



**MINISTRY OF SOCIAL
DEVELOPMENT**

TE MANATŪ WHAKAHIATO ORA

Cost-effectiveness of MSD employment assistance

Technical report for 2016/2017 financial year

November 2018

Authors

Marc de Boer, Principal Analyst, Insights MSD, Insights and Investment

Bryan Ku, Analyst, Insights MSD, Insights and Investment

Acknowledgements

We would like to thank the following people for their contributions and comments in preparing this report: Steven Su, Barbara Rackham, Ann Dostine, Laura Crespo, John Gibbs, Phillip Hall, Rob Hodgson, Eleonora Sparagna, David Rae Tim Maloney and Sarah Stacy-Baynes.

Disclaimer

The views and interpretations in this report are those of the Research and Evaluation team and are not the official position of the Ministry of Social Development.

Statistics New Zealand IDI disclaimer

The results in this report are not official statistics; they have been created for research purposes from the Integrated Data Infrastructure (IDI) managed by Statistics New Zealand.

The opinions, findings, recommendations and conclusions expressed in this paper are those of the authors, not Statistics New Zealand.

Access to the anonymised data used in this study was provided by Statistics New Zealand in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business or organisation and the results in this paper have been confidentialised to protect these groups from identification.

Careful consideration has been given to the privacy, security and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy Impact Assessment for the Integrated Data Infrastructure available from www.stats.govt.nz.

Inland Revenue IDI disclaimer

The results in this report are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must only be used for statistical purposes, and no individual information may be published or disclosed in any other form or provided to Inland Revenue for administrative or regulatory purposes.

Any person who has had access to the unit record data has certified that they have been shown, have read, and have understood section 81 of the Tax Administration Act 1994, which relates to secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes and is not related to the data's ability to support Inland Revenue's core operational requirements.

New Zealand Defence Force IDI disclaimer

The New Zealand Defence Force has consented to the release of IDI results for the Limited Services Volunteer programme to Statistics New Zealand as part of this report.

Published

Ministry of Social Development
PO Box 1556
Wellington
www.msd.govt.nz

February 2019

ISBN

REP/18/11/1533

Contents

Statistics New Zealand IDI disclaimer	2
Inland Revenue IDI disclaimer	2
New Zealand Defence Force IDI disclaimer	2
Overview	7
What are employment assistance interventions?	7
Definition of an EA intervention spell	9
Do you exclude of short duration spells or non-completers?	9
Common issues with EA intervention data	10
Case management services	11
Estimating the cost of EA interventions	12
Principles behind the cost allocation model	13
Cost allocation framework	13
Financial inputs	15
How do we estimate staff time?	15
Outcome measures	16
Income	17
Net income from all sources	18
Income support received after tax	18
Employment	19
Any time in employment	19
Time in employment while on main benefit	19
Time in employment and independent of Work and Income	19
Independent from Welfare	19
Justice	19
Any offence	20
Time spent in any corrections spell	20
Time spent in prison	20
Education qualifications	20
Qualifications achieved at NQF level 2, 3, 4	20
Highest NQF level	20
Education participation	20
Time spent in any education participation	21
Time spent in education while off benefit	21
Education participation NQF4	21

Driver licence status	21
Time spent while holding a learner's, restricted and full licence	21
Tracking outcomes longitudinally	21
Why measure outcomes from participation start?	24
Migration and mortality	24
Estimating the observed impact of EA interventions	25
Methods used to estimate the impact of interventions.....	25
Randomised Control Trial designs.....	25
Propensity matching	26
Quality of the matched groups	27
Propensity score matched historical comparison group.....	28
Natural experiments	28
Other considerations.....	28
Making multiple statistical inferences	28
Interpretation of EA impacts in the context of multiple interventions.....	29
Estimating future impact from observed impact.....	30
Step 1: Estimate the expected interval impact.....	31
Step 2: Calculate the projected cumulative impact	31
Interval impacts that do not trend towards zero	32
Step 3: Project cumulative impact from observed cumulative impact.....	32
Scaling interval impacts	33
The confidence interval for projected impact.....	33
Rating the effectiveness of interventions	33
Rating by outcome domain	33
Translating impact to an effectiveness rating	34
Rating the overall effectiveness of an intervention.....	34
Cost-benefit analysis.....	35
What does Welfare Return on Investment consist of?.....	35
Welfare ROI for case management services	37
References	38

Table of figures

Figure 1: Tracking EA intervention outcomes using administrative data.....	22
Figure 2: Interval outcomes for an example intervention.....	23
Figure 3: Cumulative outcomes for an example intervention.....	24
Figure 4: A stylised example of the relationship between interval and cumulative impact on time off benefit.....	30
Figure 5: A stylised example of projecting the interval impact.....	31
Figure 6: Forced taper in the projected impact of an intervention.....	32

Table of tables

Table 1: Classification of employment assistance interventions.....	7
Table 2: Case management services.....	11
Table 3: Cost components and their metrics.....	13
Table 4: A stylised example of a staff member's transactions in the morning.....	15
Table 5: Outcome domains that employment interventions can be expected to have an impact on.....	16
Table 6: Rating of outcome domain by the impact on outcomes.....	34

Overview

This technical report explains the approach taken by MSD to evaluate the cost-effectiveness of its employment assistance (EA) and case management (CM) service expenditure.

In this report, we describe:

- what we define as EA and CM services
- the outcome measures used to determine the effectiveness of EA
- methods used to estimate the impact of interventions
- a method for estimating unobserved future impacts
- the process used to rate the effectiveness of interventions
- how we determined the cost-effectiveness of programmes.

What are employment assistance interventions?

Employment assistance covers employment and training programmes, and services designed to help people prepare, move, sustain employment, and reduce the time they spend on income support. EA funded by MSD is primarily targeted at people eligible for income support assistance. Table 1 below provides a broad typology of EA interventions.

