention. For example, those participants in treatment group will be aware that they are receiving an intervention. But, some may consider this intervention to be merely a generic message to all households and not realise that they are being specifically targeted. Whilst others may recognise that they are receiving a special treatment.

3.8 Trial size

The size of the trial is dependent upon reaching the largest number of households that is viable given the resources available for the trial. The council have indicated that they would be able include 8 routes in the trial. This means that 4 routes will be in the treatment arm of the trial and 4 in the control group. A total of 8,581 households will be participating broken down as:

N_1: 4,296 (Treatment Group) - routes 1, 6, 10, 13
N_0: 4,285 (Control Group) - routes 11, 16, 18, 24

3.9 Statistical analysis

The outcome data will be analysed using panel data regression model which mimics a difference in difference calculation.

3.10 Ethics

There are no ethical concerns regarding this trial since none of the participants in the control group are being denied a service. Conversely, none of the participants in the treatment group are being exposed to a new type of service as they are merely receiving an additional communication from their local council.

3.11 Timeline

The trial should be conducted during an 8 week period. Given the time to get the interventions agreed, designed and deployed it is likely the trial will begin in February 2017 and run for eight weeks. Exact deployment dates can be found in Appendix D.

We should also be cognisant of the fact that there is the potential for contamination if participants in either of the trial arms (treatment or control) are exposed to recycling messages in a disproportionate way. For this reason it is important to conduct the trial at a time when there are no significant marketing campaigns that would target trial participants in a disproportionate way.

3.12 Risks

All trials contain an element of risk and this trial is no exception. Therefore, in Appendix A below we have listed some of the major risks associated with this trial and listed the actions we have taken in order to mitigate their affects.

3.13 Quality

It is vital to the success of the trial that the deployment is carried out in line with this protocol. The deployment of the door hangers will be subject to a percentage check to ensure compliance with the instructions provided in this protocol.

4. CONTACT DETAILS

If you have any comments or suggestions regarding this trial protocol and/or deployment of the trial itself the please contact a member of the Innovation Lab team. Contact details are:

Pauline Wilson
E: pauline.wilson@finance-ni.go.uk
T: 028 9081 6171

Connor Scullion
E: connor.scullion@finance-ni.gov.uk
T: 028 9081 6167

APPENDIX A - RISKS

All projects contain an element of risk and a QED is no different. In Table A below there is a list of potential risks to the project and the actions taken to minimise those risks.

TABLE A - Risk Register

RISK	PROBABILITY	IMPACT	MITIGATION
Trial attrition	MEDIUM	MEDIUM	Communication with local councils explaining the principles of the trial and expectations. Attrition rates will be monitored and reported in the findings.
Interventions not dispatched correctly	LOW	MEDIUM	Communication with council staff explaining the principles of the trial and expectations.
Failure to recruit participants	LOW	HIGH	Work with local councils to determine if sufficient route level data is available for the trial.
Councils do not follow correct trial protocols	MEDIUM	HIGH	Conduct a test period (2 weeks) for data collection to ensure compliance with trial protocols.

APPENDIX B - DOOR HANGER DESIGNS

First Intervention

FRONT **BACK**

Second Intervention

APPENDIX C - HANGER DEPLOYMENT INSTRUCTIONS

One of the key elements of this trial is the successful deployment of the door hangers on the chosen routes. Below are some pointers on how to deploy the hangers correctly.

Do:

- Give every household a hanger.
- Push the hanger fully onto the door handle.
- Put hanger through letterbox if there is no handle.
- Report any deployment issues

Don't:

- Place the hanger on a bin.
- Place the hanger on a recycling box.
- Discard any remaining hangers.

