Using a difference-in-difference control trial to test an intervention aimed at increasing the take-up of a welfare payment in New Zealand

David Rea, Dean R. Hyslop

Observational Studies, Volume 9, Issue 4, 2023, pp. 49-72 (Article)

Published by University of Pennsylvania Press

DOI: https://doi.org/10.1353/obs.2023.a906626

For additional information about this article
https://muse.jhu.edu/article/906626

For content related to this article
https://muse.jhu.edu/related_content?type=article&id=906626
Using a difference-in-difference control trial to test an intervention aimed at increasing the take-up of a welfare payment in New Zealand

David Rea
david.rea001@msd.govt.nz
Principal Advisor, Ministry of Social Development
Wellington, New Zealand

Dean R. Hyslop
dean.hyslop@motu.org.nz
Senior Fellow, Motu Economic and Public Policy Research
Wellington, New Zealand

Abstract

This paper describes a difference-in-difference control trial (DDCT) of an intervention designed to increase the take-up of an income support payment in the New Zealand welfare system. The intervention used a microsimulation model to identify potential claimants who were then contacted by either phone, email, or letter. The trial was designed as a DDCT because of ethical concerns associated with a fully randomized approach. The trial provided convincing evidence that the intervention would increase the take-up of the payment and a modified version was then implemented as an ongoing business process by the New Zealand Ministry of Social Development (MSD). The findings from the trial contribute to the literature about how best to increase the take-up of welfare payments. The study also demonstrates the value of using a difference-in-difference control trial.

Keywords: Difference-in-difference, Field trial, Welfare, Take-up

1. Introduction

Low take-up of benefit payments is a problem for welfare systems in many countries. It is often the result of limited knowledge of entitlements bought about by the complexity of eligibility rules, time-consuming and complicated application and compliance procedures, and the stigma of receiving income support. Low take-up can increase poverty and hardship among the most vulnerable people, and if it relates to payments received by those employed it may undermine work incentives (Ko and Moffitt, 2022).

Many welfare agencies have work programs aimed at increasing the take-up of payments for those who are eligible (Ko and Moffitt, 2022). Typical initiatives include providing better information, proactive outreach, reducing application and compliance costs for claimants, making payments less stigmatizing, and providing more personalised service (Currie, 2004; Remler and Glied, 2003).

In this paper we document how we tested an intervention designed to increase the take-up of a payment called Temporary Additional Support (TAS) within the New Zealand welfare system. The intervention provided proactive outreach to individuals where administrative data suggested they would be eligible for TAS payments. A randomised control trial (RCT) would have been the usual method for evaluating the intervention, but this ap-
proach was considered inappropriate by stakeholders. Instead, we developed what we call a difference-in-difference control trial (DDCT) to assess the effectiveness of the intervention.

The trial showed that the intervention increased take-up and was subsequently implemented in a modified form as an ongoing business process by the Ministry of Social Development (MSD). In addition, the study also demonstrates the value of the DDCT, and we argue that the design could be used more widely in situations where an intervention needs to be prospectively tested, but an RCT is either not feasible or not appropriate.

The rest of the paper is organized as follows. In the next section we provide background on the development of the TAS trial. Section 3 outlines the nature and methods used in a difference-in-difference control trial. Section 4 describes the design and implementation of our trial, and section 5 presents the analysis and results. Section 6 discusses the subsequent implementation of the intervention as a permanent business process following the analysis of the trial results. The paper concludes in section 7 with some reflections about the value of using a difference-in-difference control trial.

2. Background on the development of the TAS trial

We begin by providing an overview of New Zealand’s welfare policy system and context for the Temporary Additional Support (TAS) payments. We then discuss the concern about take-up of TAS payment, and the development of a trial to encourage take-up.

Overview of the New Zealand welfare system and the TAS payment

The welfare system for the working age population in New Zealand consists of a comprehensive system of income-tested welfare payments, as well as a wider system of tax credits that are mostly targeted to low- and middle-income families with children (OECD, 2020). Welfare payments are administered nationally by MSD.

Welfare payments consist of both ‘main’ as well as ‘supplementary’ benefits. Main benefits provide a basic level of income support, and ‘supplementary benefits’ provide additional income related to specific additional costs. There are separate supplementary benefits related to housing, health, disability, the care of children, as well as general hardship. Most recipients receive both a main benefit and at least one supplementary benefit at the same time.

To receive a main benefit a claimant (and their partner where applicable) must have a low income, and they must meet categorical eligibility criteria such as not being able to find full-time work, caring for a dependent child as a sole parent, being unable to work temporarily or permanently because of ill health or a disability, or looking after another adult who would otherwise require hospital care.

Supplementary benefits have different eligibility criteria and, in some instances, can also be paid to a person who is not receiving a main benefit. Temporary Additional Support (TAS) is a supplementary benefit for people who experience financial hardship. To be eligible a person must have assets less than a specified amount, and their disposable income after accounting for specific allowable costs must also fall below a specified benchmark. The shortfall in disposable income is called the “TAS deficiency”, which then determines the value of the payment up to various maximum rates that differ according to recipient characteristics.
TAS is paid for 13 weeks but can be renewed after another application. In late 2018 20% of all welfare recipients received the payment. The average value of the payment at the time was $52 per week and represented around 10% of the total value of welfare payments for these recipients.

The problem of take-up of the TAS payment

For many years advocates on behalf of benefit recipients were concerned that the number of people receiving the TAS payment was considerably less than the total who seemed to be eligible. It was argued that the welfare system was not adequately protecting families from financial hardship and poverty because of low take-up of the payment. There was additional support for the concern about take-up from research using MSD’s microsimulation (MSIM) model. The modelling suggested that only 68% of eligible people were receiving the payment, although there was uncertainty about the estimate because of the limitations of the administrative data used in the model.¹

Analysis by the Ministry of Social Development suggested several potential reasons for incomplete take-up of TAS. These reasons included that it was complicated and difficult for potential claimants to understand, required substantial effort to apply, automatically expired after 13 weeks and required a subsequent reapplication if the claimant was still eligible, and had burdensome compliance requirements related to reporting changes in circumstances. In addition, barriers to accessing TAS payments were thought to be highest among vulnerable populations including those with compromised physical and mental health.

The development of a new approach to encourage take-up

Prior to 2018 MSD had sent letters inviting people to apply for TAS where there was an indication from administrative data that they might be eligible. This automated process stopped in June 2018 due to problems with legacy computer code that had used administrative data to identify people who might be eligible. As part of restarting the outreach campaign, MSD redesigned the approach for identifying and contacting people who might be eligible for the TAS payment. Under the new approach it was decided to use the MSIM model to identify the people who were likely eligible for the payment. It was also decided that as well as the previous approach of using letters, other means of contacting people such as emails and phone calls could also be considered.

