MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

Findings from the Temporary Additional Support Campaign (2018)

Authors

David Rea, Dean Hyslop and Ainsley Smith

Acknowledgements

The paper has greatly benefited from review and feedback from a range of MSD staff, Professor Tim Maloney, and participants at the New Zealand Association of Economists Conference (2019) and the Victoria University of Wellington Econometrics Workshop (2019)

Disclaimer

The paper represents the views of the authors.

Published

Online, 19 August 2021 at Publications and resources - Ministry of Social Development (msd.govt.nz)

ISBN

978-1-99-002377-4 - PDF

Contents

Table of figures

Table of tables

Executive summary

Temporary Additional Support (TAS) is a weekly hardship payment for individuals with high ongoing costs relative to their income. The payment is available for 13 weeks, but it is often renewed.

The payment provides an important element of the welfare safety net for people facing severe financial hardship.

In late 2018 there were just over 59,000 people receiving payments. Although there was some uncertainty about the exact magnitude, modelling at the time suggested that the take-up of this payment among those eligible could be as low as 68 percent.

Low rates of take-up mean that there is more financial hardship than necessary, with associated consequences related to families not being able to afford housing, food, health services and other essentials.

Historically the Ministry of Social Development had sent letters to people who were estimated to be eligible but not receiving TAS. This mail out stopped in June 2018 as changes in the structure of benefit payments meant new computer code was needed to identify eligible individuals. The project team tasked with rebuilding a new automated process came to the view that the Ministry's microsimulation model (MSIM) should be repurposed to calculate if people might be eligible for TAS based on existing administrative data about their other payments and housing costs.

The November 2018 Temporary Additional Support campaign aimed to assess a new method of identifying people who were not currently receiving but potentially eligible for the payment. was

The campaign also aimed to test the effectiveness of different forms of proactive contact. Those with the highest level of estimated financial need were contacted by either letter, email, or phone and told they might qualify for the payment. People who decided to apply were then required to follow the normal process for TAS applications involving an initial assessment of eligibility over the phone, filling out an application form and attending an appointment at a local office.

This paper analyses the impacts of the campaign using a prospective difference-in-difference methodology. Key findings are that:

- proactive contact increased the number of people applying for and being granted the TAS payment by 10 percentage points
- email contact increased the rate of TAS receipt by 7 percentage points, letters by 11 percentage points, and outbound phone calls by 17 percentage points
- the campaign targeted individuals with the highest levels of financial need, and for those contacted who were subsequently granted TAS the payment was around $80 per week

- in the short-term the campaign did not result in people remaining on benefit for longer, although we do not know how the campaign affected time on benefit after 7 weeks

- among the population targeted by the campaign, between 23 percent and 72 percent of those identified by the modelling were likely eligible for the payment.

The campaign has been subsequently rolled out in a modified form as a continuing business process.

Background on Temporary Additional Support (TAS)

The welfare system provides a form of community support for individuals who have insufficient earnings to provide an adequate standard of living.

Welfare payments consist of income-tested main benefits and supplementary payments for those with higher costs.

The key supplementary payments are the Accommodation Supplement, Disability Allowance, Temporary Additional Support, Childcare Assistance, and the Winter Energy Payment. There are also recoverable and non-recoverable payments called Special Needs Grants that provide financial support for people incurring large one-off costs for specific goods or services. The tax system also plays a key role in the provision of income support for individuals receiving welfare payments. This is particularly the case in relation to the Best Start payment and the Family Tax Credit.

TAS provides an important last resort payment for individuals facing financial hardship due to high essential costs related to housing, health, disability and consumer durables. Most TAS recipients are also receiving a main benefit, but this is not a condition of eligibility.

To receive TAS a person must fill out an application form and provide information on their circumstances, income, and costs.

To be eligible a person must meet residency requirements. There is also a cash assets test which in November 2018 meant that a single person had to have less than $1076.96 and a couple less than $1,794.51. Individuals must also be making reasonable efforts to reduce their expenditure and increase their income. Part of this requirement is that people must claim other relevant payments including the Accommodation Supplement, Disability Allowance and the Winter Energy Payment.

The steps in the TAS payment calculation are set out in table 1 below. The payment formula is based around the concept that claimants should have a basic minimum amount of money after paying for their 'essential costs'. The amount of this basic minimum is referred to as 'standard costs' and is calculated as 70 percent of the relevant main benefit and unabated Family Tax Credit. If a person has less money than their estimated 'standard costs' after paying for their 'essential costs' then the TAS payment is available to make-up this shortfall. The maximum payment rate is 30 percent of the relevant main benefit rate, with an additional increment for those also receiving the Disability Allowance.

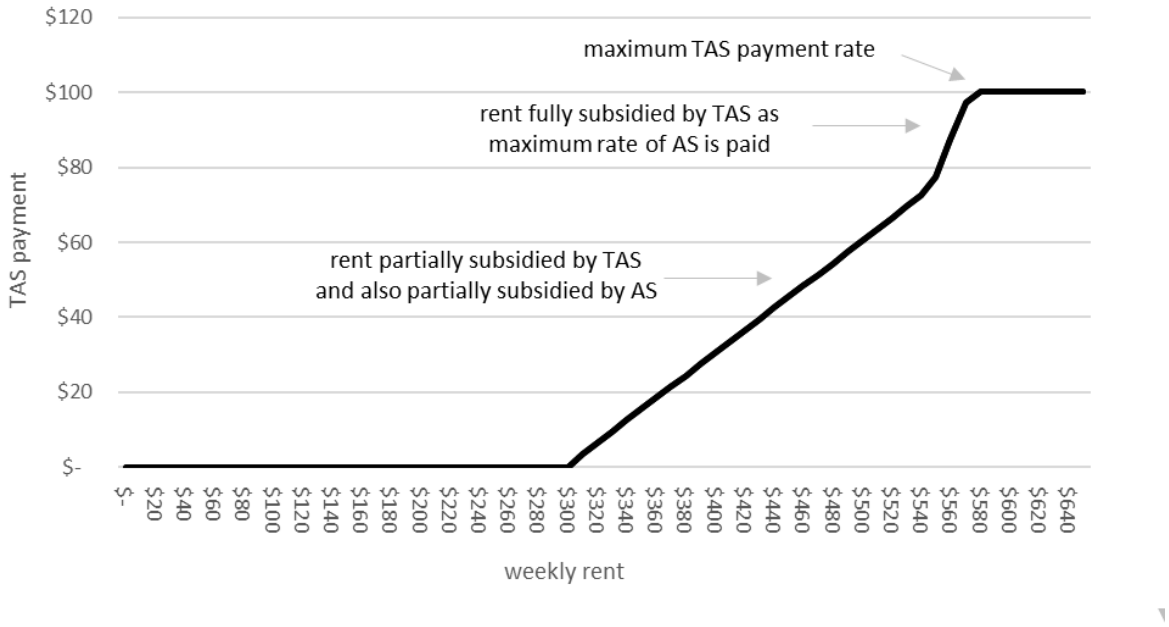
Table 1: Steps in the TAS calculation

Step 1	Determine potential eligibility	Determine eligibility related to residency, cash assets and reasonable efforts.
Step 2	Assess income	Income from all sources including main benefit, supplementary payments, tax credits, and earnings etc.
Step 3	Assess essential costs	Essential costs are defined to include those related to housing, health and disability, childcare, and costs related to household durables such as kitchen appliance, beds and laundry. There are specified

		weekly maxima for many of these elements. Costs are verified with receipts or information from a medical practitioner.
Step 4	Calculate disposable income	Disposable income = income - essential costs
Step 5	Calculate standard costs	Standard costs are 70% of rate of main rate and the relevant unabated rate of the family tax credit.
Step 6	Calculate TAS deficiency	TAS deficiency = standard costs - disposable income
Step 7	Calculate value of TAS payment	The TAS payment = either: • $0 (if the TAS deficiency is less than $1, or they are receiving Special Benefit); or • TAS deficiency (if less than maximum payment); or • maximum payment (maximum payment rate is 30% of the relevant main benefit rate, with an additional increment for those also receiving the Disability Allowance).

Figure 1 sets out an example of how the rate of payment changes as a person's essential costs increase. The figure shows the relationship between rent and the rate of the TAS payment for a sole parent with two children living in Palmerston North. As can be seen, the value of the payment increases up to a maximum of $100 in line with housing costs.

Figure 1: Relationship between rent and TAS for a Sole Parent Support recipient with two children living in Palmerston North, November 2018



Note: The modelled case assumes $20 of other essential costs.

Table 2 sets out the characteristics of TAS recipients in November 2018. At that time there were 59,495 recipients with 86 percent receiving income-tested main benefits. Around 28 percent of recipients were also supporting dependent children. On average each recipient received $52 per week, although many people also received quite small payments.

Table 2: Number of TAS recipients and value of payments, November 2018

Category	Number	Percentage of recipients	Mean weekly payment	Lower quartile weekly payment	Upper quartile weekly payment
Income-tested main benefit	51,285	86%	$50	$24	$65
NZS or Veterans Pension	6,184	10%	$58	$23	$89
Non-beneficiaries	2,026	3%	$73	$54	$100
Total	59,495	100%	$52	$24	$69
No dependent children	41,760	70%	$49	$23	$65
With dependent children	17,735	30%	$59	$29	$100
Total	59,495	100%	$52	$24	$69

Source: MSD data as at 2 November 2018

Important context is that both the number of people receiving TAS, as well as the average value of payments, was lower in November 2018 compared to the year previously. This was mainly driven by the changes to the Accommodation Supplement that were implemented as part of the 'Families Package' in April 2018. The rise in the maximum rates of payment for the Accommodation Supplement increased the disposable incomes of individuals with high housing costs. Although there was a net gain in payments for this group, the rise in income was partly offset by reduced TAS payments.

Estimating the rate of 'take-up' of Temporary Additional Support

A key issue for all welfare payments is that some people might not apply for assistance despite being eligible. This can occur because potential recipients are not aware of payments, transaction costs associated with applying for support, compliance costs associated with maintaining eligibility, stigma associated with receiving payments, preferences among potential applicants, or poor administration on the part of those administering the payment (Currie, 2004; Hernanz et al., 2004).

Studies of take-up in other countries often find evidence that the number of people receiving welfare payments is substantially smaller that the population who are potentially eligible (Bargain et al., 2012). However, there is often considerable uncertainty associated with these estimates of take-up because the underlying data used to estimate the population who are eligible is not always up-to-date, comprehensive, or accurate.

The TAS payment has many characteristics that will likely cause low rates of take-up. It is likely that the complexity of the payment makes it very difficult for potential claimants to understand if they are eligible. The payment needs to be renewed every 13 weeks, and there is a requirement to inform the Ministry if there are changes to costs, income, or personal circumstances. For some people the value of the payment might be small relative to the perceived effort. In addition, some people may experience fear and stress when applying, sometimes as a result of difficulties filling out the application form (Errington and Human, 2019).