APPENDIX D - HANGER DEPLOYMENT DATES

Hangers delivered to Council by 17 February 2017

	Trial Week No.		Wk1		Wk2		Wk3		Wk4		Wk5		Wk6		Wk7		Wk8	
	1st Hanger Deployed								2nd Hanger Deployed									
	w/c 20 Feb		w/c 27 Feb		w/c 6 Mar		w/c 13 Mar		w/c 20 Mar		w/c 27 Mar		w/c 3 Apr		w/c 10 Apr		w/c 17 Apr	
Zone	Zone A	Zone B	Zone A	Zone B	Zone A	Zone B	Zone A	Zone B	Zone A	Zone B	Zone A	Zone B	Zone A	Zone B	Zone A	Zone B	Zone A	Zone B
Collection Type	Mixed Dry	Black Bin	Black Bin	Mixed Dry	Mixed Dry	Black Bin	Black Bin	Mixed Dry	Mixed Dry	Black Bin	Black Bin	Mixed Dry	Mixed Dry	Black Bin	Black Bin	Mixed Dry	Mixed Dry	Black Bin
Treatment Group																		
Route 1	X		X		X		X		X		X		X		X		X	
Route 6	X		X		X		X		X		X		X		X		X	
Route 10	X		X		X		X		X		X		X		X		X	
Route 13	X		X		X		X		X		X		X		X		X	

X Black Bin

X Mixed Dry Recycling (green)Food/Garden (brown) Insert (glass)

Note: All treatment routes are in Zone A

Annex 3: Antrim and Newtonabbey Trial Protocol

ANTRIM and NEWTOWNABBEY BOROUGH COUNCIL & INNOVATION LAB
INCREASE DOMESTIC RECYCLING TRIAL PROTOCOL

A field experiment to determine if there is a positive effect on levels of domestic kerbside recycling when householders receive a 'nudge' intervention to thank them for their recycling efforts compared to those that do not receive an intervention.

Version 2.0, 22 February 2017

SUMMARY

Key messages

- A quasi experimental design (QED) trial will be used to determine if there is an increase in the recycling rates of households that receive an intervention compared to those that don't.
- Participants will be allocated to either a treatment group (two nudges - initial nudge followed by a reminder nudge) or a control group (no nudge received).
- Outcomes will be determined from route level data obtained from council bin collection services.
- Statistical software will then be used to analyse the data in order to establish if the nudge interventions have had a positive impact on household recycling rates.

Strengths and limitations of this trial

- A QED is the research methodology that we have decided to use for this trial. A QED is a robust experimental design. It has been chosen primarily because a full randomised controlled trial is not possible. This is due to the fact that only route level data is available and therefore sample sizes are insufficient to run a full randomised controlled trial.
- Findings will determine whether the intervention has provided a statistically significant increase in recycling rate or no increase has been detected.
- Main limitations of the study are the cost of engaging an intervention group and the availability and integrity of data obtained from local councils household kerbside collection services during the field experiment.

1. BACKGROUND

During 2014/15 the household preparing for reuse, dry recycling and composting rate was 42%. This fell short of the Northern Ireland Programme for Government and Northern Ireland Waste Management Strategy targets of 45% by 2015 and 50% by 2020. There is a clear need for corrective action at a local council level in order to encourage more domestic recycling if Northern Ireland is to achieve future waste targets.

With this in mind the Innovation Lab held a workshop in March 2016 to bring together a range of people to work with Councils to hear about user insights, experience in other parts of the UK and academic research. After clearly defining the problem areas, the workshop participants generated ideas for improvement. These ideas were prioritised and the Priority One ideas were explored further in a subsequent Behavioural Insights workshop.

During the course of the Behavioural Insights workshop it was decided that some of the new ideas generated could be tested by the Innovation Lab team by way of an experimental trial to determine which were most effective. Those ideas that were proven to be effective could then be rolled out on a larger scale across NI and thus make a positive contribution towards achieving Northern Ireland's waste targets.

2. TRIAL OUTCOMES

2.1 Primary outcomes

- To determine whether a statistically significant increase can be detected in the recycling rates of households that receive an intervention compared to those households that do not receive an intervention.

2.2 Secondary outcomes

- Establish which of the 'nudge' interventions, if any, is the most effective in increasing household recycling rate.
- Use the data produced by the trial and the cost of each intervention to conclude which intervention represents the greatest return on investment.