While the approach seemed highly plausible, there was uncertainty about whether it would work, partly because of incomplete data used in the modelling. In addition, there was also uncertainty about which of the methods of proactive contact would be the most effective at encouraging individuals to apply for the payment. For these reasons, before fully implementing the new approach the Ministry decided to conduct a trial to assess if it was effective at increasing take-up of TAS.

¹ The MSD microsimulation model calculates all the payments that a person is eligible for using the eligibility rules of the welfare system. The population and calculation of eligibility is only for people who currently receive a payment from the Ministry. For each person the model uses administrative data relating to their current and past interactions with the welfare system. This includes information on demographic characteristics (age, partnership status, number of dependent children etc), costs, payments, and other income. For the most part the model uses only actual data and imputes a limited number of characteristics where the information is not available.
A randomized control trial was one option that could have been used to assess if the new approach would increase the number of people receiving the TAS payment. However, a simple design involving randomisation would have meant some individuals predicted to be experiencing severe financial hardship would have been allocated to the control group. This was not acceptable for some stakeholders. For this reason, we looked for another approach that was both simple to implement and able to provide credible evidence about the effectiveness of the intervention. Our solution was to use what we called a ‘difference-in-difference’ control trial (DDCT). The nature of this type of trial is not widely used or understood so in the following section we provide a brief overview of the approach.

3. Overview of methods for a difference-in-difference control trial

The defining feature of any trial is that researchers design, manage and analyse an ‘experiment’ to provide evidence about the impacts of an intervention.

Most trials are conducted using randomisation to assign participants to a treatment or control condition (Glennerster and Takavarasha, 2014). With a simple RCT participants are randomly assigned to either the treatment or control group. Identification of the causal impact of the intervention is possible because the expected future outcomes of the control group represent the counterfactual for the treatment group if the treatment had not been applied (Rubin, 1974).

Non-randomised trials are an alternative approach (Handley et al., 2018; Schmidt, 2017; West et al., 2008). A DDCT is one type of non-randomised trial and was originally described by Campbell and colleagues as an interrupted time series design with a non-equivalent control group (Campbell and Stanley, 1963; Cook and Campbell, 1979; Shadish et al., 2002). We use the term DDCT to highlight the use of modern difference-in-difference techniques for causal inference (Angrist and Pischke, 2009; Card and Krueger, 1994; Wing et al., 2018).

In contrast to an RCT, in a DDCT the assignment to control or treatment is deterministic rather than randomised. Assignment to control or treatment is determined by an observable characteristic for which there is a plausible assumption that there will be a constant difference in the expected future outcomes of the two groups in the absence of treatment. The counterfactual for the treatment group is assumed to be the average outcome for the control group, adjusted for the previously observed difference in average outcomes between the treatment and control groups.\(^2\)

The standard diagrammatic representation of how the difference-in-difference strategy identifies a counterfactual is set out in Figure 1. The diagram highlights the essential idea that outcomes for the control group acts a benchmark for the treatment group in the absence of treatment.

The assumption that the ‘constant difference’ in outcomes observed in the ‘pre’ period would have occurred in the absence of intervention in the ‘post’ period is at the heart of the difference-in-difference strategy. Importantly, this assumption cannot be verified by empirical data because it relates to a state of the world that is never observed. Because of this, any trial using a difference-in-difference strategy requires considerable work to design

\(^2\) We restrict our attention here to the simple case of the “constant difference” assumption, but in line with extensions to difference-in-differences approaches it would also be possible for a trial to use constant trend differences or triple-differences.
and implement an assignment mechanism where the ‘constant difference’ assumption will be plausible.

A starting point for designing a good assignment mechanism is a theory about why there should be an enduring difference in outcomes between the proposed treatment and control groups (Kahn-Lang and Lang, 2020). It is also necessary to show evidence of a constant difference in the average outcomes of the proposed treatment and control groups before the intervention has been implemented. This can be done using a standard ‘common trends’ test and provides essential information about whether a trial using the assignment mechanism is feasible.

After the trial has been undertaken it is also important to assess if it is plausible that there would have been a constant difference in the post period without any intervention.

One approach for investigating the plausibility of the assumption is to look at changes in the relative balance of characteristics of the treatment and control group in the ‘pre’ and ‘post’ periods. Ideally the difference in the prevalence of observable characteristics across the treatment and control groups should not change ‘pre’ and ‘post’. Large shifts in the relative balance of observable covariates suggest that there may have been unmeasured shocks impacting unequally on the control and treatment groups.

Placebo tests can also be used to investigate the plausibility of the constant difference assumption in the post period. For this test there is a placebo group for which there is a constant difference in outcomes relative the control groups in the ‘pre’ period. The placebo group does not receive the intervention, and the test is simply the hypothesis that there is no difference in the difference of the outcomes for the placebo and control groups.

For the analysis of a trial the standard difference-in-difference model can be used to formally assess the impact of the intervention on the outcomes. With cross sectional unit record data covering both ‘pre’ and ‘post’ time periods the impact of the intervention on
outcome ‘Y’ can be modelled as follows:

\[ Y_i = \beta_0 + \beta_1 POST_i + \beta_2 TG_i + \beta_3 (TG_i \ast POST_i) + X_i \beta_4 + \varepsilon_i \]  

(1)

where \( Y_i \) measures some outcome of interest, the intercept \( \beta_0 \) measures outcomes for the control group during the ‘pre’ time periods, the variable \( POST_i \) is an indicator of the post time-period, and the associated parameter \( \beta_1 \) measures any change in outcomes for both the control and treatment group during this time. The variable \( TG_i \) is dummy variable that defines the treatment group and the parameter \( \beta_2 \) is an estimate of the difference in outcomes for the treatment group relative to the control. The parameter \( \beta_3 \) on the interaction between the treatment group and post time-period is the difference-in-difference estimate of the treatment effect. The parameters \( \beta_4 \) relate to the effects of the control variables in \( X_i \) and it is usual to estimate equation (1) with and without these controls.

Ideally both the control variables in \( X_i \) and the functional form of equation (1) should have been specified prior to any analysis to ensure that the ‘researcher hands are tied’ during the analysis phase (Card, 2022; Rubin, 2008). Running a trial presents a unique opportunity in this regard because it creates the possibility of analysing data from the ‘pre’ period before any intervention is implemented. This approach can not only document the existence of a constant difference which is a precondition for conducting the trial, but also establishes the structure of the final analysis. This is particularly important if any of the specified covariates have time varying effects on the outcome (Arkhangelsky, 2019; Zeldow and Hatfield, 2021).