One means of estimating take-up of TAS is to model eligibility using the Ministry of Social Development's microsimulation model called MSIM.[1] The base population for the MSIM model comprises all individuals who receive a payment from the Ministry of Social Development in any given week. For each person the MSIM model contains administrative data from their current and past interaction with the welfare system. This data includes information on demographic characteristics (age, partnership status, number of dependent children etc), costs, payments, and other income. The MSIM model is a 'rules engine' (not a predictive model) and simulates eligibility and the level of payment by applying the rules of the welfare system. The model uses only actual data and imputes a limited number of characteristics where the information is not available. For the most part the calculation of TAS eligibility uses the recorded costs related to housing and health as the cost driver of eligibility for the payment.

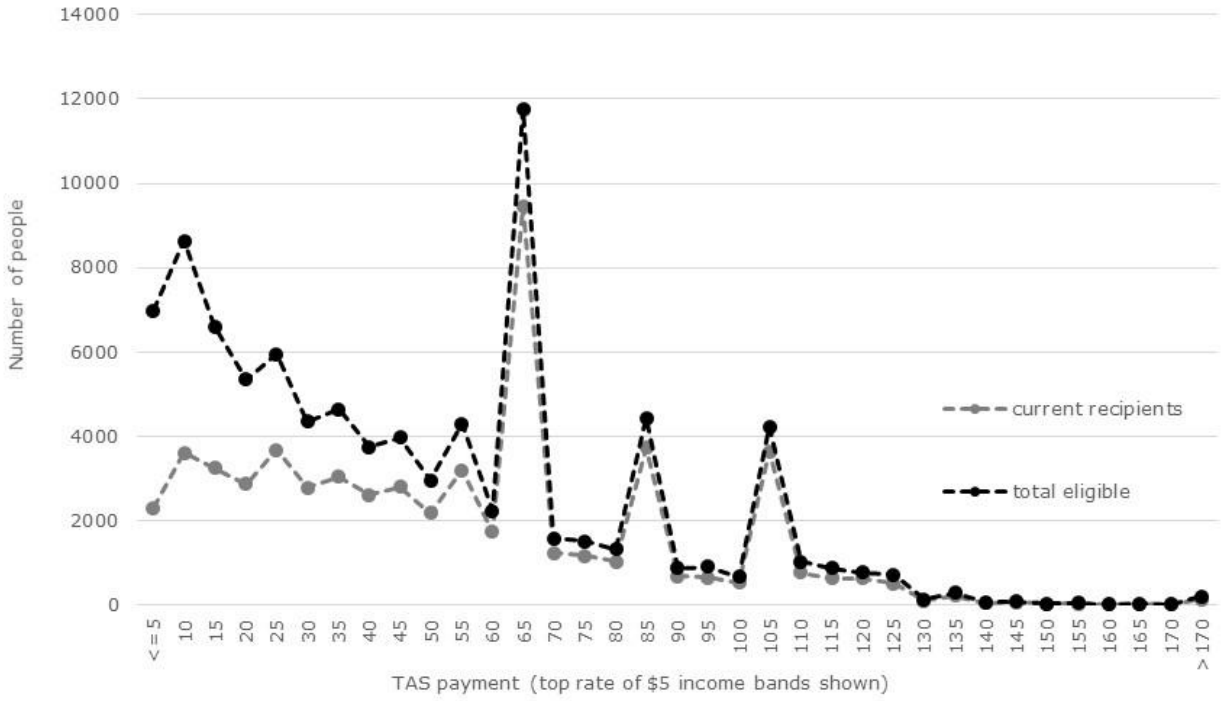
An important caveat is that the model's estimate of the amount of TAS that a person might be eligible for is not always accurate, and this is more pronounced for some sub-groups. Inaccurate estimates occur because the base information for some recipients may not be recorded. For example, for people not receiving TAS, data on some types of essential costs, particularly those related to consumer durables, are not recorded in the data. In addition, information on earnings for some groups (such as NZS recipients who are not required to report earnings) may also be missing.

[1] MSIM is short for microsimulation

In November 2018 the MSIM model estimated that take-up of the TAS payment for people 18 to 64 years in receipt of income-tested main benefits was around 68 percent. This measure of take-up is the percentage of people who are modelled as eligible for TAS who receive the payment. It is important to stress that there is some degree of uncertainty around this estimate because of the limitations of the data used in the modelling.

Figure 2 sets out the distribution of current and estimated total eligible recipients by the level of the estimated weekly TAS payment. This shows that take-up is particularly low when the predicted amount of the TAS payment is small. This gives some confidence in the simulation results as it is consistent with the notion that transactions and compliance costs deter individuals from applying when the resulting payments are smaller.

Figure 2: Current and total eligible TAS recipients in November 2018



Note: The spikes in the distribution occur because of maximum rates of TAS for some categories of recipients.

Overview of the TAS Campaign (2018)

The Ministry of Social Development has in recent years implemented a range of initiatives to respond to the low take-up of the TAS payment. These have involved provision of public information and ensuring that frontline case managers make application forms available. In previous years there has also been automated proactive contact of people who might be eligible using a mail-out of around 1,000 letters per month to people assessed as potentially eligible.

The monthly mail out stopped in June 2018 because changes in the structure of benefit payments meant new computer code was needed to identify eligible individuals. The project team tasked with rebuilding a new automated process came to the view that the Ministry's MSIM model should be used for this task. There was also consideration as to whether people should be contacted and informed about their potential eligibility by email and outbound phone calls as well as letters.

An important early issue for the project was whether to implement a business process or undertake a staged approach involving a trial. A key concern with full implementation was that if the MSIM model was not accurate then this might falsely raise expectations about increased payments and generate applications that were subsequently declined. A trial was also deemed necessary to ensure that the type of proactive contact used in the final business process was effective.

After consultation it was decided that there was enough uncertainty to warrant designing and implementing a trial to answer questions about the accuracy of the microsimulation model and the effectiveness of different forms of contact. The proposal was discussed with the National Beneficiaries Advocacy Consultative Group, and the design of the campaign was also subsequently approved by both the Ministry's independent ethics panel and the internal privacy and human rights assessment process.

The TAS campaign commenced in early November 2018 and contacted 3,000 people who had the highest level of financial need as reflected in their modelled TAS deficiency (as defined in the TAS payment calculation). Individuals were contacted using either a letter, email, or phone call. The letter was both a hardcopy sent to the recipient mail address as well as an electronic version that was viewable in myMSD. The email was sent to the recipient's private email address from the Ministry. The phone call was from the MSD call centre to the individual's preferred phone number. The three different forms of contact differed in overall unit costs with emails being least expensive, followed by letters and phone calls.

Each of the different forms of contact was designed to be as personalised as possible, easily understood, and have a clear message about how to apply for the 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.

The campaign was designed so that after initial contact the same process was used irrespective of whether the person had rung the call centre after being contacted by email or letter, or they had been phoned proactively by the call centre. After initial contact, the subsequent process involved the call centre operator assessing a person's eligibility using an online tool that recorded essential costs and income. Where this showed a person being eligible, they were then asked if they wanted to apply and were

sent application forms. After filling out the application form and providing relevant documentation an applicant then had to attend an interview at a local MSD office. Individuals were then granted the payment if they were eligible. There was also a well-defined process and training for call centre staff who were handling outbound and inbound calls about the payment.

> 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 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

Table 3 shows the number of people who were assigned and later contacted. Initial assignment occurred at the end of the first week of November 2018 and lists of 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 and phone lists were not contacted as they had cancelled their benefit in the intervening days between assignment and the time they were scheduled to be contacted.

There was also some monitoring of the extent to which contacts were received by recipients. At least 278 letters were read electronically, although it would be expected that many of the hard copies would also have been read. Approximately 64 percent of the emails that were sent were recorded as being opened by the recipient. Of people who were phoned only around 76 percent could be reached (and in a third of cases this was a message left on an answering machine).

Table 3: Overview of campaign contacts

	Individuals assigned	Adjusted to account for exits from benefit[#]	Indicative contact recorded
Letters	1,200	1,200	278*
Emails	1,200	1,125	718
Phone call	600	590	450[†]

Source: MSD data. Note: [#]exits from benefit between data extract and commencement of campaign. *record of opening of the electronic letters so likely under-represents the true level of contact. [†]includes 161 contacts that were a message left on an answerphone.

Research strategy

Research questions

The TAS campaign was designed to answer three questions:

- Does proactive contact improve take-up of TAS?
- If so, what type of contact is most effective?
- Is the MSIM model an accurate means of identifying people who are eligible for TAS?

Key measures for the trial were if someone was granted TAS in the seven weeks after being identified as eligible, as well as the weekly dollar value of any subsequent TAS payment.

The issue for measuring the impact of the campaign on these outcomes is that in the absence of the campaign, a proportion of people identified as eligible would have been granted TAS anyway.

In the previous months approximately 5 percent of individuals identified as eligible were granted the payment within seven weeks after being identified. The value of these payments averaged roughly $46 per week.

If the intervention was successful (ie the model identified people who were eligible for the payment and the proactive contact resulted in applications and grants of the payment), then this would increase the proportion of people who were subsequently granted TAS. Importantly, this is over and above what would have occurred in the absence of the intervention.

Conceptual design of the trial

Normally a randomised control trial design would have been a good option to address the research questions. However, this would have meant that some individuals experiencing quite significant financial difficulties would not have been contacted if they were randomly allocated to a control condition.

To address this issue the campaign used what we call a 'difference-in-difference' control trial design.

This involved the campaign contacting all individuals with the highest assessed level of financial need and then measuring their subsequent take-up of TAS. Individuals with a lower assessed level of financial need were not contacted, but their subsequent take-up of TAS were also measured. The impact of the campaign was measured as the difference in take-up between the two groups, after adjusting for the observed difference that had been previously observed.

Annex 1 provides more detail on the nature of causal inference with a difference-in-difference control trial design.

Financial need was measured by a person's TAS deficiency score which was calculated for everyone as part of the MSIM modelling. The campaign assigned all individuals with a TAS deficiency score of $72.45 or more per week to be contacted.

Individuals with a TAS deficiency that was at most $20 less than this cut-off were identified as the main comparison group. The observed outcomes for this group were

used to represent the counterfactual, after being adjusted for the difference in outcomes that were observed in the months prior to the trial.

The campaign was also designed to assess the effectiveness of each of the three different types of contact. For most people in the group assigned to be contacted, the type of contact was randomly assigned, and this then allowed a comparison of the relative effectiveness of the different types of contact.

Population

The population for the study – which we refer to as the 'campaign eligible population' - is a subset of people estimated by the MSIM model to be eligible but not receiving TAS. To be included in the 'campaign eligible population' a person needed to have a valid income record. They also could not have been proactively contacted about TAS in the last year, or applied for, declined, or had their TAS payment expired in the previous 120 days. Annex 2 provides more detail on these and other inclusion criteria.