2.3 Outcome measures

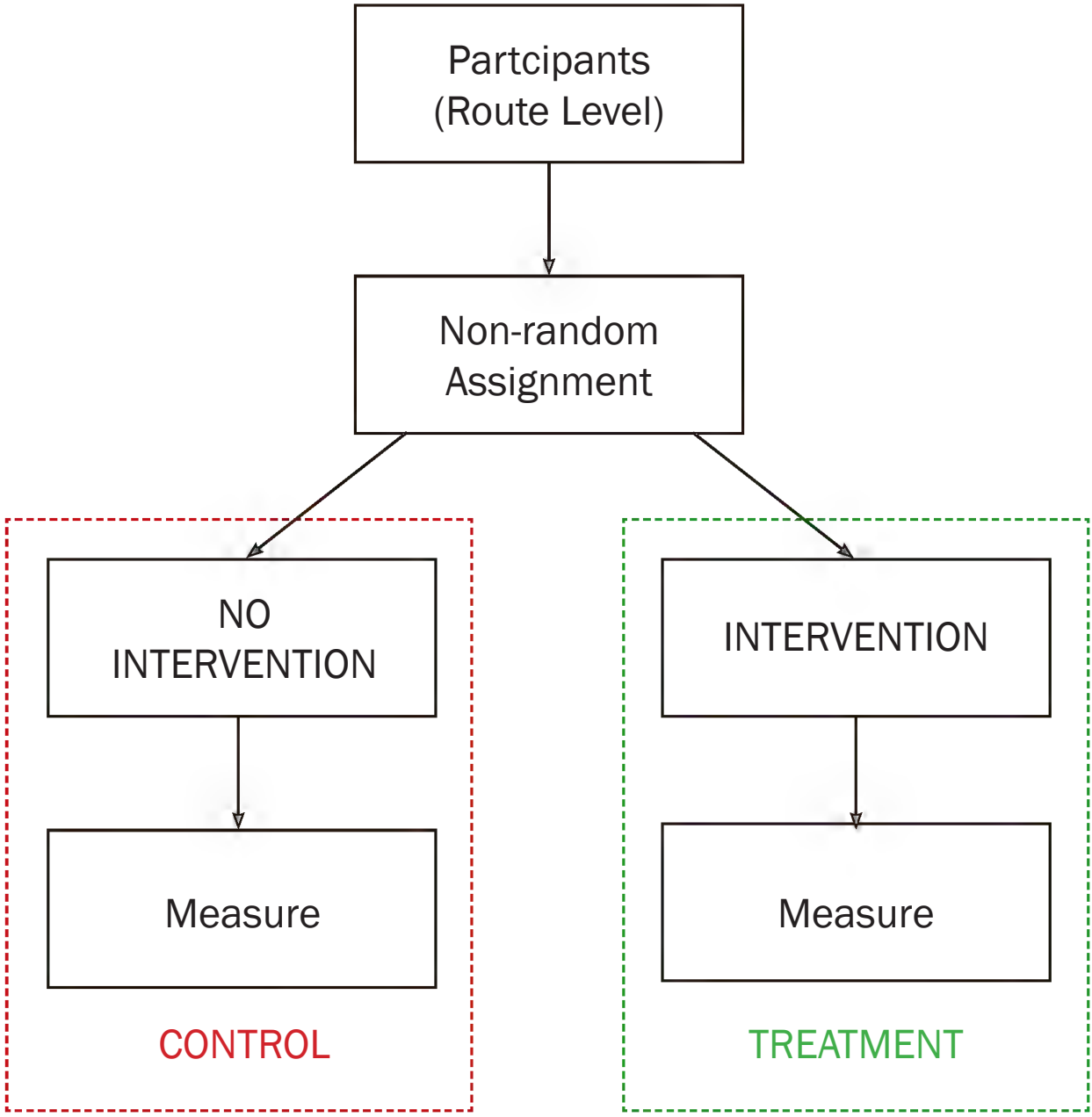
- A change in the average weight in kilograms per household of the total amount of residual waste/ dry recycling/food and compost waste for households will be the outcome measure. This will allow us to determine if there has been an increase in overall recycling and any transference between the different types of collection.

3. METHODOLOGY AND ANALYSIS

3.1 Design

The trial will be a 'Post Test Only' quasi experimental design with two trial arms that will use non random assignment of participants. This means that one group (i.e. treatment group) will receive an intervention during the same time period that the other group (i.e. the control group) will receive no intervention.