It is important to appreciate that conducting a DDCT is different to conducting either a retrospective or prospective difference-in-difference study. In particular, the researchers have the opportunity to determine who is assigned to control or treatment status, document that there is a constant difference prior to the intervention being assigned, manage the assignment process to ensure that the intervention is delivered as planned, ensure the wider conditions of the trial are controlled, and design and implement an appropriate data collection strategy. These additional design and management aspects of a trial strengthen the quality and credibility of the resulting analysis.

It is also useful to note that there are other ‘observational’ study methods that are the basis for other types of non-randomised trials. Foremost among these is a prospective trial

---

3. A prospective difference-in-difference observational study is similar but not identical to a DDCT because in a trial the researchers have control over many more aspects of what occurs. The Card and Krueger (1994) study of the impact of an increase in minimum wages provides a well-known example of a prospective difference-in-difference observational study. The researchers identified an impending natural experiment associated with a planned increase in the minimum wage in New Jersey, with no changes in neighbouring Pennsylvania. They designed and implemented a survey of fast-food restaurants in both states prior to the change in the New Jersey. This survey was then repeated with the same panel of restaurants after the minimum wage change occurred. As a thought experiment, the equivalent DDCT would look very similar, but the researchers would be able to determine the exact timing, amount, geographic coverage, and other aspects of a minimum wage change. This would allow some planning about how best to define the treatment and control groups with the purpose of studying the causal impact. In addition, prior to any minimum wage change the researchers would have an opportunity to collect data across multiple time periods to establish that there was a constant difference in the outcomes for the proposed control and treatment group. The timing of the implementation of the minimum wage change might also be randomised to ensure that it was not influenced by any considerations about future outcomes or study results.
using a regression discontinuity design (RDD). In this instance there is also deterministic assignment, with the probability of being assigned to treatment increasing abruptly at the threshold of a well-defined score (Cattaneo and Titiunik, 2022). Causal inference relies on the assumption that the expected outcomes of the control and treatment groups are continuous at the threshold (ie there is no abrupt change in measured or unmeasured covariates at the threshold).

One of the options for running a DDCT is to create an assignment rule using a cut-off on a well-defined score. As we discuss in more detail in the following section, this was the approach taken with our trial. Using a discontinuous assignment mechanism in multiple pre time periods means that it is possible to observe the difference in outcomes across the discontinuity during periods with no treatment. It also bolsters confidence in the assumption of a constant difference in the post period because individuals either side of the discontinuity are expected to be subject to very similar unmeasured shocks. The approach is also somewhat flexible in that it allows for analysis of impacts close to the threshold where the difference-in-discontinuity identification is strong, as well as for a slightly wider subset of the population which is useful for generalisability.

4. Design of the difference-in-difference control trial for the TAS campaign

Prior to the trial commencing we wrote a protocol documenting how the trial would be conducted and analysed using the methods described in the previous section. For the most part the trial was conducted according to this protocol, except for the fact that we were not able to conduct the ‘common trends’ analysis before the trial started because of delays in assembling our data. The protocol also included an assessment of the ethical considerations for the trial. These were discussed with a group called the National Beneficiaries Advocacy Consultative Group, and the design of the campaign was also approved by both the Ministry’s independent ethics panel and the internal privacy and human rights assessment process.

The protocol set out the two main research questions for the trial. First, to assess if contacting individuals modelled as eligible for TAS would increase the number of people receiving the payment. Second, to assess which of the three alternative modes of contact (phone calls, letters, and emails) was the most effective at increasing take-up.

The trial was designed to use the unit record data associated with the administration of the welfare system. A dataset was created that contained information on all main benefit recipients identified as eligible but not receiving TAS at the commencement of the campaign (2 November 2018), as well as three prior snapshot dates (6 July, 3 August, and 7 September 2018). For each person in these cross-sectional snapshots there was information on a range of demographic characteristics, housing and disability costs, and types of benefits and tax credits received.

The dataset also recorded whether TAS payments were made in the seven weeks after each snapshot date, together with the weekly dollar value of any TAS payment. The seven-week follow-up period was chosen to allow sufficient time for people to apply and then receive TAS payments. In the months prior to the campaign when there was no proactive
contact occurring, approximately 5% of all individuals modelled as eligible were granted the payment. The value of these payments averaged roughly $46 per week.

An important practical aspect of the design of the trial was an operational requirement that there would be 3000 people in total contacted, 600 of whom would receive phone calls, with the remainder receiving letters and emails.

Individuals were designated to be in either the control or treatment group based on their assessed level of financial need. This was measured by a person’s TAS deficiency score which was calculated for everyone using the microsimulation model. Individuals with a TAS deficiency score of $72.45 or more per week were allocated to the treatment group and contacted. Individuals with a TAS deficiency score that was less than this amount were not contacted. The control group were defined as those with a TAS deficiency of between $52.45 and $72.44. Our reasoning was that this group would more likely resemble individuals in the treatment group and hence be more likely to experience similar shocks in the post-period.

The exact value of the $72.45 cut-off was calculated just prior to the trial commencing based on the operational requirement that only 3,000 people could be contacted for the campaign. Among the population eligible for the campaign in November there were 3,000 people with TAS deficiency greater than $72.45. This threshold was then used to establish the cut-off to define the treatment group in both the ‘pre’ and the ‘post’ periods.

The trial was designed to use the difference in take-up between the treatment and control groups, after adjusting for the difference in take-up that had been observed in the pre time periods. In addition, the campaign was also designed to assess the effectiveness of each of the three different modes of contact. Among the treatment group, the type of contact was randomly assigned. This then allowed a comparison of the relative effectiveness of the different treatment types among those contacted.

Formally the population who were “eligible for the campaign” were all main benefit recipients who were modelled as eligible for TAS but not receiving the payment. This population also excluded those who had been proactively contacted about TAS in the past 12 months, or had applied, been declined, or expired from TAS in the previous 120 days. These exclusion criteria reflected operational decisions about who should be the focus of any future roll-out of the campaign.

A further set of criteria were used to select a subset of the population eligible for the campaign who were also eligible for randomisation. This population needed to have all their address and contact details recorded, and not be sight or hearing impaired. The standardised nature of the criteria meant that all individuals in the population eligible for the campaign were assessed as suitable for randomisation, irrespective of whether they were assigned to the treatment or control group for the trial.