All individuals in the 'campaign eligible population' were also assessed against a further set of criteria to determine if they could potentially be eligible for being randomly allocated to the three different types of contact. This was based on criteria related to having phone, email and letter contact details all recorded. A small number of individuals who were sight or hearing impaired were not eligible to be randomly assigned and were instead contacted by the most appropriate method. Further details on these criteria are also provided in Annex 2.

Analysis dataset

The analysis dataset contains snapshots of the campaign eligible population at four points in time. These dates were: 6 July 2018, 3 August 2018 and 7 September 2018 (when there was no proactive contact occurring), and 2 November 2018 (when proactive contact occurred).

The dataset is created from MSD administrative data and contains information on demographic characteristics, housing and disability costs, and all types of benefit and tax credit payments received.

Payment information is recorded at the time of the snapshot, as well as over the subsequent seven-week period. This follow-up period was chosen as it was considered that this would be enough time for people to apply and be granted TAS. Annex 2 provides more detail on the variables in the dataset.

Table 4 provides an overview of the analysis dataset. In total the dataset has 56,697 observations and contains 23,227 unique people overall.

A key issue is that the total size of each snapshot progressively increases between July and November 2018. It is difficult to be certain, but the growth in the wider group of people modelled as eligible but not receiving TAS (shown in column A) likely reflects several different factors. First, over the time period there was a growth in the population of people in receipt of benefits. Second, we suspect that the ending of the previous automated 'letters' process would also have increased non-take-up.

Column (B) of Table 4 also shows a proportionally larger growth in the 'campaign eligible population'. On top of the factors described above, this likely occurred because to be included in this group a person could not have been contacted about TAS in the last 120

days. Given the previous automated process had stopped in June this constraint covered fewer people in the later months.

Table 4: Overview of the analysis dataset

	Population (A) Total number of people not receiving but estimated eligible for TAS	Population (B) People in 'population A' who are eligible for campaign	Population (C) People in 'population B' eligible for randomisation
July 2018	27,862	11,986	8,684
August 2018	29,140	12,851	9,384
September 2018	30,585	14,365	10,581
November 2018	31,987	17,495	13,067
Total observations	119,574	56,697	41,716
Unique people	47,161	23,227	17,957

Source: MSD data

Important context for this analysis is that there were also changes to some payments in the period immediately prior to our analysis with the implementation of the 'Families Package'. One of these changes was the introduction of the Winter Energy Payment which in 2018 was paid from 1 July to 29 September.

Assignment of interventions for the campaign

Table 5 provides further information on the campaign eligible population when the campaign commenced on 2 November 2018. Of the 17,495 in the campaign eligible population on that date, it was determined that there should be 3,000 proactive contacts. This then allowed a calculation that the campaign should proactively contact individuals with a TAS deficiency score of $72.45 or more per week.

Individuals with a TAS deficiency of less than $72.45 were not contacted. In the formal analysis individuals with an estimated TAS deficiency of $52.45 to $72.44 served as the control. Those with an estimated TAS deficiency of $32.45 to $52.44 were utilised for a placebo test.

Table 5: Population eligible for the campaign in November 2018

		Population (B) Eligible for the campaign	Population (C) Eligible for campaign and randomisation	Population (D) Eligible for campaign but not randomisation
$1.00 to $32.44	No contact	11,054	8,149	2,905
$32.45 to $52.44	No contact (placebo)	2,204	1,696	508
$52.45 to $72.44	No contact (control)	1,237	932	305
$72.45 and above	-email	1,200	1,191	9
	-letter	1,200	521	679
	-phone	600	578	22
	Total treatment	3,000	2,290	710
Total people		17,495	13,067	4,428

Source: MSD data

Of the 3,000 individuals whose TAS deficiency score was $72.45 or above:

- 2,290 people were eligible to be randomised into one of three different types of contact because they had full 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.

Allocation to the eligible for randomisation group prior to the campaign meant that the intention to treat impacts were for the population with full contact details and with no restrictions on type of contact.

Randomisation was undertaken in SAS using different sampling probabilities to ensure that the total allocation of people to each channel met the predetermined operational target of 1200 letters, 1200 emails and 600 phone calls. After identifying the total number of slots available after the non-randomised individuals were allocated, the sampling probabilities were calculated to be letters (p=0.23), emails (p=0.52) and phone calls (p=0.25).

Following this procedure, deterministic lists of people to contact were then provided for a mail out, to a centralised email system, and to the MSD contact centre for outbound phone calls.

Annex 3 provides a breakdown of the characteristics of the randomised treatment group across the different forms of contact. As would be expected given the randomisation there is no evidence of imbalance.

Outline of the analysis

Our main results use only the population who are eligible for randomisation across all periods. We have also analysed impacts using the total 'campaign eligible' population, but none of these show material differences to the main results.

The analysis in this paper utilises an 'intention to treat' measure of impacts. However, because the campaign collected aggregate information on actual contact we are also able to scale our estimates to give a sense of the magnitude of impacts among people who were contacted.

The first part of our analysis looks at the questions about the relative impact of proactive contact. We provide descriptive results, as well as more formal analysis using a standard regression set-up for difference-in-difference analysis. The control group of this formal approach is individuals with a TAS deficiency that was at most $20 less than the $72.45 cut-off for eligibility.

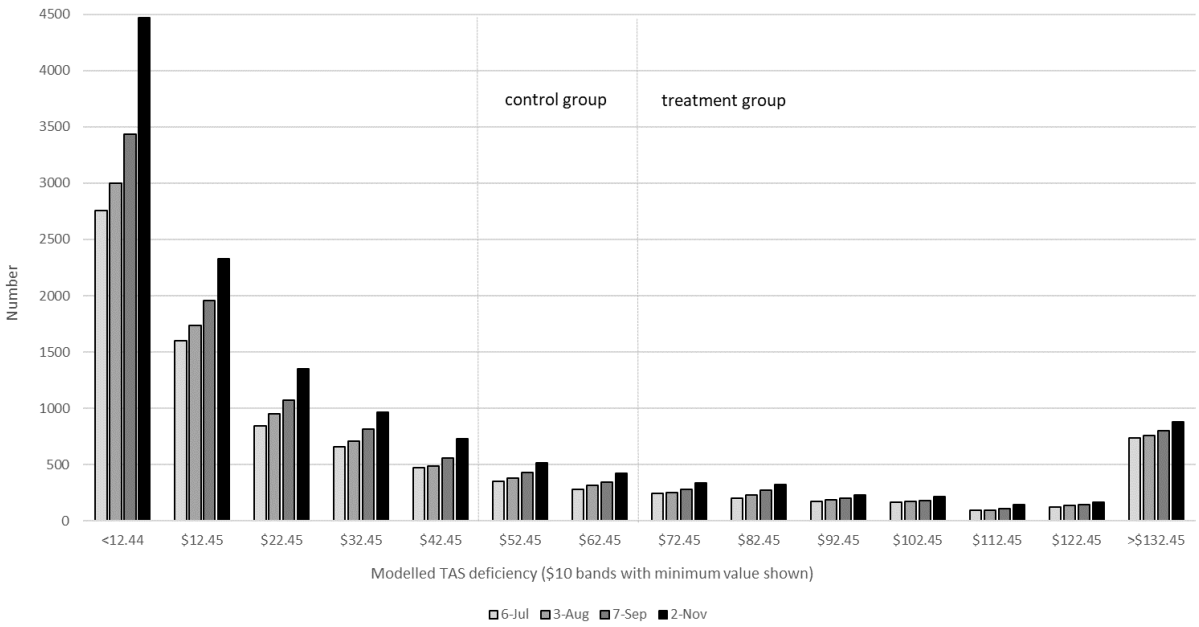
The second part of the analysis looks at the last research question about the ability of the MSIM model to successfully identify eligible individuals. As well as drawing on the impact results, we look at a subset of data of people who were phoned and assessed as eligible.

Descriptive results

In what follows we provide some descriptive graphical analysis of outcomes for the trial. This descriptive analysis uses people in the 'eligible for the campaign' population who also met the criteria for random assignment. Individuals are classified into TAS deficiency bands across the four different snapshots.

Figure 3 shows the number of people across each of the four snapshots in these bands. Roughly 22 percent of this population had a modelled TAS deficiency of more than $72.45 and in November this group were contacted.

Figure 3: Number of people eligible for TAS but not receiving the payment, by modelled TAS deficiency (July, August, September, and November 2018)



Note: The graph shows people in population C who are both 'eligible for the campaign' and meet criteria for random assignment

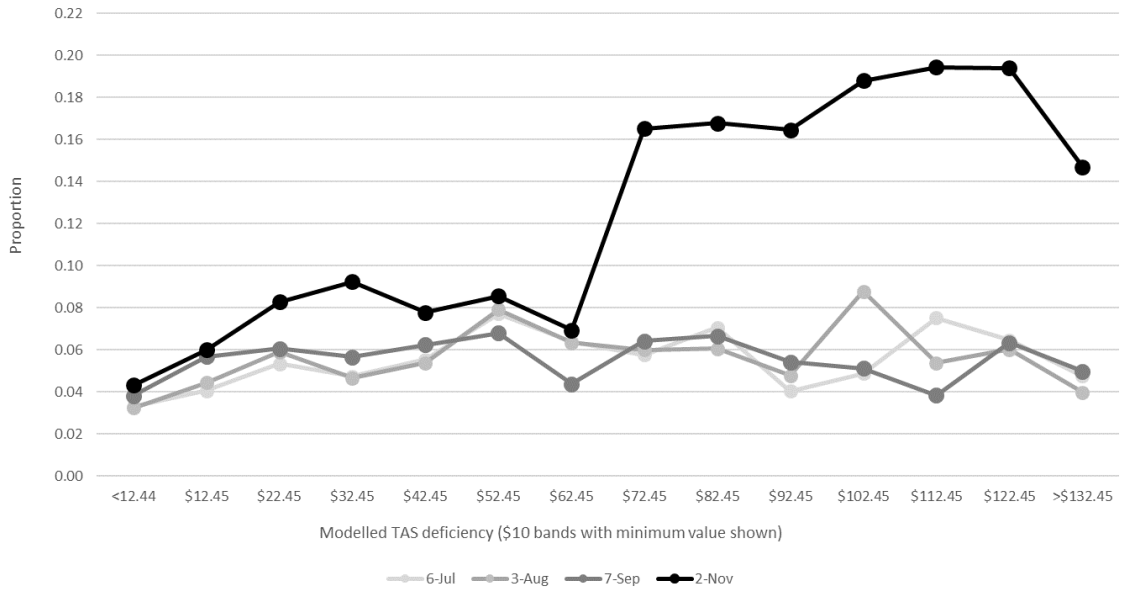
As can be seen, there was a steady increase in the number of people in this population across the months studied, and this was particularly apparent for those eligible for small TAS payments. As discussed previously, this increase likely reflected both an underlying increase in the population, as well as campaign selection criteria related to not having any contact with the Ministry about TAS in the last 120 days.