Table 1: Classification of employment assistance interventions

Type	Description	Interventions (examples)
Activation measures	Activation measures cover programmes and case management techniques designed to maintain job search activities for people expected to move into employment (eg people receiving unemployment related benefits). If people are judged not to be sufficiently engaged in job search then they can have their income support payments reduced or even cancelled.	Jobseeker Work Ready 52-week benefit reapplication Work Obligations Pre-Employment Drug Testing
Work confidence	Programmes designed at encouraging and motivating people to have the confidence to begin to move into employment.	Limited Services Volunteer Outward Bound
Case management	One-to-one meetings with a case manager to discuss and plan on how they will move back into employment. Case management can be either in-house or contracted to an external provider.	Work Focused Case Management (General) Work to Wellness
Career advice /Information Services	Career advice and counselling is a standard service provided by public employment services to help jobseekers make informed decisions about their current and future employment choices.	Careers Guidance and Counselling
Health Interventions	Providing medical and related treatment to enable people to recover from medical	PATHS Targeted Health

Type	Description	Interventions (examples)
	conditions to enable them to move back into work.	Interventions
Training (contracted)	Contracted training programmes aim to increase the foundational and vocational skills of participants to enable them to compete in the labour market.	Training for Work Driver licence programmes
Training (financial)	Financial assistance to help people access education and training programmes.	Training Incentive Allowance
Work experience	Provide people with work experience either with a private sector employer or through placements with not -for-profit organisations to help in social or environmental projects.	Mainstream Employment Programme Activity in the Community Flexi-wage Project in the Community
Job search assistance	Seminars and job clubs designed to provide jobseekers with the skills to look for work (eg searching for job leads, CV and applications and interview skills) and to provide peer support to maintain motivation.	Work Search Support
Job placement services	In-house or contracted-out services to place people into paid employment. For contracted-out services, payments are often based on a fee-for-outcome contracting model.	Employment Placement or Assistance Initiative Vacancy Placement
Hiring wage subsidies	A temporary subsidy to compensate employers who take on disadvantaged jobseekers (ie they would not have been hired by the employer in the absence of the subsidy).	Flexi-wage (Basic/Plus)
On the job training	Assistance to employees to help them gain skills whilst in work.	
Training for predetermined employment	Programmes that involve matching jobseekers to vacancies by providing short-term training to meet the specific needs of an employer.	Skills for Industry
Self-employment assistance	Assistance to help people set up their own business. Self-employment assistance can involve a combination of training, mentoring, capital grants, and a temporary subsidy to cover living costs until business cash flow is sufficient to support the participant.	Be Your Own Boss Business Training And Advice Grant Flexi-wage Self Employment (subsidy)
Incentive payments	Programmes that provide payments to people if they take up employment, in particular, to take up employment outside their local area.	3K to Work
Transition to work financial support	Financial assistance to help cover initial costs of moving into employment (eg work clothes and equipment) or to cover the period until the person is paid by the employer.	Transition to Work Payment

Type	Description	Interventions (examples)
In-work support (Financial)	Financial assistance to help people with disruptions to employment or pay to ensure they can continue in employment and avoid returning to main benefit.	Seasonal Work Assistance In-work tax credit
In-work support (Pastoral)	Programmes that contact people once they are in work to see how things are progressing and to help with any issues that might arise.	In-Work Support
Childcare assistance	Financial payments to low-income families to help cover the cost of childcare services.	Flexible Childcare Assistance Childcare Subsidy OSCAR Subsidy
Incentive payments	Payments to people who remain in employment for set periods (ie 3, 6 and 12 months).	In Work Payment Work Bonus
Vocational Services	Externally-contracted employment assistance to people with ill-health or disabilities.	Vocational Services Employment
Youth Programmes	Assistance targeted at teens (usually under the age of 18) to help them remain in education, training or employment.	Youth Services
Migrant assistance	Assistance targeted at new migrants and refugees.	Migrant Employment Assistance
New Initiatives	Locally developed initiatives that cannot be easily categorised.	New Initiatives

Definition of an EA intervention spell

For the purposes of our analysis we define EA interventions as discrete events that have the following attributes:

- Person identifier: system id that determines who received the intervention
- Intervention name: name of the intervention
- Start date: calendar date the person started the intervention.

These elements are the minimum necessary to evaluate the effectiveness of the intervention. Additional attributes that we also try to collect include:

- End date: the date the person finished the intervention
- Referral date: when they were referred to the intervention
- Provider: who delivered the intervention, especially when it is contracted out
- Cost: how much it cost, this is covered in the section on the cost of employment interventions (see page 12).

Do you exclude of short duration spells or non-completers?

In the analysis of EA interventions, we include all participant starts and do not exclude any apparently short participation spells or those recorded as not having completed the programme or course. There are two reasons for taking this position. The first is the difficulty in having reliable participation end dates or information on who successfully completed the intervention. As discussed below, this information may or may not be

recorded depending on the source system or the diligence of staff in recording these types of outcomes in the administrative system.

The second reason is that we consider early exits or non-completion is a core feature of the intervention and should be included in its assessment of its effectiveness. Therefore, if an intervention is being poorly run such that many participants either exit soon after starting or fail to complete the programme, then this should be reflected in its performance.

Common issues with EA intervention data

Because EA information is located in more than seven MSD administrative systems compiling information about EA interventions is not always straightforward. As with most administrative data, there are a number of issues with how well EA intervention data is recorded.

Duplicate participation events

Participation events are defined as any recorded participation spell. In some cases, a person may participate in two different interventions on the same day. This occurs where a person may receive different forms of assistance (eg a Job Plus and Work Start Grant or Enterprise Allowance and Enterprise Capitalisation). However, there are duplicate participation spells for the same person, intervention and start date. When these occur we select only one event.

In a small number of instances, we also take an intervention's participation from only one IT application source. These occur where business processes indicate that this is the primary system for recording intervention participation spells.