Table 1 shows the number of people identified by the MSIM model as eligible for TAS at each of the ‘pre’ and ‘post’ intervention dates. The table documents the impact of the criteria used to select the population. In the first column, Population ‘A’ is all main benefit clients estimated to be eligible but not receiving TAS. In the next column, ‘B’ is the population eligible for the campaign which reflect the exclusion criteria for the trial. Individuals in population ‘C’ were also eligible for randomisation and are the focus of the impact analysis in what follows. Population ‘D’ are those individuals who were eligible for the campaign but not able to be randomly assigned.
Table 1: Overview of the trial population and restrictions

<table>
<thead>
<tr>
<th></th>
<th>Population (A)</th>
<th>Population (B)</th>
<th>Population (C)</th>
<th>Population (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-intervention:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>July 2018</td>
<td>27,862</td>
<td>11,986</td>
<td>8,684</td>
<td>3,302</td>
</tr>
<tr>
<td>August 2018</td>
<td>29,140</td>
<td>12,851</td>
<td>9,384</td>
<td>3,467</td>
</tr>
<tr>
<td>September 2018</td>
<td>30,585</td>
<td>14,365</td>
<td>10,581</td>
<td>3,784</td>
</tr>
<tr>
<td>Post-intervention:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>November 2018</td>
<td>31,987</td>
<td>17,495</td>
<td>13,067</td>
<td>4,428</td>
</tr>
<tr>
<td>Total pre and post:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total observations</td>
<td>119,574</td>
<td>56,697</td>
<td>41,716</td>
<td>14,981</td>
</tr>
<tr>
<td>Unique people</td>
<td>47,161</td>
<td>23,227</td>
<td>17,957</td>
<td>5,270</td>
</tr>
</tbody>
</table>

Notes: Population-A consists of people who were not receiving but estimated eligible for TAS (by MSIM); Population-B consists of people in ‘A’ who were eligible for the campaign; Population-C consists of people in ‘B’ who were eligible for randomisation; and Population-D consists of people in ‘B’ who were not eligible for randomisation.

Source: MSD data.

Implementation

Table 2 shows the population eligible for the campaign on 2 November 2018 when the trial commenced. By design there were 3,000 people with a TAS deficiency score of $72.45 or more per week on this date, and 17,495 in the population eligible for the campaign. Of the 3,000 individuals whose TAS deficiency score was $72.45 or above, 2,290 people were eligible to be randomized into one of three different types of contact because they had complete contact details and no restrictions on what form of contact was appropriate. The remaining 710 people were not eligible for randomisation and were allocated to a specific type of contact.

Randomisation was undertaken in SAS using different sampling probabilities to ensure that the total allocation of people to each type of contact met the predetermined operational requirement that there be 600 phone calls, 1,200 letters, and 1,200 emails. After identifying the total number of slots available after the non-randomized individuals were allocated, the sampling probabilities were calculated to be letters (p=0.23), emails (p=0.52) and phone calls (p=0.25). Appendix Table A1 provides a breakdown of the characteristics of the randomized treatment group across the different forms of contact. Consistent with randomisation there is no evidence of imbalance across the characteristics.

Following allocation of individuals to the control and treatment groups, lists of people were provided for a mail out, to MSD’s centralised email system, and to the Ministry’s contact centre for outbound phone calls. Individuals were contacted over several weeks using either a letter, email, or phone call. The letter was both a hardcopy sent to the recipient mail address as well as electronic version that was viewable in a personalised online Ministry account. The email was sent to the recipient’s private email address from the Ministry and, as with the letter, was also viewable online. The phone call was from the Ministry’s call centre to the person’s preferred phone number.
Table 2: Campaign population in November 2018

<table>
<thead>
<tr>
<th>TAS deficiency</th>
<th>Group</th>
<th>Population (B)</th>
<th>Population (C)</th>
<th>Population (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1.00 to $32.44</td>
<td>No contact</td>
<td>11,054</td>
<td>8,149</td>
<td>2,905</td>
</tr>
<tr>
<td>$32.45 to $52.44</td>
<td>No contact</td>
<td>2,204</td>
<td>1,696</td>
<td>508</td>
</tr>
<tr>
<td></td>
<td>(placebo)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$52.45 to $72.44</td>
<td>No contact</td>
<td>1,237</td>
<td>932</td>
<td>305</td>
</tr>
<tr>
<td></td>
<td>(control)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$72.45 and above</td>
<td>-email</td>
<td>1,200</td>
<td>1,191</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>-letter</td>
<td>1,200</td>
<td>521</td>
<td>679</td>
</tr>
<tr>
<td></td>
<td>-phone</td>
<td>600</td>
<td>578</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Total treatment</td>
<td>3,000</td>
<td>2,290</td>
<td>710</td>
</tr>
<tr>
<td>Total people</td>
<td></td>
<td>17,495</td>
<td>13,067</td>
<td>4,428</td>
</tr>
</tbody>
</table>

Notes: see notes to Table 1 for description of populations.
Source: MSD data.

Each of the different forms of contact were designed to be as personalised as possible, easily understood, and with a clear message about how to apply for a TAS payment. The box below provides the text of the letter and email. The introduction to the outbound phone call was also scripted in a similar manner.

Tēnā koe Aroha

Call us – you may qualify for extra help
I’m getting in touch because it looks like you may be able to get Temporary Additional Support. This helps pay for essentials when it’s hard to make ends meet. You don’t have to pay it back. Call us now on 0800 559 009 so we check you’re getting you all the help you can. If we find you can get extra help, we’ll pop an application form for Temporary Additional Support in the post. The call will only take around five minutes and you could be much better off.

Nāku iti noa, nā

Awhina

The relatively simple nature of the intervention and the use of a list of people to contact meant that assignment was well managed and occurred as planned. Initial assignment occurred at the end of the first week of November 2018 and lists of the 3,000 people were distributed to the three different ‘channels’ assigned to manage the contact. The letters were all sent early in the following week, while the distribution of emails and outbound calls were phased over the first fortnight. A small number of individuals on the email (75) and phone lists (10) were not contacted as they had cancelled their benefit in the intervening days between assignment and the time they were scheduled to be contacted.
Table 3 shows estimates of the number of people who were assigned and subsequently contacted. At least 278 letters were read electronically, although it would be expected that many of the hard copies would also have been read. Approximately two-thirds (64%) of the emails that were sent were recorded as being opened by the recipient. Of people who were phoned about three-quarters (76%) could be reached, and in a third of these cases a message was left on voice mail.

5. Analysis of results

The analysis set out below seeks to identify the causal impact of the campaign on receipt and income from the TAS payment. Our focus is the average treatment effect on all those assigned, irrespective of whether they received any contact. We firstly document descriptive patterns before providing more formal difference-in-difference estimates of the intention to treat impacts.

Descriptive overview of findings

Figure 2 shows the proportion of people who were granted a TAS payment within the subsequent 7 weeks across $10 TAS deficiency bands. Each of the four snapshot dates are shown separately. For the snapshot dates before the campaign commenced the proportion who were granted TAS was relatively similar irrespective of the level of TAS deficiency.