Figure 4 shows the proportion of people who went on to apply and be granted a TAS payment within the subsequent seven weeks across the different deficiency bands. As can be seen there was an increase in the proportion granted TAS from the November snapshot. Among those with a modelled TAS deficiency of more than $72.45 the rate of being granted TAS increases from an average of around 6 percent to around 17 percent

Another feature is that there is some evidence of a general lift of around 2 percent in the proportion being granted TAS among those who were *not* contacted in November (a TAS deficiency of $72.45 or less). This could be due to seasonal differences, a change in the composition of people in the eligible population, or an increase in more 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 hence other service delivery changes were not unexpected.

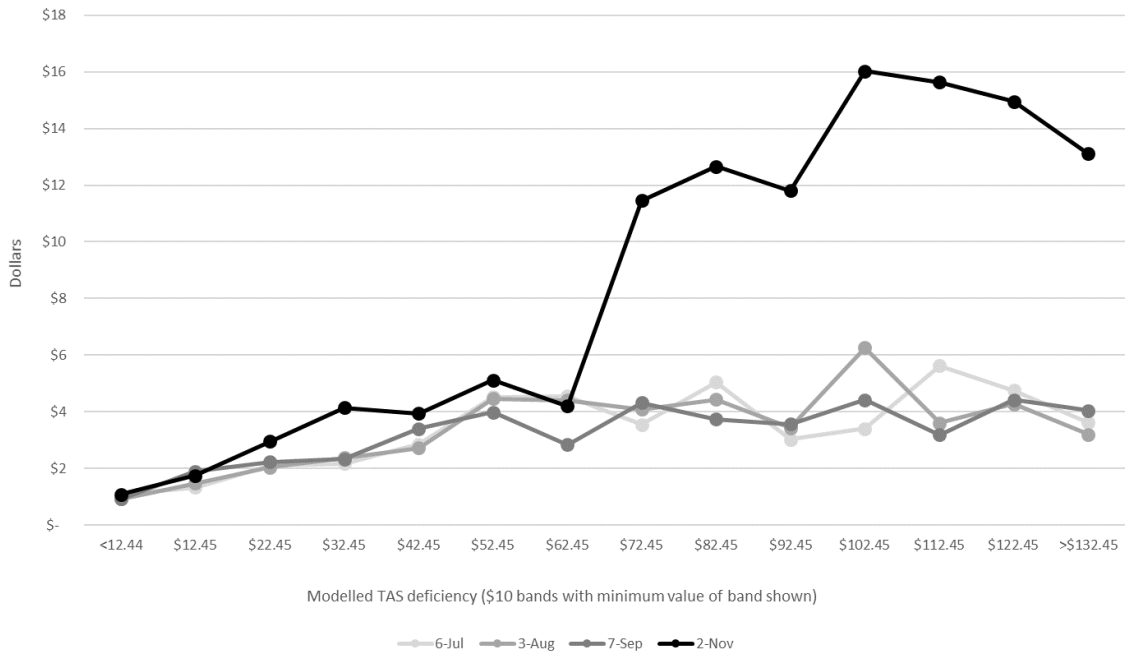
Figure 4: Proportion of people granted TAS within subsequent seven weeks, by modelled TAS deficiency (July, August, September and November 2018)



Note: The graph shows people in population C who are both 'eligible for the campaign' and meet criteria for random assignment

Figure 5 shows the average weekly amounts of TAS income after seven weeks for each of the groups described above. Note this is averaged across everyone in each TAS deficiency band, irrespective of whether they received TAS. This also shows a clear increase in TAS income for those contacted.

Figure 5: Average weekly TAS income after seven weeks, by modelled TAS deficiency (July, August, September and November 2018)



Note: The graph shows people in population C who are both 'eligible for the campaign' and meet criteria for random assignment

As context it is also useful to note that after the campaign commenced in November, for individuals in the treatment group who were subsequently granted TAS the average payment was around $80 per week.

It is also useful to note that for those granted TAS the amount of the payment was reasonably close to the level predicted by the MSIM model. Table 6 shows this across the different TAS deficiency bands.

Table 6: Predicted and actual TAS payment among all individuals granted TAS (July, August, September, and November 2018)

MSIM estimated TAS deficiency band	Average MSIM estimated TAS entitlement	Actual TAS payment for those granted
Less than $12.45	$7	$27
$12.45 to $22.44	$17	$32
$22.45 to $32.44	$27	$36
$32.45 to $42.44	$38	$45
$42.45 to $52.44	$47	$52
$52.45 to $62.44	$56	$58
$62.45 to $72.44	$64	$66
$72.45 to $82.44	$70	$68
$82.45 to $92.44	$76	$71
$92.45 to $102.44	$77	$71
$102.45 to $112.44	$85	$81
$112.45 to $122.44	$82	$78
$122.45 to $132.44	$79	$75
Greater than or equal to $132.45	$87	$85
Total	$41	$49

Source: MSD data. Note 1: The table uses population C who are 'eligible for the campaign' and meet criteria for random assignment. Note 2: The rate of TAS is calculated as a person's TAS deficiency up to various maximum rates of payment. The first column is the TAS deficiency estimated by the MSIM model. The second column is the average estimated TAS entitlement for each band. This reflects the fact that at higher TAS deficiencies payments are constrained by various maximum rates. The third column is the actual TAS payment calculated for those who applied and were granted TAS.

Formal analysis of impacts

The graphical representation of the impacts of the campaign set out in figures 3 and 4 summarise the increase in TAS payments resulting from the campaign. In this section we more formally quantify and assess the robustness of these impacts, and also show how the magnitude of the impacts differed by the type of contact.

For the formal analysis we use the standard difference-in-difference approach that the campaign was designed to utilise.

The main outcomes estimated are both the probability of a TAS grant within seven weeks, as well as the value of TAS payments. In addition, we also look at the impact of the campaign on the probability of remaining in receipt of any payments as it might be expected that increased payments might have increased the financial incentive to remain in receipt of benefits.

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).[2]

Our basic model has the form:

$$Outcome_{it} = \beta_0 + \beta_1 NOV_{it} + \beta_2 T_{it} + \beta_3 (NOV_{it} * T_{it}) + X_{it}'\beta_4 + \varepsilon_{it}$$

The dummy variable 'NOV' defines the November time period when the campaign occurred. The variable 'T' identifies that a person had a TAS deficiency of more than $72.45 and was part of the potential and actual treatment group. The matrix 'X' includes individual month effects, a variety of demographics characteristics, and a variety of TAS and other benefit-related controls. The difference-in-difference estimator β_3 identifies the impact of the campaign conditional on the covariates.

An important extension of the model is that we are also able to estimate the effect of each type of treatment. This is implemented by interacting the NOV*T variable with separate indicators for the different randomly assigned treatments.

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.

In what follows we firstly describe the characteristics of individuals in the dataset and set out the formal results. In subsequent sections we then assess the assumptions and robustness of the identification strategy.

Characteristics of control and treatment groups

Table 7 sets out the broad characteristics of the treatment and control group in the months pre and post the campaign. We return to the analysis of these characteristics when we investigate the validity of the assumptions of the trial.

[2]We constrained the control group to individuals with an estimated TAS deficiency of up to $20 less than the $72.45 threshold on the basis that they will more likely resemble the treatment group. However, in practice widening the control group to also include individuals in the campaign population with a TAS deficiency of less than $52.45 does not change the results.

Table 7: Characteristics of people in the formal analysis dataset

	Control (pre)	Treatment (pre)	Control (post)	Treatment (post)
Men	0.40	0.41	0.37	0.39
	(0.49)	(0.49)	(0.48)	(0.49)
Women	0.60	0.59	0.63	0.61
	(0.49)	(0.49)	(0.48)	(0.49)
Māori	0.28	0.23	0.30	0.26
	(0.45)	(0.42)	(0.46)	(0.44)
European	0.40	0.42	0.38	0.42
	(0.49)	(0.49)	(0.49)	(0.49)
Pacific	0.05	0.06	0.06	0.06
	(0.22)	(0.23)	(0.23)	(0.24)
Other ethnicity	0.24	0.26	0.22	0.24
	(0.43)	(0.44)	(0.42)	(0.43)
Unspecified ethnicity	0.03	0.03	0.03	0.03
	(0.18)	(0.17)	(0.18)	(0.17)
Under 25 years of age	0.09	0.05	0.10	0.06
	(0.29)	(0.22)	(0.30)	(0.23)
Age 25 to 44 years	0.40	0.32	0.42	0.37
	(0.49)	(0.47)	(0.49)	(0.48)
Age 45 to 64 years	0.38	0.46	0.37	0.42
	(0.49)	(0.50)	(0.48)	(0.49)
Above 65 years of age	0.12	0.17	0.11	0.16
	(0.33)	(0.37)	(0.31)	(0.37)
Single no dependent children	0.48	0.55	0.46	0.53
	(0.50)	(0.50)	(0.50)	(0.50)
Couple no dependent children	0.19	0.22	0.17	0.19
	(0.40)	(0.41)	(0.37)	(0.39)
Couple with dependent children	0.07	0.07	0.06	0.07
	(0.26)	(0.25)	(0.25)	(0.25)
Sole parent with dependent children	0.25	0.16	0.31	0.21
	(0.43)	(0.37)	(0.46)	(0.41)

Table 7: Characteristics of people in the formal analysis dataset (continued)

	Control (pre)	Treatment (pre)	Control (post)	Treatment (post)
Job Seeker Support payment	0.42	0.40	0.41	0.39
	(0.49)	(0.49)	(0.49)	(0.49)
Supported Living Payment	0.16	0.19	0.15	0.17
	(0.37)	(0.39)	(0.36)	(0.37)
Sole Parent Support	0.17	0.11	0.22	0.17
	(0.37)	(0.31)	(0.41)	(0.37)
NZS or Veterans Pension	0.13	0.17	0.11	0.15
	(0.33)	(0.37)	(0.32)	(0.36)
Non-beneficiary	0.10	0.12	0.07	0.11
	(0.30)	(0.32)	(0.25)	(0.31)
Modelled TAS deficiency	$61.28	$132.50	$61.16	$129.57
	(5.59)	(57.18)	(5.64)	(55.61)
Modelled TAS payment	$60.52	$88.06	$60.42	$87.32
	(5.50)	(33.04)	(5.52)	(28.76)
Assets	$34.82	$40.17	$22.47	$37.06
	(182.46)	(200.69)	(142.56)	(194.77)
TAS payment in previous 24 months	0.41	0.39	0.47	0.43
	(0.49)	(0.49)	(0.50)	(0.50)
Total observations (N)	2,095	5,529	932	2,290

Source: MSD data. Note: Population 'C' restricted to individuals with a TAS deficiency of $52.45 or more. Control are individuals with TAS deficiency of between $52.45 and $72.44. Treatment group has a TAS deficiency of $72.45 or more. 'Pre' refers to July, August and September 2018, while 'post' refers to November 2018. Standard deviations in parenthesis.