Inconsistent system information

In a substantial number of cases, there is more than one source of information for a variable for a given EA participation record. Important examples include the name of the intervention, provider name and participation start and end dates. We identify those records where these inconsistencies occur. The general approach to resolving these inconsistencies has been to favour the source that is most closely associated with the event itself. For example, if a contract system end date differs from the front line application end date, we take the front line system end date.

Participation end dates

One difficult area of EA participation is an accurate recording of participation end dates. Either end dates are missing or they are miss-keyed, giving either implausibly long participation spells or end dates that are earlier than start dates. In many instances, it is necessary to impute end dates where they are currently null or implausible (eg a seminar lasting eight years). If there is information about the expected end date, then we use this when the actual end date is missing. If no suitable end date is available, we estimate the end date based on the observed duration that people spend on the intervention or a similar intervention, if required.

Referrals

From an evaluation perspective, we are interested in when people are referred to interventions to identify who might have been approached about participating as well as to identify likely effects of being referred to an intervention. For example, do we see a

lock-in effect or the reverse, people exiting benefit in the period between referral and intervention start?

Referral information is generally unreliable for this type of analysis. In many instances, formal recording of referral occurs after an informal discussion and conversation with the intervention providers. Under these conditions, referral captures the point when a person is confirmed as intending to participate in the intervention. What is missed are those people where the case manager may have discussed the opportunity with the participant and the individual turned it down or where a provider had screened potential participants out.

Case management services

From October 2012 in selected sites, and from July 2013 for all sites, people receiving income support (main benefit or supplementary only) are automatically allocated to a case management service. The allocation process is managed centrally, with the site and case managers sent lists of people in each case management service. In addition, the service allocation process updates front-line systems of the individual's service status.

The establishment of case management allocation has enabled MSD to better understand who is receiving different levels of case management assistance. Table 2 summarises the case management services included in the case management allocation process.

Table 2: Case management services

Name	Description	Start Date
General Case Management	People NOT assigned to any other service are allocated to GCM. For GCM there is no caseload restriction.	September 2012
Mental Health Employment Service	A voluntary contracted-out job placement service targeted to people on Jobseeker Support with mild to moderate mental health issues (stress, depression or anxiety).	September 2013 to July 2016
Sole Parent Employment Service	A voluntary contracted-out job placement service targeted to sole parents with full or part-time work obligations.	July 2013 to July 2016
Work Focused Case Management (pilot)	One-to-one case management, with caseloads, capped at 108 participants for every case manager. WFCM case managers provided both income support and employment assistance.	September 2012 to July 2013
Work Focused Case Management General	One-to-one case management with a caseload cap of 121 participants (excluding partners) for each case manager. WFCM Gen clients are mainly made up of sole parents, jobseekers and a small proportion of jobseekers with a health condition or disability (with part-time work obligations).	July 2013
Work Focused Case Management Health Condition and Disability	One-to-one case management with a caseload of 100 participants who were exclusively jobseekers with a health condition or disability and had deferred	July 2013

Name	Description	Start Date
	work obligations (predominantly work preparation obligations).	
Work Focused Case Management Integrated Service	One-to-one case management that began with a caseload of 50 but was later increased to 80, working with people under 25 years of age at selection, with priority given to people exiting from YP and YPP. In addition, these case managers could choose to work with up to 20 other people, nominated by either themselves or other case managers.	July 2013
Work Focused Case Management Integrated Service ex-Youth Payment Young Parent Payment	People who came onto a main benefit from Youth Payment or Young Parent Payment were prioritised for the WFCM IS service, but may have received the WFCM Gen service if WFCM IS was unavailable.	July 2013
Work Focused Case Management Intensive Client Support (ICS)	An internal case management service with a caseload of no more than 40 people for each case manager. WFCM ICS case managers work with high needs clients who have complex issues or barriers to employment.	March 2015
Work Search Support (pilot)	A one-to-many case management service, with caseloads capped at 215 participants for every case manager. Alongside administering income support, WSS involved a structured sequence of job search seminars. In addition to job search assistance, WSS clients also participated in employment programmes and services. People on main benefit for less than seven weeks were left to manage their own job search.	September 2012 to July 2013

Estimating the cost of EA interventions

We use the individual Cost Allocation Model (iCAM) to estimate the cost of EA interventions for each financial year (MSD, 2017). Insights MSD created iCAM to provide a view of how spending to date has been allocated to outputs at the individual level. Here we define outputs as activities that MSD does to assist people. Outputs can include activities, such as a face-to-face meeting, a main benefit application, or an EA intervention.

Principles behind the cost allocation model

The cost allocation model works on the following principles:

1. Include all financial costs for Service Delivery (the operational arm of MSD): the model starts with appropriation¹ expenditure for all outputs delivered by Service Delivery. Excluded at this time are income support payments designed to reduce income inadequacy. The reason behind this principle is to make sure we do not exclude any costs that are already recorded in the Ministry's financial systems.
2. Reconcile allocated expenditure to financial totals: for each appropriation, the model reconciles (as far as it is possible) the allocated expenditure back to the appropriation amount in each financial year. At the very least, the sum of the allocated expenditure in each financial year should not exceed the appropriation amount.
3. Disaggregate costs down to the individual output level: to provide the highest level of accuracy and flexibility, the model disaggregates costs down to outputs (see the Cost allocation framework section below) at the person-event level. By doing so, we can accurately assess the amount of expenditure for individuals as well as have the flexibility to summarise costs for any group of people. By building the model this way, we can estimate the variability in output costs.
4. Apply the same approach over all financial years: by applying the same approach across financial years (from 2001/2002 onwards) it is possible to identify trends in the cost of Service Delivery outputs across groups of people. However, this also means it is not possible to compare results across different versions of reports or cost models.

Cost allocation framework

In this report, we will briefly describe how the cost model works by using an example of an in-house seminar delivered by MSD. For a more detailed description, please refer to the iCAM technical report (MSD, 2017).