In November when the campaign was undertaken there was a large increase in the proportion granted TAS among the treatment group with a TAS deficiency of more than $72.45. For the treatment group, the proportion granted TAS increased nearly threefold from about 6% in the pre periods to around 17% in the post period.

Another feature is that there is some evidence of a general lift of around 2% in the proportion being granted TAS among those who were not contacted in November (a TAS deficiency of $72.44 or less). This may be due to seasonal differences, a change in the composition of people in the eligible population, or an increase in more ad hoc proactive discussions about TAS eligibility by MSD staff. In relation to this last point, the campaign was a product of a wider concern about take-up, and it is possible that this also affected the
control group. To the extent this occurred, we expect it would have impacted equally across the control and treatment groups, and hence should not bias the difference-in-difference estimator.

Figure 3 shows the average weekly amounts of TAS income after seven weeks for each of the groups described above. This is the unconditional average across everyone in each TAS deficiency band, irrespective of whether they received TAS. The figure again shows a clear increase in TAS income for the treatment group in November. As context, it is also useful to note that after the campaign commenced in November, the average payment was around $80 per week for individuals in the treatment group who were granted TAS.

**Formal analysis of the impacts of the trial**

The graphical representation of the impacts of the campaign set out in Figure 2 and Figure 3 summarise the increase in TAS payments resulting from the campaign. In this sub-section we quantify and assess the robustness of these impacts and show how the magnitude of the impacts differed by the type of contact.

For the formal analysis we use the standard difference-in-difference set-up that our protocol specified. The main outcomes evaluated were the probability of a TAS grant within seven weeks, as well as the average value of TAS payments across everyone including non-recipients.
We use the data from all four snapshots, and in most cases restrict the analysis to only people with an estimated TAS deficiency of $52.45 or more per week. We characterise people into the control group (TAS deficiency from $52.45 to $72.44), and the treatment group (TAS deficiency $72.45 and above).

Our general model has the form:

$$ Y_i = \beta_0 + \beta_1 Nov_i + \beta_2 TG_i + \beta_3 (Nov_i \ast TG_i) + X_i \beta_4 + \varepsilon_i $$

where $Y_i$ is an outcome of interest (either TAS receipt or TAS payments) for person-$i$; $Nov_i$ is a dummy variable that defines the November date when the campaign occurred; and $TG_i$ is a dummy variable for whether a person was assigned to the treatment group. $X_i$ is a vector of controls which include the exact amount of a person’s TAS deficiency, trial months, demographics characteristics including gender, ethnicity, age and family type, and a variety of TAS and other benefit-related variables such as prior TAS receipt. The coefficient $\beta_3$ identifies the difference-in-difference impact of the campaign conditional on these covariates. We estimate the relationships using OLS and standard errors are clustered on the individual given that the observations are not independent across each of the cross-sectional snapshots.

We also consider two extensions to this model. First, equation (2) restricts each treatment mode of contact (letter, email, phone) to have the same effects. This restriction is
relaxed to estimate the effect of each type of treatment by interacting the \((N_{ov_i} \times T_{G_i})\) variable with separate indicators for the different randomly assigned treatments (letter, email, or phone call).

Second, an important feature of our trial is that we used a specific cut-off on the TAS deficiency score to determine who was in the treatment group. This ensured that there was controlled assignment, but it also meant we could undertake a difference-in-difference analysis at a discontinuity. As an extension to this analysis, we restrict the treatment group to only those whose TAS deficiency was at most $20$ above the cut-off.\(^4\)

**Estimates of impacts**

Table 4 and 5 reports the estimated impact of the campaign on the proportion of the treatment group who were subsequently granted TAS and the value of these TAS payments. Models 1 and 2 report the overall treatment effects without and with controls. Models 3 and 4 allow separate treatment effects by method of contact without and with controls. Model 5 restricts the treatment group sample to those with a TAS deficiency between $72.45$ and $92.44$.

Table 4 shows that the campaign increased receipt of the payment by 10 percentage points. Our preferred specification for looking at the effectiveness of the different types of contact is model 4 which controls for individual month effects, demographics and benefit related characteristics. This shows that ‘email’ contact increased the proportion of people being granted TAS by 7 percentage points, ‘letter’ contact by 11 percentage points, and ‘outbound phone call’ contact by 17 percentage points.

Model 5 restricts the treatment group sample to only those individuals who are at most $20$ above the cut-off. This difference in discontinuity analysis, which uses about half the sample, shows that after controlling for the specific value of a person’s TAS deficiency as well as other covariates, the estimated treatment effect around the threshold was broadly in line with the impact observed across the wider sample. However, the estimated impact of ‘email’ contact was reduced.

Table 5 reports the estimates of how the campaign changed average weekly TAS payments. These shows that across everyone in the treatment groups (irrespective of whether they were subsequently granted TAS) the average amount of income from TAS increased by almost $9$ per week. Our preferred estimates of the impact of each of the three different types of contact are from model 4 which controls for individual month effects, demographics, and benefit-related characteristics. For individuals who were emailed, the average impact was almost $6$ per week, while for those sent letters gained just over $9$ per week on average, and those who were phoned gained almost $15$ per week on average. Model 5 uses a restricted population for the treatment group and shows slightly smaller impacts for both ‘email’ and ‘phone call’ contacts.

---

4. Our protocol specified that we would undertake the trial using difference-in-difference methods. We could of course have conducted a trial without any ‘pre’ time-period data and used a regression discontinuity design. Figures 2 and 3 show a ‘jump’ in the outcomes at our assignment threshold in the ‘post period’ which suggest treatment impacts that are the same order of magnitude as from our analysis. However, it is useful to reflect that a single period RDD approach would not have been as robust because having ‘pre’ period data where treatment was not allocated provides good evidence of the continuity of the regression function across the threshold.
Table 4: Estimated impact of campaign on TAS-grants within 7 weeks

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total impact ($\beta_3$)</td>
<td>0.10***</td>
<td>0.10***</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Email ($\beta_{3e}$)</td>
<td>—</td>
<td>—</td>
<td>0.06***</td>
<td>0.07***</td>
<td>0.04**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Letter ($\beta_{3l}$)</td>
<td>—</td>
<td>—</td>
<td>0.11***</td>
<td>0.11***</td>
<td>0.14***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Phone ($\beta_{3p}$)</td>
<td>—</td>
<td>—</td>
<td>0.16***</td>
<td>0.17***</td>
<td>0.17***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Covariates</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.03</td>
<td>0.05</td>
<td>0.03</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>No. observations</td>
<td>10,846</td>
<td>10,846</td>
<td>10,846</td>
<td>10,846</td>
<td>5,154</td>
</tr>
</tbody>
</table>

Note: Estimation based on population ‘C’ restricted to individuals with a TAS deficiency of $52.45 or more. Standard errors are clustered on individuals. All models are versions of equation 2. Model 5 restricts the treatment group to individuals with a TAS deficiency of between $72.45 to $92.44. Covariates include TAS deficiency, month, demographic and benefit controls. Impacts estimated with a linear probability model. The conclusions about the magnitudes and significance of the parameters were also confirmed using logistic regression.