Figure 1: General outline of the 'Post Test Only' QED



3.2 Allocation ratio

Eligible participants in each arm will be allocated as close as possible to a 1:1 ratio to the intervention groups or control group to receive one of the following:

- Two 'Thank you' nudges over an 8 week period (treatment group)
- No communication (control group)

3.3 Setting

Data will be collected in the field at a route level by the council. All types of bins will be collected and the collection vehicle weighed after the completion of each route to determine if there is an increase in recycling and/or transference from one type of collection to another.

3.4 Participant recruitment and eligibility

Active recruitment of participants will not be necessary as the data will be captured from the local council bin collection service as part of the normal collection process and does not require any personal input and/or interaction. For this reason written consent of participants will not be required as no personally identifiable information will be used. The only eligibility criteria for inclusion in the trial will be that the household is on the route identified by the local council as participating in the trial.

3.5 Intervention

There will be two 'nudges'. The first nudge will be deployed in the week prior to the beginning of the trial period. It will take the form of a door hanger which will contain a reaffirming message that we are doing well, as a community, at recycling and we want to maintain/improve our recycling rate even further.

The second nudge will follow in the fourth week of the trial period and will consist of a thank you for your efforts message to encourage households still further.

Hanger designs are included in Appendix B and there are some useful deployment 'Dos and Donts' available in Appendix C.

3.6 Data collection

Data collection method to produce the baseline data will involve obtaining historical household recycling data from local councils prior to commencement of the trial. This data will be sourced from the same population and cover a period of time that matches the trial period as closely as possible.

Outcome data will be collected from the bin collection service on a weekly basis during the trial period. The data will then be collated and quality checked then used for the analysis phase. It should be noted that the availability, reliability and validity of the council collection services data will be verified before any trial can be conducted.

3.7 Blinding

Both the Innovation Lab and council staff will know which households have been assigned to which particular treatment arm as they will not be blinded to the allocation process. The allocation procedures will be carried out in such a way that the control and treatment groups will be as similar as possible.

It will however, be difficult to determine to what extent participants will be blinded to the intervention. For example, those participants in treatment group will be aware that they are receiving an intervention. But, some may consider this intervention to be merely a generic message to all households and not realise that they are being specifically targeted. Whilst others may recognise that they are receiving a special treatment.

3.8 Trial size

The size of the trial is dependent upon reaching the largest number of households that is viable given the resources available for the trial. The council have indicated that they would be able include 10 routes in the trial. This means that 5 routes will be in the treatment arm of the trial and 5 in the control group. A total of 8,914 households will be participating broken down as:

> N_1: 4,455 (Treatment Group) - routes Antrim (Tuesday C & IB, Tuesday D & JB, Wednesday D & JB) & Newtownabbey (The Brambles, Elmfield)
> N_0: 4,459 (Control Group) - routes Antrim (Thursday A & IA, Thursday B & JA) & Newtownabbey (Richmond, Mossley, Ballyclare)

3.9 Statistical analysis

The outcome data will be analysed using panel data regression model which mimics a difference in difference calculation.

3.10 Ethics

There are no ethical concerns regarding this trial since none of the participants in the control group are being denied a service. Conversely, none of the participants in the treatment group are being exposed to a new type of service as they are merely receiving an additional communication from their local council.

3.11 Timeline

The trial should be conducted during an 8 week period. Given the time to get the interventions agreed, designed and deployed it is likely the trial will begin in March 2017 and run for eight weeks. Exact deployment dates can be found in Appendix D.

We should also be cognisant of the fact that there is the potential for contamination if participants in either of the trial arms (treatment or control) are exposed to recycling messages in a disproportionate way. For this reason it is important to conduct the trial at a time when there are no significant marketing campaigns that would target trial participants in a disproportionate way.

3.12 Risks

All trials contain an element of risk and this trial is no exception. Therefore, in Appendix A below we have listed some of the major risks associated with this trial and listed the actions we have taken in order to mitigate their affects.

3.13 Quality

It is vital to the success of the trial that the deployment is carried out in line with this protocol. The deployment of the door hangers will be subject to a percentage check to ensure compliance with the instructions provided in this protocol.