The process of delivering a seminar can be broken down into several components as listed in Table 3. We can think of components as the activities that are involved in delivering the EA. For example, the time staff spend the booking an appointment for a participant to attend the seminar. Also, if the seminar is run internally, staff have to spend time administering and running the seminar. We first determine the total expenditure (see the Financial inputs section below) for each of these components for that seminar in a financial year. See the table below for a list of components in the iCAM.

Table 3: Cost components and their metrics

Component	Definition	Metric
Appointment	Scheduling an appointment	Staff time
Benefit administration	Assessing and maintaining entitlement to income support assistance	Staff time
Benefit payments	Bank fees for payment of income support benefits	Pay weeks

¹ We use the term here to refer to how public money is spent, see: <https://treasury.govt.nz/publications/guide/guide-appropriations-html#section-1>

Component	Definition	Metric
Client contact	Contact with individuals to help them plan and move into employment or updating their records	Staff time
Contract Administration	Administration of contracts, including tendering, negotiation, payment and managing the performance of contracted providers	Contract amount
Contract payment	Payment of contracts	Contract amount
Grant	Financial transfer to people to assist them with undertaking further training or with transitioning into employment	Grant amount
Grant Administration	Assessing and administering grant applications	Staff time
Integrity (fraud and debt)	Identification of benefit fraud and the collection of outstanding debt	Staff time
Placement Opportunity	Time spent by contact centre staff and work brokers to identify and establish vacancies with employers	Starts
Referral	Time spent by case managers in referring people to employment vacancies, employment programmes, or training programmes	Staff time
Seminar	Staff time in administering and running seminars	Staff time
Study Assistance	Time in assessing and maintaining entitlement to student loans and allowances	Staff time
Wage Subsidy	Payments made to employers or sponsors for hiring wage subsidy, work experience, or self-employment programmes	Subsidy payments
Wage Subsidy Administration	Cost of administering wage subsidy assistance	Starts
Unallocated Service Delivery	Unallocated frontline staff time costs for Service Delivery	Duration on income support or student allowance
Overhead costs	IT, corporate services, property, and support staff costs	Departmental cost of each output

The next step is to find a metric related to each component so that we can assign a dollar value to that component. We define metrics as quantitative information about an output's component. For example, for the appointment component, we can use the number of minutes that staff spend to book participants for each seminar. Multiplying the number of minutes spent by staff by a per-minute staff rate will give us the appointment cost for each seminar attendee.

Finally, we add the cost of each component to arrive at a total cost for the seminar. The variation in the cost of each seminar output for the financial year will depend on the variability in the cost of its components.

Financial inputs

Having identified the outputs, their cost components, and how to assign costs to them, the next question is where we source the financial costs for Service Delivery. We can access records of Service Delivery expenditure through the Ministry's financial accounting system called KEA. These records capture expenditure information down to the cost centre and nominal level. Some cost components have a relatively straightforward link to the financial inputs – for example, the wage subsidy payments for a wage subsidy programme.

How do we estimate staff time?

Table 3 shows that staff time is a commonly used metric in the model. However, obtaining the data is not straightforward. In this section, we summarise how we collect staff time data. The source of this information is staff system transactions on MSD's various IT admin systems² to estimate what frontline staff are doing through the workday, and with which individual. The key information for these transactions are:

- A unique ID for a staff member
- A unique ID for an individual
- A start time
- An end time
- What the action was.

This allows us to construct a transaction-based view of the staff member's day. Table 4 below shows a stylised example of three consecutive transactions performed by a staff member.

Table 4: A stylised example of a staff member's transactions in the morning

Staff ID	Client ID	Start time	End time	Activity
xxx	my	03/04/2012 8.30 am	03/04/2012 8.34 am	Look up client details
xxx	yyy	03/04/2012 8.35 am	03/04/2012 8.44 am	Book appointment for the seminar
xxx	zzz	03/04/2012 8.45 am	03/04/2012 8.57 am	Update client details

We then link transactions to outputs that have components with staff time as a metric. These transactions should occur around the start date of the output, or within the start date and end date of the output, depending on the type of component. For example, the second transaction related to booking a person for a seminar in Table 4 could be linked to a seminar output that happened a week later for a person with the client ID yyy.

² For a list of IT admin systems that we currently track, please refer to the iCAM technical report.

Outcome measures

When reporting on effectiveness, we measured the impacts of EA interventions across a range of outcome domains. We focus on those domains that we expect employment assistance to have a direct impact on as shown in Table 5 below. However, we acknowledge that we do not have all outcomes that interventions could reasonably be expected to impact (eg outcomes, such as children’s short- and long-term outcomes, health status and household income). The absence of an outcome measure is often because the SNZ IDI either lacks this information or there are issues with the data that need to be resolved. Over successive reports, we have expanded the number of outcomes included in the analysis. In this year’s report, we have included education qualifications and participation and justice outcomes to the analysis. In subsequent reports, we plan to further expand the range of outcomes included in the analysis, such as mortality and periods spent overseas.

Table 5: Outcome domains that employment interventions can be expected to have an impact on

Outcome domain	Included	Comment
Employment	Yes Inferred from tax data	By definition, all employment interventions have either a long- to medium-term goal of increasing time in employment.
Income	Yes Labour market income and transfers to individuals	While employment is important, interventions should not result in a reduction in overall income. Currently, we have not developed a measure of household income because of the difficulty of defining households within the IDI.
Education and training	Yes Government funded training and education	Many interventions have the goal of helping participants take up further training or education.
Qualifications gained	Yes Only includes formal qualifications	In general, returns from education and training depend on the achievement of qualifications, especially higher level qualifications. The gain in qualifications assumes an increase in human capital.
Confidence and motivation	No Difficult to measure from current data sources	A common objective of case management and related programmes is to increase participant’s confidence and motivation. The assumption is that increases in confidence and motivation will move