*** p-value < 0.01 ** p-value < 0.05 * p-value < 0.1.

Source: MSD data.

The impacts described above reflect both the extent to which the different methods of contact were able to reach the treatment group (as set out in table 3), as well as the possibly different effect of take-up once contacted. Given that the trial was aiming to assess the overall effectiveness of the different modes of contact these intention-to-treat provide useful information.

Was there a constant difference in outcomes prior to the trial commencing?

The analysis of the impact of the trial assumes there was a constant difference in outcomes between the treatment and control groups prior to the trial commencing. For our trial we wanted to test the assumption of a constant difference prior to commencing the campaign, but delays in assembling outcome information meant that this was not possible, and we were only able to undertake the analysis after the trial had started. If we had been able to undertake the analysis earlier, it would have strengthened our conclusions and reduced the risk of embarking on a poorly designed trial. As it turned out these risks did not eventuate because there was evidence of a constant difference in the ‘pre’ periods. The trial also had a sufficient sample size to detect impacts despite the relatively modest levels of actual engagement among those contacted.

Our test of the assumption of a constant historical difference in outcomes between the treatment and control groups uses data from the July, August and September monthly snapshots before the trial started. We estimate a model of the form:

$$Y_i = \beta_0 + \beta_1 Aug_i + \beta_2 Sep_i + \beta_3 TG_i + \beta_4 Aug_i \times TG_i + \beta_5 Sep_i \times TG_i + X_i' \beta_6 + \varepsilon_i.$$  \hspace{1cm} \text{(3)}
### Table 5: Estimated impact of campaign on weekly TAS-income after 7 weeks

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total impact (\beta_3)</strong></td>
<td>$8.72***$</td>
<td>$8.90***$</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$(0.98)$</td>
<td>$(0.97)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email (\beta_3e)</td>
<td>—</td>
<td>—</td>
<td>$5.73***$</td>
<td>$5.93***$</td>
<td>$3.49**$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(1.09)$</td>
<td>$(1.09)$</td>
<td>$(1.49)$</td>
</tr>
<tr>
<td>Letter (\beta_3l)</td>
<td>—</td>
<td>—</td>
<td>$8.92***$</td>
<td>$9.08***$</td>
<td>$9.70***$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(1.55)$</td>
<td>$(1.55)$</td>
<td>$(2.49)$</td>
</tr>
<tr>
<td>Phone (\beta_3p)</td>
<td>—</td>
<td>—</td>
<td>$14.70***$</td>
<td>$14.70***$</td>
<td>$13.93***$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(1.70)$</td>
<td>$(1.70)$</td>
<td>$(2.75)$</td>
</tr>
<tr>
<td>Covariates</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.03</td>
<td>0.05</td>
<td>0.04</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>No. observations</td>
<td>10,846</td>
<td>10,846</td>
<td>10,846</td>
<td>10,846</td>
<td>5,154</td>
</tr>
</tbody>
</table>

Note: Estimation based on population ‘C’ restricted to individuals with a TAS deficiency of $52.45 or more. Standard errors are clustered on individuals. All models are versions of equation 2. Model 5 restricts the treatment group to individuals with a TAS deficiency of between $72.45 to $92.44. Covariates include TAS deficiency, month, demographic and benefit controls. Impacts estimated with a linear probability model. The conclusions about the magnitudes and significance of the parameters were also confirmed using logistic regression.

*** p-value < 0.01 **p-value < 0.05 *p-value < 0.1.

Source: MSD data.

Within this set-up the parameter \(\beta_3\) measures the difference between the treatment and control groups. Our assessment of ‘common trends’ uses an F-test of the linear restriction that the estimate \(\beta_4\) and \(\beta_5\) are jointly zero.

As reported in appendix Table A2, after controlling for covariates there was no practical or statistically significant difference between the treatment and control group. Importantly, there was no evidence to suggest that the assumption of a constant difference was not valid.

Important context for the assessment of the constant differences is that the size of the population progressively increased between July and November 2018 (shown in column A of Table 1). The growth in the total number of people eligible but not receiving TAS likely reflected several different factors including growth in the underlying population of people in receipt of income-tested main benefits. Column (B) of Table 1 also shows a proportionally larger growth in the ‘eligible for the campaign’ population. This likely occurred because to be included in the ‘eligible for the campaign’ group a person could not have been contacted about TAS in the last 120 days and the cessation of the previous automated process will have reduced the numbers of people who were affected by this specific exclusion criteria.

As mentioned, the analysis also makes the unverifiable assumption that the previously observed difference in the outcomes between the treatment and control group would also have continued in the intervention time-period. We assess the plausibility of this assumption by looking at the relative balance of covariates across the treatment and control groups. Changes in the relative balance of observed covariates between the ‘pre’ and ‘post’ time-periods are suggestive that the treatment and control groups are experiencing different shocks, and there could be unobserved confounding (Wing et al., 2018). To test this, for
Table 6: Placebo test regressions

<table>
<thead>
<tr>
<th></th>
<th>TAS grants</th>
<th>TAS payments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Placebo impact ($\beta_3$)</td>
<td>0.02</td>
<td>$0.91$</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.79)</td>
</tr>
<tr>
<td>Month, demographic and benefit controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>No. observations</td>
<td>8,412</td>
<td>8,412</td>
</tr>
</tbody>
</table>

Note: Estimation is based on population ‘C’ restricted to individuals with a TAS deficiency of $32.45 to $72.44. Clustered standard errors are in parentheses. Demographic and benefit controls relating to sex, ethnicity, age group, family type, benefit type, and TAS-related variables.