4. CONTACT DETAILS

If you have any comments or suggestions regarding this trial protocol and/or deployment of the trial itself the please contact a member of the Innovation Lab team. Contact details are:

Pauline Wilson
E: pauline.wilson@finance-ni.go.uk
T: 028 9081 6171

Connor Scullion
E: connor.scullion@finance-ni.gov.uk
T: 028 9081 6167

APPENDIX A - RISKS

All projects contain an element of risk and a QED is no different. In Table A below there is a list of potential risks to the project and the actions taken to minimise those risks.

TABLE A - Risk Register

Trial attrition	MEDIUM	MEDIUM	Communication with local councils explaining the principles of the trial and expectations. Attrition rates will be monitored and reported in the findings.
Interventions not dispatched correctly	LOW	MEDIUM	Communication with council staff explaining the principles of the trial and expectations.
Failure to recruit participants	LOW	HIGH	Work with local councils to determine if sufficient route level data is available for the trial.
Councils do not follow correct trial protocols	MEDIUM	HIGH	Conduct a test period (2 weeks) for data collection to ensure compliance with trial protocols.

APPENDIX B - DOOR HANGER DESIGNS

First Intervention

FRONT **BACK**

Second Intervention

APPENDIX C - HANGER DEPLOYMENT INSTRUCTIONS

One of the key elements of this trial is the successful deployment of the door hangers on the chosen routes. Below are some pointers on how to deploy the hangers correctly.

Do:

- Give every household a hanger.
- Push the hanger fully onto the door handle.
- Put hanger through letterbox if there is no handle.
- Report any deployment issues

Don't:

- Place the hanger on a bin.
- Place the hanger on a recycling box.
- Discard any remaining hangers.

APPENDIX D - HANGER DEPLOYMENT DATES

Hangers delivered to Council by 10 March 2017

Table 1: Newtownabbey Routes

	Trial Week No.			Wk1			Wk2			Wk3			Wk4			Wk5			Wk6			Wk7			Wk8		
	1st Hanger Deployed												2nd Hanger Deployed														
	w/c 13 March			w/c 20 March			w/c 27 March			w/c 3 April			w/c 10 April			w/c 17 April			w/c 24 April			w/c 1 May			w/c 8 May		
Collection Type	Black	Brown	T stack	Black	Brown	T stack	Black	Brown	T stack	Black	Brown	T stack	Black	Brown	T stack	Black	Brown	T stack	Black	Brown	T stack	Black	Brown	T stack	Black	Brown	T stack
Treatment Group																											
The Brambles	X	X	X			X	X	X	X			X	X	X	X			X	X	X	X			X	X	X	X
Elmfield	X	X	X			X	X	X	X			X	X	X	X			X	X	X	X			X	X	X	X

X Black Bin

X Brown Bin

X Triple Stack

Hangers delivered to Council by 10 March 2017

Table 2: Antrim Routes

Trial Week No.	Wk1			Wk2			Wk3			Wk4			Wk5			Wk6			Wk7			Wk8		
	1st Hanger Deployed									2nd Hanger Deployed														
	w/c 13 March			w/c 20 March			w/c 27 March			w/c 3 April			w/c 10 April			w/c 17 April			w/c 24 April			w/c 1 May			w/c 8 May		
Collection Type	Black	Brown	Blue	Black	Brown	Blue	Black	Brown	Blue	Black	Brown	Blue	Black	Brown	Blue	Black	Brown	Blue	Black	Brown	Blue	Black	Brown	Blue	Black	Brown	Blue
Treatment Group																											
Tuesday C & IB		X	X	X				X	X	X				X	X	X				X	X	X				X	X
Tuesday D & JB		X	X	X				X	X	X				X	X	X				X	X	X				X	X
Wednesday D & JB		X	X	X				X	X	X				X	X	X				X	X	X				X	X

X Black Bin

X Brown Bin

X Blue Bin

Contact

Innovation Lab
Public Sector Reform Division
Department of Finance
Clare House
303 Airport Road West
Belfast
BT3 9ED

e-mail: ilab@finance-ni.gov.uk

follow us: @iLab_NI

search for: Innovation Lab NI

Cover Photograph by Denys Nevozhai on Unsplash