Outcome domain	Included	Comment
		people towards employment.
Health care use	No Data is available and we plan to include this outcome in later reports	A number of employment interventions involve the purchase of health care or help participants to access health care. The assumption is that resolving health needs will enable participants to move back into employment.
Health status	No Limited data that can be used for impact evaluation	For health-focused interventions, the impact of health status would be an important immediate outcome. For other EA interventions, health status is likely to be more of a medium to long-term outcome. We may be able to look at acute health care as an indicator of poor health.
Justice-offending	Yes	An expected indirect impact of employment is a reduction in criminal behaviour.
Children's immediate outcomes	No	EA interventions may have impacts on the immediate outcomes of participants' children. For example, improvement in income may result in better health.
Children's long-term outcomes	No	A long-term impact of EA interventions targeted at sole parents may be seen in the adult outcomes of their children. We are beginning to reach follow up periods (eg 18 years) where this analysis may be feasible.
Mortality	No Include in next update	While most interventions are not intended to impact on mortality directly, this could be a long-term impact.
Time overseas	No Include in next update	One consequence of increased employment may be a reduction in the probability of moving overseas (eg to Australia to find employment).

Income

Total income is an important measure of family's overall wellbeing. In the current analysis, we are restricted in looking at the income of individuals only. Ideally, we would like to measure the household income to better account for the overall material

wellbeing of individuals (eg supporting children or non-working household members). However, we have not yet developed a suitable measure of household income that can be used for the evaluation of EA interventions.

Net income from all sources

Net income from all sources is the main outcome measure. It includes all sources of income but excludes the drawdown of student loans. Income is net of tax. The measure was based on Inland Revenue (IR) and MSD data provided to the Statistics New Zealand Integrated Data Infrastructure (SNZ IDI).

The current income measures include:

- **Employer Month Schedule (EMS):** New Zealand operates a Pay As You Earn tax system. Accordingly, all employers provide IR with monthly schedules of the earnings of all their employees. In addition to employee earnings, the EMS also includes taxable income support (main benefit), Accident Compensation Corporation (ACC) and pension payments.
- **Self-employment and company earnings:** people who run their own business or company are also required to file annual tax returns. In the analysis, these annual returns are converted into monthly spells with annual totals split equally across the 12 months of the tax year. There can be considerable lags in the lodging of self-employment earnings that can mean measures of income for the most recent periods underestimate actual income. Note, however, because we update the analysis on a regular basis the results incorporate these lags in reported earnings in subsequent updates.
- **Non-taxable income support payments:** not all income support payments are subject to tax. In particular, second-tier assistance, such as the Accommodation Supplement and a third tier or hardship assistance such as Emergency Food Grants are not taxed. For hardship payments, we exclude recoverable assistance, as these are advances on main benefits. Recoverable payments will either be reflected in lower main benefit payments or, if the person moves off main benefit, in the form of an income support debt. At present, we do not have reliable data on income support debt.

Income sources not covered by the current measure:

- **Tax credits:** in the current analysis we have not included tax credits. IR has recently supplied tax-credit datasets to the IDI and we are in the process of developing business rules to extract this information.
- **Child Support:** transfer payments between custodial and non-custodial parents that are administered by IR.
- **Non-taxable income support payments to people over 65:** current income support data supplied by MSD to the IDI exclude payments to people on New Zealand Superannuation or Veteran's benefit.
- **Illegal and undeclared income:** the IDI data doesn't cover income from informal or illegal activities, including tax evasion.

Income support received after tax

Income support payments are both taxable and non-taxable. For consistency, we calculate the total amount of income support a person receives after tax. Because of

data limitations, income support only includes second and third tier income support payments for working age people. By using IR data we can include income support payments for people receiving New Zealand Superannuation payments. As noted above, we have not yet included tax credits into this measure.

Employment

Any time in employment

Employment is based on the period that people declare income from employment or from self-employment. Note that employment spells are based on either monthly or annual periods so we may be over or understating the actual time a person is in employment depending on where in the month or tax year they started or ended employment. At present, we have not attempted to adjust for this (eg by looking at the following or subsequent month to identify the likely start and end periods).

There are also lags in lodging tax returns, with these most pronounced for annual returns. We choose not to censor our analysis period to accommodate these lags and instead rely on regular updates to the analysis to incorporate delayed tax data into the results.

Time in employment while on main benefit

Here we include periods where a person is both on main benefit and receiving employment income. Sole parent benefits, in particular, are designed to allow people to remain entitled while earning relatively high levels of income from employment. Similarly, people on health and disability-related benefits may only be able to work part-time.

Time in employment and independent of Work and Income

Here we are interested in employment without support from main benefits or employment assistance. This measure is particularly useful when looking at subsidy-based interventions that mean participants are in employment but are supported indirectly through a subsidy.

Independent from Welfare

We measured the time people are dependent on welfare assistance by the period they were entitled to a main benefit and whether they were participating in EA interventions. The inclusion of the latter is to cover instances where people are receiving employment assistance while off main benefit (eg a wage subsidy).

A limitation of this measure is that it fails to account for negative destinations. For example, people who move from main benefit into prison would appear to be off welfare assistance.

Justice

We have two sources of information on justice outcomes: police offending and periods under Corrections supervision. These data are also obtained from SNZ IDI.

Any offence

This measure is based on Police data of people who are arrested for an offence (but may not result in a prosecution). Note that offending data is only available from 2009.

Time spent in any Corrections spell

This includes any spell under Corrections supervision and covers periods of custodial and non-custodial supervision (such as prison, Community Service, home detention, remand, parole and Periodic Detention).

Time spent in prison

An important subset of Corrections supervision spells is time spent in prison. Accordingly, we also include time spent in prison as a separate outcome.

Education qualifications

Educational achievement information is based on secondary and tertiary qualifications achieved. We include school, tertiary, industry training and targeted training qualifications data. There is a considerable reporting lag for qualifications data in the SNZ IDI, and normally qualifications data are out of date by over 12 months. In addition, qualification data only provides the year the qualification was attained. In our analysis, we make the assumption that the qualification was attained at the end of the year (ie 31 December). For analysis of intervention impacts, we exclude qualifications gained in the year the participant started the intervention as we cannot know whether they achieved the qualification before or after starting the intervention.