*** p-value < 0.01 ** p-value < 0.05 * p-value < 0.1.
Source: MSD data.

Insights into the predictive accuracy of the MSIM model from the trial

As well as understanding the relative impact of different forms of contact, an important rationale for the campaign was to assess if the MSIM microsimulation model would accurately identify individuals who were eligible to receive TAS. Table 7 reports the outcomes from the campaign in November and shows that overall, 8% of individuals identified by the
Table 7: Overall outcomes from the November 2018 campaign

<table>
<thead>
<tr>
<th>Group</th>
<th>TAS deficiency</th>
<th>Number of people</th>
<th>Granted TAS within seven weeks</th>
<th>Average TAS grant value for those granted</th>
<th>Average predicted TAS grant for those granted</th>
<th>Average prediction error (RMSE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No contact</td>
<td>$1.00-$32.44</td>
<td>8,149</td>
<td>0.05</td>
<td>$29</td>
<td>$15</td>
<td>$27</td>
</tr>
<tr>
<td>No contact (placebo)</td>
<td>$32.45-$52.44</td>
<td>1,696</td>
<td>0.09</td>
<td>$47</td>
<td>$41</td>
<td>$21</td>
</tr>
<tr>
<td>No contact (control)</td>
<td>$52.45-$72.44</td>
<td>932</td>
<td>0.08</td>
<td>$60</td>
<td>$60</td>
<td>$18</td>
</tr>
<tr>
<td>Treatment</td>
<td>$72.45+</td>
<td>2,290</td>
<td>0.16</td>
<td>$80</td>
<td>$83</td>
<td>$14</td>
</tr>
<tr>
<td>-email</td>
<td></td>
<td>1,191</td>
<td>0.13</td>
<td>$79</td>
<td>$83</td>
<td>$15</td>
</tr>
<tr>
<td>-letter</td>
<td></td>
<td>521</td>
<td>0.17</td>
<td>$78</td>
<td>$82</td>
<td>$16</td>
</tr>
<tr>
<td>-phone</td>
<td></td>
<td>578</td>
<td>0.23</td>
<td>$84</td>
<td>$85</td>
<td>$12</td>
</tr>
<tr>
<td>Total</td>
<td>$1 or more</td>
<td>13,067</td>
<td>0.08</td>
<td>$52</td>
<td>$47</td>
<td>$22</td>
</tr>
</tbody>
</table>

Note: Estimation is based on population ‘C’.
Source: MSD data.

model went on to be granted TAS. Of those who were assigned to the phone call (which was the most effective form of contact), 23% were subsequently granted TAS.

The 23% take-up rate for the phone call establishes a conservative lower bound measure of the accuracy of the model among the treatment group. It is conservative because not everyone in the phone call group was contacted and informed about their potential eligibility. In addition, not everyone who was identified as eligible during their phone call followed through with the application process for the TAS payment.

A less conservative estimate of accuracy uses the detailed information recorded by the call centre for the campaign. Of all the individuals who were able to be contacted on the phone (n=282), approximately 203 were identified as eligible for the payment after the initial assessment was undertaken by the Call Centre staff. This suggests that around 72% of the phone call group may have been eligible and represents a better estimate of the accuracy of the model.

6. Implementation of the TAS campaign as an ongoing business process

Following the analysis of the results from the trial there was a decision by the Ministry of Social Development to implement an ongoing process of proactive contact to increase the take-up of TAS. There were some changes to the process used in the original campaign, and some of information provided to potential recipients was refined following a small qualitative study (Errington and Human, 2019). The new ongoing process continued to use the MSIM model to identify people who were potentially eligible. Approximately 3,000 people per month were contacted, and the new business process progressively worked through the list of people eligible according to the level of modelled TAS deficiency.
Individuals were contacted by email or letters, and sometimes a combination of both approaches. Outbound phone calls were not used due to high demands of other work. In addition to the new proactive business process, there were also other initiatives aimed at improving uptake of TAS. One of these was the provision of a prompt (based on information from the MSIM model) for case managers when interacting with clients. This ‘proactive client entitlement’ initiative was piloted and expanded to several different regions during 2019.

7. Conclusion

Our trial aimed to test a new approach for increasing the take-up of a welfare payment in New Zealand. The payment called Temporary Additional Support (TAS) provides additional support for people in financial hardship, and there was credible evidence that not everyone who was eligible was claiming the payment. The New Zealand Ministry of Social Development wanted to assess if proactively contacting individuals who were modelled as eligible would increase the number of people receiving the payment.

A randomised trial was an obvious choice for testing the new approach. However, there was reluctance from stakeholders to use this type of trial because it would have assigned some individuals with very little disposable income to a control condition. Our solution was to develop and implement what we called a difference-in-difference control trial to test the new approach. The trial used a measure of financial hardship and a cut-off to determine who would receive an invitation to apply for the payment. Individuals above the cut-off (the group experiencing the most severe financial hardship) received the intervention. Individuals with a score just below the cut-off became the control group.

In multiple time periods prior to the intervention occurring there was no statistically significant difference in receipt of the TAS payment between the two groups. In the ‘post’ period after the intervention was implemented there was a 10-percentage point increase in take-up for the treatment group relative to the control, and income from TAS was increased by almost $9 per week among the treatment group.

The trial provided compelling evidence for the effectiveness of the new intervention and was implemented in a modified form as an ongoing business process. The results from the trial contribute to a growing international evidence base about how best to tackle the problem of low take-up of welfare payments (Bargain et al., 2012; Currie, 2004; Hernanz et al., 2004; Ko and Moffitt, 2022). Our results suggest that proactive contact based on modelled eligibility can be effective strategy to increase take-up.

This study also demonstrates the value of using a DDCT design for undertaking trials. Our use of a DDCT was motivated by the need to test a new intervention without using an RCT. This problem occurs in other areas of public policy when randomisation is either not ethically appropriate or not feasible due to cost (Formoso et al., 2013). Although the inference from a DDCT is not as rigorous as an RCT, it can be a good choice if a randomised trial is not possible. The DDCT is not often used, and we think more widespread use of the design would expand the range of interventions that are subject to rigorous testing prior to wider implementation.

One possible reason for there being so few examples of DDCT’s is that there is not a lot of guidance about how such trials should be conducted and reported. We think that
developing some practical guidelines would likely encourage a greater uptake and use of this important approach for testing interventions and generating evidence. This guidance might also usefully consider the merits of using a DDCT relative to other non-randomised approaches such as ex ante RDD for conducting trials.

Acknowledgments

We would like to thank the editors and two anonymous referees for many insightful comments. The paper has also benefited from review from a range of staff at the Ministry of Social Development including Kavitha Ilanko, Ainsley Smith, Tim Maloney, Hugh Webb, Marc de Boer, Tracey McIntosh, Margaret McArthur, and Kahukore Baker. We also received valuable feedback from participants at the New Zealand Association of Economists Conference and the Victoria University of Wellington Econometrics Workshop.