Impacts

Our analysis of impacts uses a standard difference-in-difference set-up to formally quantify the impacts.

Table 8 reports estimates of the impact of the campaign on the proportion of the treatment group who were subsequently granted TAS within seven weeks. Our preferred estimate is model 4 which controls for individual month effects, demographics and benefit related characteristics. This shows that the 'email' increased the proportion of people being granted TAS by 7 percentage points, the 'letter' by 11 percentage points, and the 'outbound phone call' by 17 percentage points.

Model 5 restricts the treatment group sample to only those individuals who are within $20 of the threshold. This discontinuity analysis shows that after controlling for differences in covariates, the local average treatment effect around the threshold was broadly in line with the impact observed across the wider sample, however the estimated impact of the 'email' was reduced.

Table 8: Estimated impact of the campaign on the proportion of people granted TAS within 7 weeks

	Model 1	Model 2	Model 3	Model 4	Model 5
Total impact (β_3)	0.10***	0.10***	-	-	-
	(0.01)	(0.01)	-	-	-
Email (β_{3e})	-	-	0.06***	0.07***	0.04**
	-	-	(0.01)	(0.01)	(0.02)
Letter (β_{3l})	-	-	0.11***	0.11***	0.14***
	-	-	(0.02)	(0.02)	(0.04)
Phone (β_{3p})	-	-	0.16***	0.17***	0.17***
	-	-	(0.02)	(0.02)	(0.04)
Month, demographic and benefit controls	No	Yes	No	Yes	Yes
r^2	0.03	0.05	0.03	0.05	0.04
N	10,846	10,846	10,846	10,846	5,154

Source: MSD data. Note: Population 'C' restricted to individuals with a TAS deficiency of $52.45 or more. Model 5 has restricted treatment group. *** p value<0.01 **p value <0.05 *p value <0.1. Impact estimated with a linear probability model. Standard errors are clustered on individuals. The magnitudes and significance of the parameters were similar when estimated using logistic regression.

Table 9 reports estimates for how the campaign changed average weekly TAS payments. This 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 the 'email' and the 'phone call'.

Table 9: Estimated impact of the campaign on income from TAS ($ per week 7 weeks after the campaign commenced)

	Model 1	Model 2	Model 3	Model 4	Model 5
Total impact (β_3)	$8.72***	$8.90***	-	-	-
	(0.98)	(0.97)	-	-	-
Email (β_{3e})	-	-	$5.73***	$5.93***	$3.49**
			(1.09)	(1.09)	(1.49)
Letter (β_{3l})	-	-	$8.92***	$9.08***	$9.70***
			(1.55)	(1.55)	(2.49)
Phone (β_{3p})	-	-	$14.70***	$14.70***	$13.93***
			(1.70)	(1.70)	(2.75)
Month, demographic and benefit controls	No	Yes	No	Yes	Yes
r^2	0.03	0.05	0.04	0.06	0.05
N	10,846	10,846	10,846	10,846	5,154

Source: MSD data. Note: Population 'C' restricted to individuals with a TAS deficiency of $52.45 or more. Model 5 has restricted treatment group. *** p value<0.01 **p value <0.05 *p value <0.1. Impact estimated with OLS. Standard errors are clustered on individuals.

We also looked at the extent to which the campaign changed the proportion of people continuing to receive any form of income support payment after seven weeks. Prior to the trial around 95 percent of all individuals would still be in receipt of any welfare payments after seven weeks. It might be expected that the financial incentives associated with increased uptake of TAS would have meant that this rate would have been higher for those who were proactively contacted during the trial period. However, as shown in table 10 there was some very weak evidence that the relationship may have been the opposite. Although not significant, the estimates suggest the campaign reduced the proportion of people in receipt of payments in the treatment group by around 1 percentage point (model 2). This small impact was also larger for individuals who received the more effective forms of contact. Individuals who received outbound phone calls were 2 percentage points less likely to be receiving any payment after the campaign (model 4). We consider that this unexpected finding could have been caused by a small sub-group of individuals who were not eligible for *any* payments cancelling their benefits because of being contacted. It should also be noted that seven weeks is a relatively short time-period over which to observe these effects.

Table 10: Estimated impact of the campaign on the proportion of people receiving any income support payments after 7 weeks

	Model 1	Model 2	Model 3	Model 4	Model 5
Total impact (β_3)	-0.01 (0.01)	-0.01 (0.01)	- -	- -	- -
Email (β_{3e})	-	-	0.00 (0.01)	0.00 (0.01)	-0.02 (0.02)
Letter (β_{3l})	-	-	-0.02 (0.02)	-0.02 (0.01)	-0.01 (0.02)
Phone (β_{3p})	-	-	-0.02 (0.01)	-0.02* (0.01)	-0.04 (0.02)
Month, demographic and benefit controls	No	Yes	No	Yes	Yes
r^2	0.00	0.04	0.00	0.04	0.04
N	10,846	10,846	10,846	10,846	5,154

Source: MSD data. Note: Population 'C' restricted to individuals with a TAS deficiency of $52.45 or more. Model 5 has restricted treatment group. *** p value<0.01 **p value <0.05 *p value <0.1. Impact estimated with a linear probability model. Standard errors are clustered on individuals. The magnitudes and significance of the parameters were similar when estimated using logistic regression.

As part of the analysis we also looked at the impacts of the trial for Māori (versus non-Māori). The results across each of the outcome areas were broadly similar between these two sub-groups. We also looked at impacts separately by gender and found larger impacts for women compared to men. This was mostly driven by low uptake of TAS by men when sent letters or emails. There was a similar response for phone calls.

Evidence of a constant difference prior to the trial commencing

The difference-in-difference analysis of the impact of the trial assumes there was a constant difference in outcomes between the treatment and control groups in the pre-trial period.

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

$$Outcome_{it} = \beta_0 + \beta_1 AUG_{it} + \beta_2 SEP_{it} + \beta_3 T + \beta_4 (AUG_{it} * T) + \beta_5 (SEP_{it} * T) + X_{it}' \beta_6 + \varepsilon_{it}$$

Within this set-up the parameter β_3 measures the difference between the treatment and control groups. We use the standard 'parallel trends' approach of an F-test of the linear restriction that the estimate β_4 and β_5 are jointly zero.

As reported in table 11, once controlling for covariates there was no practical or statistically significant difference between the treatment and control group. Importantly, there was no evidence to contradict the assumption of a constant difference.

Table 11: Test of constant difference between treatment and control groups in the pre-period

	TAS grants model 1	TAS payment model 2	Benefit receipt model 3
Treatment group difference (β_3)	-0.01	-$0.25	-0.01
	0.01	$0.83	0.01
Difference-in-difference in August (β_4)	0.00	$0.15	0.01
	0.01	$0.88	0.01
Difference-in-difference in September (β_5)	0.02	$1.21	0.00
	0.01	$1.01	0.01
p value of f-test that β_4 and β_5 =0 (parallel trends)	0.55	0.41	0.71
Demographic and benefit controls	Yes	Yes	Yes
r^2	0.03	0.02	0.04
N	7,624	7,624	7,624

Source: MSD data. Note: Population 'C' restricted to individuals with a TAS deficiency of $52.45 or more. Model 3 has restricted treatment group. *** p value<0.01 **p value <0.05 *p value <0.1. Impacts estimated with OLS. Standard errors are clustered on individuals. Demographic and benefit covariates relating to sex, ethnicity, age group, family type, benefit type, and TAS-related variables. For model 1 and 3 the results were comparable when estimated using a logistic model.

We had originally planned to undertake the novel step of testing the assumption of a constant difference prior to commencing the campaign. However, the rapid implementation of the trial and some small delays in assembling the historical data meant that we were only able to conduct this analysis after the trial had been undertaken.[3]

How reasonable is the assumption of that the constant difference would have continued in the absence of treatment?

The difference-in-difference analysis makes the unverifiable assumption that the previously observed difference in the outcomes between the treatment and control group (after controlling for observable characteristics) would also have continued in the time-period during which the treatment was implemented.

[3]Testing the assumption of a constant difference prior to undertaking the trial is not usually done, but we suggest it would be useful to provide assurance, prior to the trial commencing, that the trial is feasible and sufficiently powered. This would reduce the risk of undertaking a trial that is unable to identify an impact because there is not a constant difference. Relatedly, it would also reduce the risk of undertaking a trial with an insufficient sample size because an analysis of pre-trial trends provides data for an ex ante estimate of the standard error of the difference-in-difference estimator. These risks did not eventuate for the trial as implemented. An additional benefit of testing the assumption of a constant difference prior to the trial commencing is that it establishes the functional form of the regressions for the ex post analysis. It is useful to note in this regard that we used the functional form that we originally planned.

There are two types of reasons why this assumption of no unobserved confounding might not be correct:

- the intervention may have been assigned to the treatment group based on unobservable characteristics associated with the future outcome (typically referred to as endogeneity)
- other unmeasured factors, policy changes or service delivery interventions might impact on relative outcomes. Importantly, in a policy context, other interventions might be motivated by the same rationale as the intervention being tested.

While the assumption of a constant future difference in outcomes cannot be verified, it is possible to investigate the reasonableness of the assumption in different ways. This involves both assessing 'threats' in the broader environment in which the trial was conducted, as well more formal quantitative analysis.

One issue is that the campaign was conducted within an environment of increased concerns about take-up of TAS and other payments. It is possible that this led to increased conversations between staff and case managers about the payment. This could have increased take-up rates across everyone and is one possible explanation for an uptick in TAS grants for the control group during November compared to earlier months (apparent in Figures 4 and 5). If this effect did occur, we assume that it effected both the treatment and control groups equally, as information about 'treatment' status was not widely available.

We use two different tests in our formal strategy to assess the reasonableness of the no unobserved confounding assumption.

The first test looks at any statistically significant shifts in the relative balance of the characteristics described in table 8 across the treatment and control groups. While the analysis controls for changes in these observed covariates, changes in the relative balance of covariates between the pre and post periods are also suggestive that there could be unobserved confounding (Wing et al., 2018). We operationalise this with a regression of the form:

$$Covariate_{it} = \beta_0 + \beta_1 NOV_{it} + \beta_2 T_{it} + \beta_3 (NOV_{it} * T_{it}) + X_{it}'\beta_4 + \varepsilon_{it}$$

The dummy variable 'NOV' defines the November time period when the campaign occurred. The variable 'T' identifies that a person had a TAS deficiency of more than $72.45 and was part of the potential and actual treatment group. The matrix 'X' includes other demographic characteristics, and a variety of TAS and other benefit-related controls. The parameter β_3 measures any difference-in-difference in the relative balance of covariates.

Table 12 reports the estimates and standard errors for β_3 from these regressions. There are very few statistically significant changes in the relative balance of covariates, although the difference in the average TAS deficiency between the treatment and control groups did reduce slightly. We interpret these findings as not raising any major concerns about the trial assumptions.