Here we make the assumption that gains in education qualifications reflect improvements in human capital. This may not always be the case. For example, analysis by Crichton (2013) found people on income support who achieved low-level qualifications (NQF3 and below) appeared not to gain any benefit in terms of subsequent employment or income. Similarly, this measure ignores any human capital gained through informal means.

Qualifications achieved at NQF level 2, 3, 4

For each person, we construct spells when they have achieved a specified minimum NQF level. NQF levels start at one (first national school assessment) through to nine (doctorate). For each individual, we identify the date they first achieved the specific NQF level.

Highest NQF level

The highest NQF level is the highest NQF level achieved by a person at a specific date. From this measure, we can calculate the average NQF level achieved by the participant and comparison group of an EA intervention.

Education participation

Participation in further education and training provides an early indication of whether people are engaged in developing their human capital. The unit of measurement for this

outcome is the number of days enrolled. However, people may not be attending training even when they are enrolled.

Time spent in any education participation

For any education participation, we combine all education spells in school, tertiary, industry training and targeted training.

Time spent in education while off benefit

Education participation spells where a person is also off main benefit (based on benefit entitlement spells).

Education participation NQF4

Time spent participating in education courses at NQF level 4 or above (broadly equivalent to University degree level).

Driver licence status

We are able to measure progress through the New Zealand driver licence graduation system. The system started in 1984 and has been through several changes. People go through learner and restricted stages before getting a full licence.

Time spent while holding a learner's, restricted and full licence

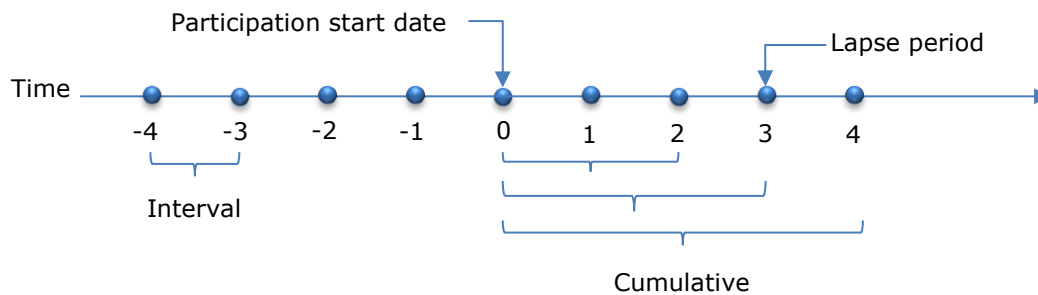
We are able to track the number of days spent at each driver's licence stage that enables us to accurately track people's progression towards a full driver's licence over time.

Tracking outcomes longitudinally

It is useful at this point to explain how we analyse the outcomes relative to participation in EA interventions. The outcomes described in the previous section are all longitudinal in nature. Therefore, we have the ability to measure outcomes at multiple points in time rather than being limited to a small number of measurement periods as would be the case for survey-based outcome measures.

This flexibility allows us to track outcomes relative to participation start dates as shown in Figure 1. The first point to make is that we measure outcomes from when people start an intervention, and this is defined as zero on our timeline. Why we choose the start date as the zero point is explained below. From the zero point, we can then create a series of lapse periods that represent the periods before and after the participation start date. Based on this timeline, we can measure outcomes in two ways: interval and cumulative.

Figure 1: Tracking EA intervention outcomes using administrative data

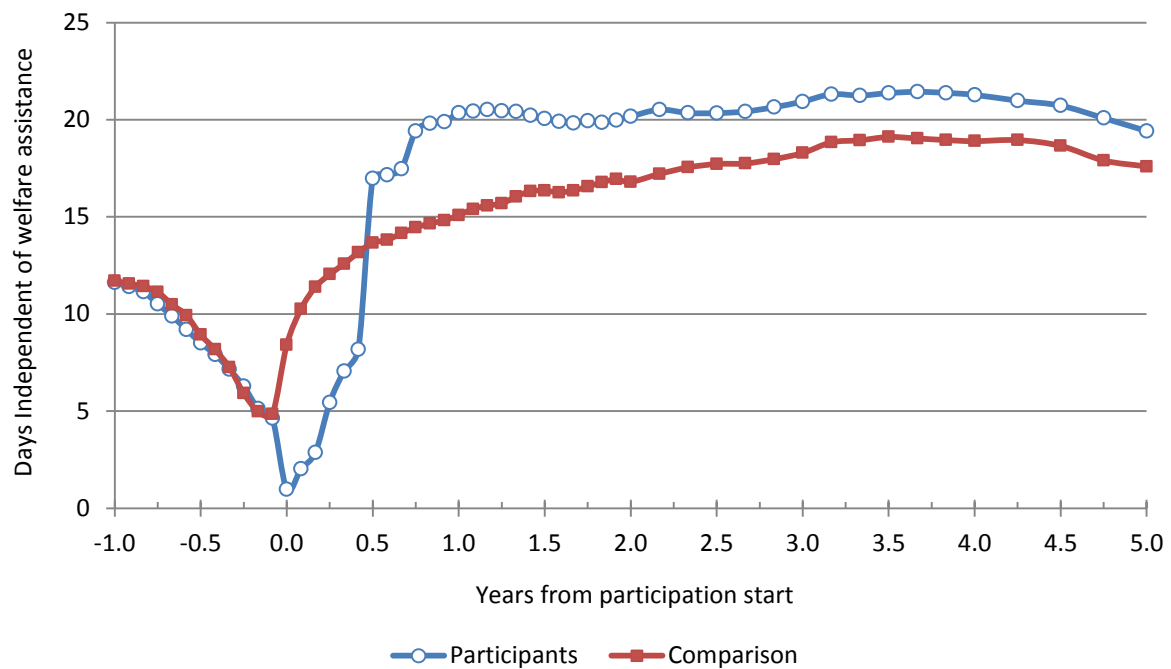


Interval outcomes

Interval outcomes are measured within a discrete lapse period, say the amount of income a person earned in the 12th month after starting an intervention. These intervals can vary in duration from one day to any period, but for EA interventions we usually use 30-day intervals. Figure 2 below shows the number of days in each lapse interval participant and comparison groups spend independent of welfare for an illustrative EA intervention. In the example, at one year before starting the intervention, both participants and comparison spent 12 days of the 30-day interval independent of welfare assistance (ie not on main benefit or receiving employment assistance).