References


**Appendix: Additional tables**

<table>
<thead>
<tr>
<th>Table A1: Characteristics of individuals in the randomized contact groups</th>
<th>Letter (n=521)</th>
<th>Email (n=1191)</th>
<th>Phone (n=578)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>0.60</td>
<td>0.62</td>
<td>0.59</td>
<td>0.37</td>
</tr>
<tr>
<td>Māori</td>
<td>0.26</td>
<td>0.26</td>
<td>0.25</td>
<td>0.82</td>
</tr>
<tr>
<td>European</td>
<td>0.42</td>
<td>0.41</td>
<td>0.43</td>
<td>0.63</td>
</tr>
<tr>
<td>Pacific</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.83</td>
</tr>
<tr>
<td>Other ethnicity</td>
<td>0.24</td>
<td>0.25</td>
<td>0.22</td>
<td>0.56</td>
</tr>
<tr>
<td>Unspecified ethnicity</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.62</td>
</tr>
<tr>
<td>Under 25 years of age</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.75</td>
</tr>
<tr>
<td>25 to 44 years of age</td>
<td>0.36</td>
<td>0.38</td>
<td>0.35</td>
<td>0.42</td>
</tr>
<tr>
<td>45 to 64 years of age</td>
<td>0.43</td>
<td>0.41</td>
<td>0.43</td>
<td>0.69</td>
</tr>
<tr>
<td>65 years of age and over</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>0.91</td>
</tr>
<tr>
<td>Single person</td>
<td>0.54</td>
<td>0.53</td>
<td>0.52</td>
<td>0.75</td>
</tr>
<tr>
<td>Couple no children</td>
<td>0.18</td>
<td>0.18</td>
<td>0.21</td>
<td>0.45</td>
</tr>
<tr>
<td>Couple with children</td>
<td>0.07</td>
<td>0.07</td>
<td>0.05</td>
<td>0.23</td>
</tr>
<tr>
<td>Sole parent with children</td>
<td>0.20</td>
<td>0.21</td>
<td>0.22</td>
<td>0.71</td>
</tr>
<tr>
<td>Non beneficiary</td>
<td>0.10</td>
<td>0.12</td>
<td>0.09</td>
<td>0.13</td>
</tr>
<tr>
<td>Job Seeker Support</td>
<td>0.43</td>
<td>0.38</td>
<td>0.37</td>
<td>0.05*</td>
</tr>
<tr>
<td>Supported Living Payment</td>
<td>0.15</td>
<td>0.17</td>
<td>0.19</td>
<td>0.22</td>
</tr>
<tr>
<td>Sole Parent Support</td>
<td>0.16</td>
<td>0.16</td>
<td>0.18</td>
<td>0.48</td>
</tr>
<tr>
<td>NZS or Veterans Pension</td>
<td>0.14</td>
<td>0.16</td>
<td>0.15</td>
<td>0.58</td>
</tr>
<tr>
<td>Modelled TAS deficiency</td>
<td>$127.35</td>
<td>$129.27</td>
<td>$132.20</td>
<td>0.33</td>
</tr>
<tr>
<td>Estimated TAS entitlement</td>
<td>$85.96</td>
<td>$86.91</td>
<td>$89.39</td>
<td>0.18</td>
</tr>
<tr>
<td>Value of assets</td>
<td>$29.47</td>
<td>$41.10</td>
<td>$35.59</td>
<td>0.47</td>
</tr>
<tr>
<td>TAS receipt (previous 24 months)</td>
<td>0.44</td>
<td>0.42</td>
<td>0.44</td>
<td>0.69</td>
</tr>
</tbody>
</table>

*Note: Sample is Population ‘C’, November 2018. F-test used to calculate the p-value for any difference in the proportions.*

*Source: MSD data.*
### Table A2: Test of constant ‘pre’ period differences between treatment and control groups

<table>
<thead>
<tr>
<th></th>
<th>TAS grants model</th>
<th>TAS payment model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment group difference ($\beta_3$)</td>
<td>-0.01 (0.01)</td>
<td>-0.25 (0.83)</td>
</tr>
<tr>
<td>Difference-in-difference in August ($\beta_4$)</td>
<td>0.00 (0.01)</td>
<td>0.15 (0.88)</td>
</tr>
<tr>
<td>Difference-in-difference in September ($\beta_5$)</td>
<td>0.02 (0.01)</td>
<td>1.21 (0.01)</td>
</tr>
<tr>
<td>$H_0: \beta_4 = \beta_5 = 0$ (F-test p-value)</td>
<td>0.55</td>
<td>0.41</td>
</tr>
<tr>
<td>Demographic and benefit controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>No. Observations</td>
<td>7,624</td>
<td>7,624</td>
</tr>
</tbody>
</table>

Note: Population ‘C’ restricted to individuals with a TAS deficiency of $52.45 or more. Impacts estimated with OLS. Standard errors clustered on individuals are in parentheses. Demographic and benefit controls for sex, ethnicity, age group, family type, benefit type, and TAS-related variables. For the TAS grants model the results were comparable when using a logistic model.

*** p-value < 0.01 ** p-value < 0.05 * p-value < 0.1.

Source: MSD data.
Table A3: Testing for relative covariate balance across the treatment and control groups

<table>
<thead>
<tr>
<th></th>
<th>Proportion / average $ value</th>
<th>Difference-inifference ($\beta_3$)</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>0.00</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>0.00</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Māori</td>
<td>0.00</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>European</td>
<td>0.01</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Pacific</td>
<td>-0.01</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>Other ethnicity</td>
<td>0.00</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>Under 25 years of age</td>
<td>0.00</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>Age 25 to 44 years</td>
<td>0.03</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Age 45 to 64 years</td>
<td>-0.03</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Above 65 years of age</td>
<td>0.00</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>Single no dependent children</td>
<td>0.01</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Couple no dependent children</td>
<td>0.00</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>Couple with dependent children</td>
<td>0.01</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>Sole parent with dependent children</td>
<td>-0.01</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Job Seeker Support payment</td>
<td>-0.01</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Supported Living Payment</td>
<td>-0.01</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>Sole Parent Support</td>
<td>0.00</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>NZS or Veterans Pension</td>
<td>0.00</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>Non-beneficiary</td>
<td>0.02*</td>
<td>(0.01)</td>
<td></td>
</tr>
<tr>
<td>Average modelled TAS deficiency</td>
<td>-$2.81***</td>
<td>(0.87)</td>
<td></td>
</tr>
<tr>
<td>Average modelled TAS entitlement</td>
<td>-$0.64</td>
<td>(0.52)</td>
<td></td>
</tr>
<tr>
<td>Average assets</td>
<td>$9.24*</td>
<td>(5.17)</td>
<td></td>
</tr>
<tr>
<td>Proportion with TAS in prior 24 months</td>
<td>-0.03</td>
<td>(0.02)</td>
<td></td>
</tr>
</tbody>
</table>

Note: N=10,846. Note: Population ‘C’ restricted to individuals with a TAS deficiency of $52.45 or more. Clustered standard errors.

*** p-value < 0.01 ** p-value < 0.05 * p-value < 0.1.

Source: MSD data.