Table 12: Difference-in-difference in covariates balance between treatment and control groups

Proportion / average $ value	Estimate difference-in-difference (β_3)	Standard error
Men	0.00	(0.02)
Women	0.00	(0.02)
Māori	0.00	(0.02)
European	0.01	(0.02)
Pacific	-0.01	(0.01)
Other ethnicity	0.00	(0.01)
Unspecified ethnicity	0.00	(0.01)
Under 25 years of age	0.00	(0.01)
Age 25 to 44 years	0.03	(0.02)
Age 45 to 64 years	-0.03	(0.02)
Above 65 years of age	0.00	(0.01)
Single no dependent children	0.01	(0.02)
Couple no dependent children	0.00	(0.01)
Couple with dependent children	0.01	(0.01)
Sole parent with dependent children	-0.01	(0.02)
Job Seeker Support payment	-0.01	(0.02)
Supported Living Payment	-0.01	(0.01)
Sole Parent Support	0.00	(0.01)
NZS or Veterans Pension	0.00	(0.01)
Non-beneficiary	0.02*	(0.01)
Average modelled TAS deficiency	-$2.81***	(0.87)
Average modelled TAS entitlement	-$0.64	(0.52)
Average assets	$9.24*	(5.17)
Proportion with TAS in prior 24 months	-0.03	(0.02)

Source: MSD data, 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

Our second formal test of the unconfoundedness assumption is a placebo test using individuals with a slightly lower TAS deficiency than the control group (from $32.45 to $52.44 per week). We would be concerned about the reasonableness of the unconfoundedness assumption if there was a difference in the relative outcomes of the control and placebo groups during the treatment period. We implement the placebo test with the usual difference-in-difference set-up using data on only individuals in the placebo and control groups.

$$Outcome_{it} = \beta_0 + \beta_1 NOV_{it} + \beta_2 PLACEBO + \beta_3 (NOV_{it} * PLACEBO) + X_{it}'\beta_4 + \varepsilon_{it}$$

Table 13 sets out the results of the placebo test regressions for both outcomes related to TAS, as well as continued receipt of payments. The null hypothesis of the test is that the parameter β_3 is zero, and there was little evidence to suggest that this was not the case.

Table 13: Placebo test regressions

	TAS grants Model 1	TAS payments Model 2	Receipt of any payments Model 3
Placebo impact (β_3)	0.02	$0.91	0.00
Clustered standard errors	0.01	$0.79	0.01
Month, demographic and benefit controls	Yes	Yes	Yes
r^2	0.02	0.02	0.03
N	8,412	8,412	8,412

Source: MSD data. Note: Population 'C' restricted to individuals with a TAS deficiency of $32.45 to $72.44. 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. The results were broadly similar when estimated using a logit for model 1 and 2, although the β3 parameter was significant at the 10% level of significance.

Assessment of the accuracy of the MSIM model

A key rationale for the campaign was to assess how accurate the MSIM microsimulation model was at identifying individuals eligible to receive TAS.

Table 14 reports the overall outcomes from the campaign for the population who met the criteria for eligibility for the campaign and randomisation. This shows that overall, 8 percent of individuals identified by the 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 percent were subsequently granted TAS.

Table 14: Outcomes from the November 2018 campaign

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

Source: MSD data. Note: Population 'C'.

The 23 percent take-up rate for the phone call establishes a conservative lower bound measure of the accuracy of the model among the treatment group for the study population. 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.[4]

A less conservative estimate of accuracy uses more detailed information recorded by the call centre for the campaign. Of all the individuals who were able to be contacted and talked to 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. Not all of this group subsequently applied for the payment. But this suggests that around 72

[4]Qualitative evidence undertaken after the trial suggests that the time and stress associated with the application process may have been a factor in this.

percent of the phone call group may have been eligible and represents an upper bound estimate of the accuracy of the model.

Table 14 also provides information on how accurate the MSIM model was in predicting the weekly amount of TAS payments. For those who ended up being granted TAS, the table reports the average amount predicted and the average amount granted. As can be seen, these were quite close suggesting that on average the model was good at predicting. However, while on average the model was close to the target, the prediction was not always precise. The average prediction error (RMSE) was around $22 for everyone in the campaign population, but slightly less at higher levels of TAS deficiency.

The campaign appeared to make a similar impact for Māori and non-Māori which suggests that the MSIM model may have been equally good at predicting eligibility for these different groups.

Overall, we conclude that the campaign demonstrated that for the population studied, the MSIM model provided valuable information for proactive targeting of messaging to improve the uptake of TAS.

Roll-out of an ongoing process of proactive contact

After analysing the result of the campaign there was a decision to implement an ongoing process of proactive contact to increase take-up of the TAS payment.

There were some changes to the process used in the campaign, and some of the information provided to potential recipients was refined following a small qualitative study (Errington and Human, 2019).

The new ongoing process uses the MSIM model to identify people who are potentially eligible. Approximately 3,000 people per month are contacted, and the campaign progressively works through the list of people eligible according to the level of modelled TAS deficiency.

Individuals are contacted by email or letters, and sometimes a combination of both. Outbound phone calls are not used due to pressures on the call centre for other work. However, clients are encouraged to phone the contact centre to discuss their potential entitlement.

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 in a number of different regions during 2019.

Conclusion

The TAS payment provides temporary financial support for people experiencing severe financial hardship. Modelling suggests there are a reasonable number of people in receipt of income-tested main benefit payments who would likely qualify for TAS but have not applied.

There are several features of the TAS payment that are likely to result in low take-up. These include the complexity of the payment (which makes it difficult to understand if you are eligible), the time-consuming nature of the application process, and the fact that the payment is only temporary and needs to be renewed every 13 weeks.

In 2018 the Ministry of Social Development designed and tested a new way of identifying and proactively contacting people who might be eligible for the payment. The Ministry's existing microsimulation model was used to calculate if current benefit recipients might be eligible for the payment. The calculation used existing data on each recipient's payments and housing costs to estimate possible eligibility for TAS.

The Ministry then conducted a campaign of contacting people who might qualify for the payment with the goals of assessing if the model was able to accurately identify people who were eligible, and to also establish what were the most effective channels for proactive contact.

A key rationale for testing the new method of proactively contacting people was to ensure that the approach was well targeted and effective, and not resulting in large numbers of people spending time making unsuccessful applications for the payment.

The campaign contacted individuals with the highest level of calculated financial need, and the impact of the campaign was measured using a difference-in-difference approach. The key findings of the trial were that:

- proactive contact increased the number of people applying for and being granted the TAS payment
- email contact increased the rate of TAS receipt by 7 percentage points, letters by 11 percentage points, and outbound phone calls by 17 percentage points
- the campaign targeted individuals with the highest levels of financial need, and for those contacted who were subsequently granted TAS the payment was around $80 per week
- in the short-term the campaign did not result in people remaining on benefit for longer, although we do not know how the campaign affected time on benefit after 7 weeks
- among the population targeted by the campaign, between 23 percent and 72 percent of those identified by the modelling were likely eligible for the payment.

The campaign has been subsequently rolled out in a modified form as a continuing business process.

Annex 1: The difference-in-difference control trial

Purpose

The difference-in-difference strategy is a well-established method for estimating causal impacts. One of the earliest documented examples is John Snow's analysis of the London cholera epidemics of 1849 and 1854. Typically, the strategy is used in observational studies, but it can also be used experimentally to assess the impacts of an intervention using what we call a 'difference-in-difference' control trial.

The advantages of experimental studies

Experimental studies are based on researchers actively designing and implementing a test of an intervention. Observational studies use historical data on interventions and outcomes to estimate causal effects.

It is often implied that the use of purposeful random assignment is the essence of the distinction between these two types of study. However, while randomised control trials are the main experimental method in public policy, it is possible for an experimental study to be conducted without randomisation. Fully controlled experiments in the physical sciences do not always require randomisation, and there are also other types of experimental studies that do not rely on randomisation as the basis for causal inference.

The essence of the distinction between an experimental and observational study is the extent to which researchers are able to plan and manage both assignment of the intervention and the collection of appropriate data (Cochran, 1965). In an experiment the assignment mechanism is known and controlled by the researchers, whereas with observational studies, the assignment mechanism is not fully known or controlled by researchers (Imbens and Rubin, 2015; Shadish et al., 2002).

Experimental studies or trials have several advantages compared to observational studies.

First, they allow the researcher to test a specific intervention with a specific population. Often this relates to assessing the effectiveness of a new intervention, or an existing intervention with a different population. Observational research is by definition constrained by the need to study previously implemented interventions.

Second, experimental studies generate evidence with high levels of internal validity because of the ability to design and document the assignment mechanism. Retrospective studies provide weaker evidence of impacts due to the absence of a designed and documented assignment of individuals to an intervention. Uncertainty about the assignment mechanism, and the possibility that selection is correlated with unobservable characteristics that matter for outcomes, is the key issue for observational research on causal impacts.

Third, an experimental trial creates an opportunity for the researcher to assess and invest in the collection of appropriate data. While this opportunity does not come with any guarantee, it stands in contrast to observational studies where the existence of a useable dataset is often a matter of luck. Observational studies are hampered by missing data on the exact nature of the assignment process, as well as important covariates and outcomes for treatment and comparison groups.

Fourth, prospective studies also have the advantage that they are designed and implemented without access to outcome data. They are a more rigorous 'test' of the intervention because there are fewer opportunities for researchers to implicitly or explicitly influence the results of the study as part of the research process (Rubin, 2008).

Despite these advantages, there are also many situations where experimental studies are not appropriate or as good as an observational study. It is not ethical to undertake some prospective studies, either because consent has not been obtained, or because the treatment or control group will be harmed. Some experimental studies are simply not feasible to implement, or can only be conducted on a narrow range of the population. In some cases an observational study may be able to assess impacts across a wide range of the population. Where there are long-term impacts an observational study may be able to produce results more rapidly than a new trial.

The potential outcomes framework

In public policy the aim of experimental and observational studies is to measure the causal impact of an intervention on the lives of people. Measuring causal impact means assessing the extent to which interventions in the fields of education, health, criminal justice, employment, income support or economic productivity improve outcomes with minimal adverse side effects.

The potential outcomes framework provides a simple exposition of what is meant by causal impacts (Rubin, 1978). For a given population, the average causal impact of an intervention is the difference in average outcomes that occur with and without the intervention being assigned. In the standard notation:

Average treatment effect = $E[Y1i \mid Di=1] - E[Y0i \mid Di=1]$ eqn 1

The fundamental problem of causal inference is that there is no data on the outcomes for the treatment group in the absence of the intervention being received. The key issue is that $E[Y0i \mid Di=1]$ is never observed, and the essence of causal research is to use a robust strategy to estimate this potential outcome.

One strategy is to measure outcomes for a group who are exactly equivalent to the control group who do not receive the intervention.