Tracking interval outcomes is most useful in understanding the dynamic relationship between the intervention and the outcome in question. The purpose of EA interventions is to change the outcome trajectories of participants. Looking at how outcomes change in each lapse interval before and after commencing an intervention provides important information on the likely behavioural responses to the intervention. To return to the example intervention in Figure 2, we can see that the independent of welfare assistance outcomes of the participants are less than that of the comparison over the initial six months after starting the intervention (lock-in effect). However, over later intervals, the outcomes of the participants exceed that of the comparison group. In other words, after completing the intervention participants are more likely to be independent of welfare assistance than the comparison group.

Figure 2: Interval outcomes for an example intervention



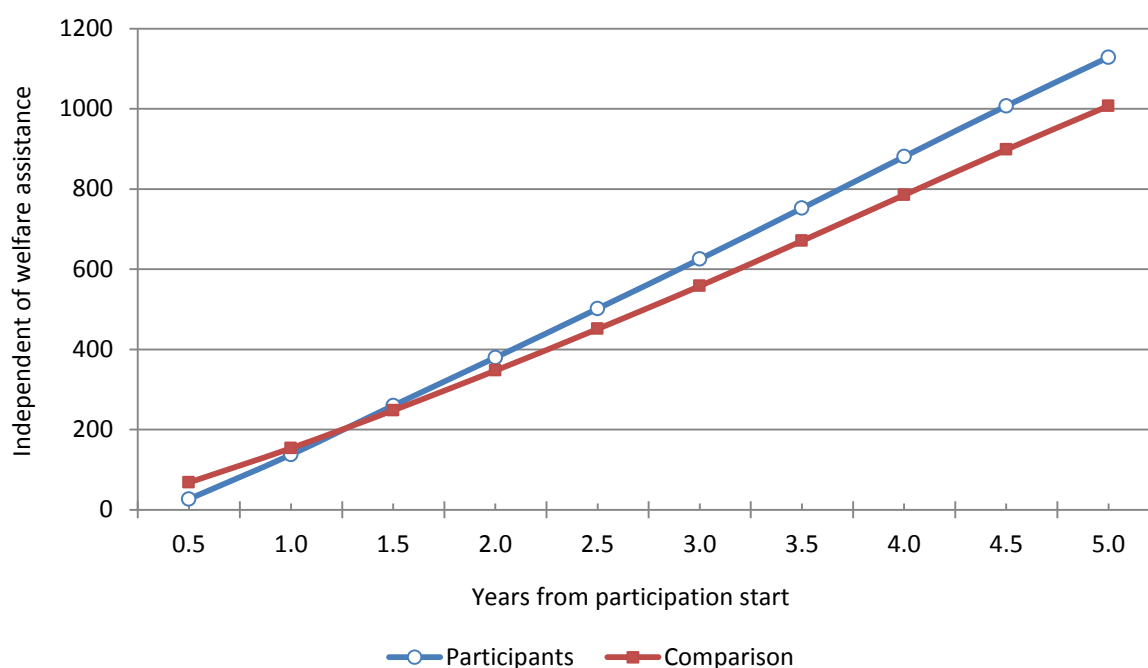
Independent of welfare assistance: not on a main benefit or receiving employment assistance, such as a wage subsidy during each 30-day interval.

Cumulative outcomes

While interval outcomes are useful to understand how outcomes and impacts change relative to when people start an intervention, they do not allow us to quantify the overall impact of an intervention. To make summative judgements we use cumulative outcomes. As Figure 1 (previous page) shows, cumulative outcomes start from period 0 to each subsequent follow-up period. For example, how much income did participants receive over the 12 months after starting the intervention?

Figure 3 shows the cumulative outcomes for the same intervention illustrated in Figure 2. By definition cumulative outcome measures only cover the period after participation start and not before. Figure 3 shows the average days over each successive period after starting the intervention that participants and comparison spent independent of welfare assistance. At the end of two years, participants spent an average of 379 days, while over five years this increased to 1,128 days. In other words, the cumulative outcomes are simply the sum of the outcomes achieved in each successive interval after the intervention started. Therefore, after five years there was a difference of 121 days between the participant and comparison group.

Figure 3: Cumulative outcomes for an example intervention



Independent of welfare assistance: not on a main benefit or receiving employment assistance, such as a wage subsidy since participation start.

Why measure outcomes from participation start?

A common question is why we measure outcomes from when people start an intervention, rather than when they finish. There are two reasons. The first is practical, (as discussed in the section on measuring, when people participate in EA interventions) as when people finish an intervention is often poorly recorded. Therefore, the date when people actually finish participating in an intervention is much less certain than the date they started.

The second reason is the importance of capturing the full impact of an intervention. As Figure 2 above shows, the period while a person is on a programme can have an impact on their outcomes. The most common impact is referred to as the lock-in effect. As the name suggests, while people are participating in an intervention they are less likely to achieve an outcome, such as moving into employment. This can occur for a number of reasons. One is simply the reduction in time participants have to look for work. And, for training programmes, the need to complete the course to gain a qualification provides an incentive to turn down job opportunities if they do arise. If we did not include these effects, we run the risk of overstating the positive impact of interventions.