A simple version of this occurs when individuals in a defined population are randomly assigned to either receive or not receive an intervention. Random assignment means the control and treatment groups are in expectation equivalent across both observed and unobserved characteristics. If there is no spill-overs between individuals in the treatment and control groups, the expected outcomes of the control group are a proxy for the treatment group without treatment. Specified in potential outcomes notation:

$E[Y0i \mid Di=1] = E[Y0i \mid Di=0]$ eqn 2

Outcomes for the control group can then be substituted into equation 1 to calculate the average treatment effect.

A more complex version of creating an equivalent control group can occur where assignment is partly random, and partly determined by other observable and unobservable factors. The critical issue is that there is an assignment mechanism that is probabilistic, but also uncorrelated with any unobserved characteristics that will affect outcomes. This last aspect is referred to as unconfoundedness or conditional

independence (Imbens and Rubins, 2015). In potential outcomes notation this means that after adjusting for any differences in observable characteristics, average outcomes of the control group are a proxy for the treatment group without treatment. In potential outcomes notation:

$E[Y0i \mid Di=1, Xi] = E[Y0i \mid Di=0, Xi]$ eqn 3

Intentional random assignment is the essence of a randomised control trial.

The assumption of accidental randomisation is the basis for many observational techniques including natural experiments, matching, and instrumental variables. These observational techniques assume that the treatment is either strictly random, or partially random but unconfounded after adjusting for observed covariates. The key issue for these observational techniques is that this assumption may not be plausible. In many studies there is often considerable uncertainty about how assignment occurred. There are often multiple decision makers making choices based on factors that are not well measured in data. If the assumption of uncounfounded assignment is not valid then the control group outcomes are not a good proxy for the untreated treatment group.

An alternative strategy for causal inference measure outcomes for non-equivalent groups (Shadish et al., 2002). Rather than creating a control group that is equivalent to the treatment group, this strategy finds a group that is a reliable benchmark for the untreated treatment group. In potential outcomes notation this means that:

$E[Y0i \mid Di=1] = E[Y0i \mid Di=0] + \text{difference}$

For this strategy to be successful the difference needs to be highly predictable and able to be estimated precisely.

The strategy involving non-equivalent groups relies on deterministic rather than probabilistic assignment. This means that allocation of the intervention is a function of observable characteristics.

Assignment also needs to be unconfounded, meaning that interventions cannot also be allocated to groups as a function of unobserved covariates that affect the outcomes.

Using a non-equivalent control group is the basis for the difference-in-difference strategy (Card and Kreuger, 1994). Assignment of the intervention is determined strictly by characteristics related to group membership, and this allocation cannot be confounded with any other unobserved variables that matter for outcomes. Moreover, to be a viable approach there needs to have been a stable difference between the two groups in the historical data. Inference relies on the assumption that the difference would have continued during the time that the treatment was applied. Although this is sometimes referred to as the 'common shocks' assumption, it is also the assumption that there is no unobserved confounding between the treatment and the difference.

A non-equivalent control group is the basis for the regression discontinuity design (Thistlethwaite and Campbell, 1957). In this instance treatment is determined by being above a threshold on a well-defined assignment score. Inference uses the outcomes for the control group after adjusting to account for the difference in outcomes from the extrapolated regression function across the threshold. While groups either side of the discontinuity will be very similar, they are not equivalent due to marginal differences in the covariates that give rise to differences in the assignment score.

Figure 6 represents a typology of empirical methods based on both the form of study and the mode of causal inference. Within each quadrant we highlight examples of specific approaches. Broadly the light shaded quadrants can be classified as 'quasi experimental' studies, which differ from the purposeful designed randomisation of the gold standard approach (Shadish et al., 2002).

Figure 6: Typology of approaches for studies that enable causal inference

	Equivalent control group	**Non-equivalent control group**
Experimental	Randomised control trial	Difference-in-difference control trial Discontinuity control trial
Observational	True natural experiments Matching Instrumental Variables Regression on matched samples	Difference-in-differences Regression discontinuity design Regression on unmatched samples

The focus of this paper is what we call the difference-in-difference control trial, which is in the quadrant of an experimental study but using a non-equivalent control group strategy.

It is also useful to note that within the quadrant there are other types of approach, one of which might be called a regression discontinuity control trial. This assigns treatment based on being above a defined score, and then analyses the difference in outcomes across the discontinuity (Cappelleri and Trochim, 1994).

Each of the types of studies highlighted above provide empirical tools that are suited to different situations.

A difference-in-difference control trial is most suited to a situation where the impact of a new intervention needs to be assessed, but the traditional randomised control trial is not appropriate or feasible. This may be due to ethical concerns related to the appropriateness of using randomisation to allocate the intervention. Randomisation may also not be feasible due to costs. A good example of this occurs with some public health interventions where interventions are delivered to communities. For example, implementing a randomised trial of a social marketing intervention across multiple communities can be very costly, and a difference-in-difference strategy can be a more cost-effective approach (Formoso et al., 2013).

The inference associated with a difference-in-difference trial is not as robust as a randomised trial because it requires more maintained assumptions. However, in many cases it is better than not having any evidence at all, particularly if there is a need to compare evidence across different types of interventions. Only relying on evidence about interventions that can be assessed using a randomised control trial introduces a particular type of selection bias into scientific knowledge.

Implementing a difference-in-difference control trial

The broad outline of a difference-in-difference control trial was described by Campbell and colleagues several decades ago (Campbell and Stanley, 1963; Cook and Campbell, 1979). In this older tradition it was often described as an interrupted time series design with a non-equivalent control group (Shadish et al., 2002). Actual use of the design has remained comparatively rare, with only a few examples from the fields of psychology, education and health (Handley, 2018; West, 2008).

We think that the approach could be used more widely. Over recent decades there has been considerable development of the toolkit for how to successfully conduct a randomised control trial (Glennerster and Takavarasha, 2013). Much of this toolkit can also be readily adapted for the difference-in-difference version.

In addition, there has also been considerable development of the difference-in-difference method for observational studies in recent decades (Athey and Imbens, 2006; Angrist and Pischke, 2009; Bertrand et al., 2004; Wing et al., 2018). Many of these developments are readily transferable to an experimental context and can strengthen the quality of inference about causal impacts.

Like a traditional randomised control trial, conducting a difference-in-difference control trial involves three stages of planning, implementation and analysis. Importantly, the trial involves more than just using difference-in-difference method at the analysis stage after a trial or pilot has been conducted. For the approach to be successful, the trial needs to be specifically planned and implemented in a manner that will enable valid inference using the difference-in-difference method. This is the same as with a randomised trial, where the validity of the empirical analysis hinges on the trial being well designed and properly implemented.

In what follows we briefly describe the components of the three phases of sequential activity necessary to conduct a difference-in-difference control trial. This discussion is based on our reading of established best practice for a randomised control trial, adapted for use with difference-in-difference analysis. In the second part of this section we describe how a modern approach to difference-in-difference analysis can be adapted to an experimental setting.

A broad overview of the process of conducting a difference-in-difference control trial

The planning stage is the first phase of activity required for a successful trial. There are several tasks that need to be undertaken at this stage, and the key elements of this stage should be documented and published in a trial protocol.

Defining the research question for the study: As with any research there is a need to be clear about the precise research question. In this case the question is about the impacts of an intervention on a defined population. Importantly, both the intervention and the population need to be precisely described, and it is necessary to be able to articulate a well-defined intervention logic founded on existing evidence.

Trial design: This requires consideration of the exact design for the trial including the choice of control group, how assignment will occur, the data collection strategy, and the various design features that can be used to improve both the power and robustness of the inference. Importantly, this stage should demonstrate that there is a 'constant difference' between the proposed control and treatment group. The nature of the

robustness tests to be used at the analysis stage should also be planned and specified. More detail on these and the design features that can improve the robustness of the inference are discussed in the following section.

Ethics, consent and privacy: There needs to be explicit and independent consideration of the ethics of undertaking the research, with specific emphasis on the risk of harm to participants, consent and also privacy.

Prototyping: An important and often overlooked element of undertaking a trial is to ensure that the planned intervention can be provided and is able to function as intended. This frequently requires prototyping the intervention and seeking feedback from both recipients and those tasked with delivering the intervention. If this is not done there is a risk that a trial will be undertaken with a poorly conceived and badly delivered intervention.

Project management: The key to successfully conducting a trial is to ensure that it is implemented according to plan, and this requires explicit consideration of governance, tasks, timelines and responsibilities.

The second stage of activity is the implementation of the trial. Implementation in a manner that adheres to the protocol is essentially a project management task. The aim is to ensure that there is controlled assignment of the intervention, that this is delivered according to what was intended, and there is reliable collection of data. As with a randomised control trial, the validity of the analytical results requires successful implementation of these tasks.

If the design and implementation phases have been well conducted, then the final analysis stage should be relatively simple. This then includes:

- clear description of the intervention
- documentation of the trial design and process by which treatment was assigned
- information on characteristics of the treatment and control group populations, and
- simple difference-in-difference analysis with tests that assess threats to internal validity.

Difference-in-difference estimation in an experimental setting

The essence of the inference for a difference-in-difference control trial is that the outcome for a control group, adjusted to account for a constant difference in relation to the treatment group, provides an estimate of unobserved potential outcome for the treatment group.

In an experimental study the standard difference-in-difference approach can usefully be separated into two parts.

In the first part, which should be conducted as part of the planning stage, it is necessary to identify a possible control group, and then demonstrate that there exists a constant and reasonably precisely estimated difference in outcomes with the treatment group.

Undertaking the analysis prior to the trial allows the comparison group to be selected to improve the chance that the trial can measure impacts successfully. Conceptually this involves identifying a control group who will experience the same unobserved shocks as the treatment group. For example, if the control group is geographically close to the treatment group it is expected that they will likely experience common unobserved

shocks.[5] Similarly, if it is possible to use a well-defined score to assign treatment, then individuals who score just below the threshold for assignment should be very similar to the treatment group.

After identifying a control group, it is necessary to demonstrate that there is a stable difference in the study outcome between this group and the treatment group in the historical data. A precondition for a trial to be undertaken is that there is a stable difference (or 'common trends'). If a stable difference can be shown, then the precision with which the difference is estimated describes the ability of the future experiment to detect impacts.

The simple two period model to analyse the historical difference prior to any treatment is:

$Outcome_{it} = \beta_0 + \beta_1(\text{time dummy}) + \beta_2(\text{treatment\_group\_ind}) + \beta_3(\text{time\_dummy} * \text{treatment\_group\_ind}) + X' \beta_4 + \varepsilon_{it}$

In this set-up a test for a stable difference is that $\beta_3=0$. In addition, the standard error of β_3 provides insights into the expected power of the experiment. Where these parameters are imprecisely estimated only large treatment effects will be able to be detected.

A consequence of estimating the difference prior to the trial being undertaken is that the covariates and functional form of the outcome equation are established. This provides an ability to identify which specification is most suited to the analysis, and also makes the final analysis more robust in the sense that it limits the ability of researcher to allow the final results to inform the choice of specification.

After the trial has been implemented the standard difference-in-difference estimation of treatment effects can be undertaken as part of the analysis stage.

Despite demonstration of common trends prior to the trial commencing, the validity of the difference-in-difference estimator still rests on a maintained assumption of "common shocks". The common shocks assumption is not possible to empirically assess, but there are some tests that can highlight areas of concern.

One of these is the difference-in-difference of covariates for the treatment and control groups. As part of the analysis it is useful to provide a table showing the difference-in-difference in covariates. Large shifts in the relative distribution of covariates are suggestive that unmeasured factors might be a factor in any observed difference in outcomes.

There are robustness tests that can also provide additional assurance about the validity of the difference-in-difference impact estimates of the study. Robustness tests can use other outcomes, or groups, to assess the results and use the standard difference-in-difference set-up. Examples include:

- analysing outcomes that are not thought to be affected by treatment, and showing that there is no relative change in these outcomes, and
- identifying other placebo groups (either controls or treatments) for whom it is possible to assess if there is any difference-in-difference impact on the study outcome measure.

[5]Although this also creates a threat to validity if there are spill-overs from the treatment to the control areas.

Annex 2: Description of the analytical dataset

The dataset for this study was drawn from administrative datasets maintained by the Ministry of Social Development.

Table 15 sets out key variables in the analytical dataset. The unit of measurement is the person and their associated partner and dependent children. The data is cross-sectional snapshots on 6 July 2018, 3 August 2018, 7 September 2018, and 2 November 2018. Outcomes are measured over the subsequent seven-week window from the snapshot dates. The data is 'as measured on the date', as opposed to the 'current view' of the data which can be different due to backdating changes.

Table 15: Characteristics of individuals in the randomised contact groups (November 2018)

Variable	Definition
SWN	Unique Social Welfare Number of recipient
Extract date	Date of snapshot
Sex	Male and Female
Age	Derived from date of birth
Ethnicity	Self-reported ethnicity which has been prioritised. Categories include Māori (first priority category), Pacific, European, Other and Unspecified
Family structure	Based on administrative information recorded in SWIFTT. Categorised according to partnership status and the presence of dependent children included in the benefit. Main categories are single without dependent children, couple without children, couple with dependent children, and sole parent with dependent children
Main benefit type	Type of main benefit received and categorised as Job Seeker Support, Supported Living Payment, Sole Parent Support, New Zealand Superannuation or Veterans Pension, and Non-beneficiary
Supplementary payment type	Details on nature of supplementary payments
Dollar value of payments	Weekly dollar amount of each payment received at extract date
TAS related calculation variables	Variables related to the TAS calculation including housing costs, disability costs and value of assets

TAS deficiency	Modelled TAS deficiency
Receipt of TAS in previous 24 months	Indicator variable relating to prior TAS receipt
Eligible for campaign	Variables defining eligibility for campaign based on indicator variables related to having a valid income record, not applied, cancelled, suspended or expired TAS in last 120 days, partner not applied for TAS in last 120 days, not previously contacted in previous 120 days about TAS, partner not suspended or expired benefit, not sanctioned, not receiving Residential Care Subsidy or in Hospital, not receiving Student Allowance or Veterans Pension.
Eligible for randomisation	Variables defining eligibility for randomisation including not being hearing impaired, blind, not having restricted records, valid email, valid phone, and address contact details
Weekly payments at end of 7 weeks	Value of all payments 7 weeks after extract date
TAS grant after 7 weeks	Granted TAS after 7 weeks
Value of TAS grant	Weekly amount of TAS grant within 7 weeks of extract date

Annex 3: Balance across randomised treatment groups

Table 16 shows the characteristics of individuals in the three randomised treatment groups. For example, women made up 60 percent of those who were randomly allocated to receive letters, 62 percent of the email group, and 59 percent of the phone group.

Table 16: Characteristics of individuals in the randomised contact groups (November 2018)

	Letter (n=521)	Email (n=1191)	Phone (n=578)	p-value difference
Women	0.60	0.62	0.59	0.37
Men	0.40	0.38	0.41	0.37
Māori	0.26	0.26	0.25	0.82
European	0.42	0.41	0.43	0.63
Pacific	0.06	0.06	0.06	0.83
Other ethnicity	0.24	0.25	0.22	0.56
Unspecified ethnicity	0.02	0.03	0.03	0.62
Under 25 years of age	0.05	0.06	0.06	0.75
25 to 44 years of age	0.36	0.38	0.35	0.42
45 to 64 years of age	0.43	0.41	0.43	0.69
65 years of age and over	0.16	0.16	0.16	0.91
Single person	0.54	0.53	0.52	0.75
Couple no children	0.18	0.18	0.21	0.45
Couple with children	0.07	0.07	0.05	0.23
Sole parent with children	0.20	0.21	0.22	0.71
Non beneficiary	0.10	0.12	0.09	0.13
Job Seeker Support	0.43	0.38	0.37	0.05*
Supported Living Payment	0.15	0.17	0.19	0.22
Sole Parent Support	0.16	0.16	0.18	0.48
NZS or Veterans Pension	0.14	0.16	0.15	0.58
Modelled TAS deficiency	$127.35	$129.27	$132.20	0.33
Estimated TAS entitlement				
Value of assets	$29.47	$41.10	$35.59	0.47
Proportion TAS receipt in previous 24 months	0.44	0.42	0.44	0.69

Source: MSD data. Population 'C'.

As would be expected given randomisation there is little evidence of statistically significant differences. The one exception is the slightly higher proportion of Job Seeker

Support among the letter group compared to the others, which we interpret as imbalance occurring by chance.

Annex 4: Glossary

Temporary Additional Support (TAS) - A temporary 13-week payment for individuals with high levels of essential expenses compared to income.

Accommodation Supplement (AS) – Supplementary payment for people with moderate to high housing costs

Ministry of Social Development (MSD) – Administers welfare payments and active employment assistance.

Microsimulation Model (MSIM) – The population for this model is all people receiving payments from the Ministry at a point in time. The model has data on actual payments received and can be used to simulate eligibility and rate of payments.

Welfare payments – these include any payments made by the Ministry of Social Development. It is useful to distinguish between income tested main benefits (eg Job Seeker Support), non-income tested main benefits (eg NZS), tax credits (eg Family Tax Credit) and supplementary payments. Individuals do not need to be receiving a main benefit to be receiving supplementary payments.

SWIFTT – the main MSD computer system that is used in the delivery of welfare payments.

References

Angrist, J. D., and Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press. https://doi.org/10.1111/j.1475-4932.2011.00742.x

Athey, S., and Imbens, G. (2018). Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption. Retrieved from http://arxiv.org/abs/1808.05293

Bargain, O., Immervoll, H., and Viitamäki, H. (2012). No claim, no pain. Measuring the non-take-up of social assistance using register data. *Journal of Economic Inequality*, 10(3), 375–395. https://doi.org/10.1007/s10888-010-9158-8

Campbell, D., and Stanley, J. (1963). Experimental and quasi-experimental designs for research on teaching. In N. Gage (Ed.), *Handbook of Research on Teaching*. Chicago: Rand-McNally

Cappelleri, J., and Trochim, W. (1994). An illustrative statistical analysis of cutoff-based randomized clinical trials. *Journal of Clinical Epidemiology*, 47, 261–270

Card, D., and Kreuger, A. (1994). Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania. *American Economic Review*, 84, 772–784

Cochran, W. (2015). Observational Studies. *Observational Studies*, 1, 126–136. Retrieved from http://obsstudies.org/journal.php?id=24

Cook, T., and Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally

Currie, J. (2006). The take-up of social benefits. In A.J. Auerbach et al. [eds.] (2006). *Public Policy and the Distribution of Income* (pp. 80–148). Russell Sage Foundation.

Errington, L. and Human, J. (2019). Temporary Additional Support: Client Feedback to MSD's Proactive Contact, unpublished report

Formoso, G., Paltrinieri, B., Marata, A. M., Gagliotti, C., Pan, A., Moro, M. L., and Magrini, N. (2013). Feasibility and effectiveness of a low-cost campaign on antibiotic prescribing in Italy: Community level, controlled, non-randomised trial. *BMJ (Online)*, 347(7926), 1–10. https://doi.org/10.1136/bmj.f5391

Glennerster, R., and Takavarasha, K. (2013). Running Randomized Evaluations: A Practical Guide. Princeton University Press. https://doi.org/10.2307/j.ctt4cgd52

Handley, M. A., Lyles, C., McCulloch, C., & Cattamanchi, A. (2018). Selecting and Improving Quasi-Experimental Designs in Effectiveness and Implementation Research. SSRN. https://doi.org/10.1146/annurev-publhealth-040617-014128

Hernanz, V., Malherbet, F., & Pellizzari, M. (2004). Take-Up of Welfare Benefits in OECD Countries: A Review of the Evidence. OECD Social, Employment and Migration Working Papers. https://doi.org/10.1787/525815265414

Imbens, G. W., and Rubin, D. B. (2015). Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge, Cambridge University Press https://doi.org/10.1017/CBO9781139025751

Kahn-Lang, A., and Lang, K. (2018). The Promise and Pitfalls of Differences-in-Differences: Reflections on '16 and Pregnant' and Other Applications. https://doi.org/10.3386/w24857

Lechner, M. (2010). The Estimation of Causal Effects by Difference-in-Difference Methods. Foundations and Trends in Econometrics, 4(3), 165–224. https://doi.org/10.1561/0800000014

Meyer, B. D. (1995). Natural and Quasi-Experiments in Economics. *Journal of Business and Economic Statistics*, 13(2), 151–161. Retrieved from http://www.tandfonline.com/doi/full/10.1080/07350015.1995.10524589

Ross, H., Campbell, D. and Glass, G. (1970) 'Determining the social effects of a legal reform: The British breathalyser crackdown of 1967'. *American Behavioural Scientist*, 13: 493-509

Rubin D. B. (1994). Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701. https://doi.org/10.1037/h0037350

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics*, 2(3), 808–840. https://doi.org/10.1214/08-AOAS187

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston, Houghton Mifflkin Company

Thistlewaite, D. L., and Campbell, D. T. (1960). Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment. *Journal of Educational Psychology*, 51, 309–317. Retrieved from https://obsstudies.org/wp-content/uploads/2017/01/regression_discontinuity_all_comments-1.pdf

West, S. G., Duan, N., Pequegnat, W., and others (2008). Alternatives to the randomized controlled trial. *American Journal of Public Health*, 98(8), 1359–1366. https://doi.org/10.2105/AJPH.2007.124446

Wing, C., Simon, K., & Bello-Gomez, R. A. (2018). Designing Difference in Difference Studies: *Best Practices for Public Health Policy Research*, 39, 453–469. Retrieved from https://www.annualreviews.org/doi/10.1146/annurev-publhealth-040617-013507