Migration and mortality

In the current analysis, we have not adjusted for people moving out of New Zealand or dying. These events would, over longer follow-up periods, reduce the denominator for each of the above outcome measures (ie right censoring). We plan to adjust for right censoring in subsequent updates to this analysis.

Estimating the observed impact of EA interventions

To rate the effectiveness of EA interventions we need to determine the impact of EA interventions on outcomes to date. In this analysis, we estimate effectiveness using counterfactual designs. The term counterfactual refers to the question: what would have happened in the absence of the intervention?³ By definition, it is not possible to observe the counterfactual outcomes of participants. The solution is to identify a proxy for the counterfactual, usually a group of non-participants whose outcomes we use for comparison purposes. The challenge is to ensure that the comparison outcomes are an accurate representation of participants' counterfactual outcomes. Specifically, other than programme participation, are there other reasons for any differences between the outcomes of participants and those of the comparison group (ie selection bias)?

Various methods are able to control for selection bias to a greater or lesser degree. To assist readers in judging the robustness of a particular counterfactual design, we categorise methods according to the Scientific Maryland Scale (SMS). The SMS scale ranks counterfactual designs from 1 (least robust) to 5 (most robust). Robust in this context refers to the level of confidence we have that the impact estimate of a design accurately quantifies the causal effect of the intervention on the outcome.

In the current report, we have four designs: randomised control trial (SMS 5), propensity-matched comparison group (SMS 3), propensity-matched historical comparison group (SMS 3(-)) and natural experiments (SMS 3) designs. We outline each in turn.

Methods used to estimate the impact of interventions

Rating the effectiveness of EA interventions is a three-step process. The first step is to estimate the observed impact of an intervention on participants' outcomes to date. The second step is to estimate the long-term impact based on observed short- and medium-term impacts. The final step is to apply standard rules to determine the effectiveness rating of each intervention.

Randomised control trial designs

Interventions covered are: Mental Health Employment Service Trial, Sole Parent Employment Service Trial, Work Focused Case Management (General), Work Focused Case Management (pilot), Work Focused Case Management HCD, Work Focused Case Management ICS (Entrenched), Work Focused Case Management Integrated Services (IS), Work Search Support, Work Search Support (pilot)

Randomised control trial (RCT) designs are the most robust counterfactual designs as they require the fewest assumptions and therefore can make the strongest quantitative

³ It is important to emphasise that quantitative counterfactual designs are not the only or primary evaluation method. To fully understand the effect of an intervention requires a mixed method approach. Specifically, we need additional information to help understand the context and operation of the intervention itself to fully explain why the intervention has the impacts that it does. Similarly, not all outcomes are always quantified in a way suitable for impact evaluation.

statements about the causal relationship between intervention and outcomes. RCTs in the context of MSD EA interventions have been used most extensively to evaluate the impact of case management services, such as Work Focused Case Management or Investment Approach Trials (MSD, 2018a).

Propensity matching

Propensity matching is the main method we use to estimate the impact of EA interventions. Propensity matching is a common alternative to randomisation. It estimates the counterfactual by constructing a matched group of non-participants who have the same (or similar) characteristics as the participants. These non-participants are drawn from the population who have received income support. Propensity Score Matching is one of a group of methods referred to as quasi-experimental that attempts to replicate the same conditions as a randomised control trial. However, in all instances, quasi-experimental designs rely on additional assumptions that make them less robust than RCTs.

Before outlining propensity matching, it is useful to think of an intuitively appealing alternative of exact matching. Exact matching, as the term suggests, is to match a participant to a comparison with the same characteristics (eg same age, gender, benefit history and so on). However, exact matching is limited by the probability that two people share the same set of observable characteristics (and is also unnecessarily restrictive).⁴ The more characteristics included in the exact match, the less likely it is to find a comparison person with the exact same characteristics for each participant. As a result, these methods require the arbitrary selection of only a few matching variables.

Propensity matching overcomes this problem by using a logistic regression model to relate observable characteristics to programme participation. The logistic regression produces an estimate of how likely a given individual would be a participant in a programme. It is possible to use this likelihood (called "the propensity score") to match participants and non-participants based on the similarity of their propensity scores. If the propensity score is properly specified, the participants and matched comparison groups will have a similar observable characteristic profile (eg similar duration, benefit type, age, number of children).

Conditional Independence Assumption

A key assumption for propensity score matching is the Conditional Independence Assumption (CIA), which states that controlling for differences in observable characteristics between the participant and comparison groups also controls for unobserved differences between the two groups. Estimating the impact by controlling for observable characteristics requires that the CIA holds. If it holds, the only statistically significant difference between the participant and comparison groups will be their participation in the programme. Any resulting estimates would be unbiased. In other words, the only explanation for differences in outcomes between the two groups would be whether they participated in the programme. If the CIA fails, the estimates will be biased. Here differences in outcomes could be due to unobserved differences between participants and their comparisons, as well as the impact of the programme.

⁴ Within a randomised control trial, the treatment and control groups share the same statistical profile, not that each treatment group member has an identical twin in the control group.

The main limitation of the propensity matching method is that it relies on available and measurable information about people likely to participate in the EA intervention. It is rare that comprehensive information exists about the types of people who participate in the programme or those who could form part of the comparison group. The analysis relies on the information available on MSD's administrative databases. This increases the risk of biased estimates. To address this, we have started to use a wider range of information available in the IDI to match comparison groups. The first test case for this is the evaluation of driver licensing programmes. In the future, we plan to match comparison groups for other programmes in the IDI. The second limitation of the CIA is that it is not possible to determine whether it has been violated or, if it has, to what extent.

Quality of the matched groups

While we cannot test if the CIA has been violated, we can check to see if the comparison group has a similar profile to the participants, on average. This refers to the balancing test. This section of the report summarises our approach and is covered in detail in the technical report on Propensity Matching MSD Interventions (MSD, 2018b).

A quick way of checking for this is to count the number of variables that show a statistically significant difference between the participant and comparison group. We can then apply a rule of thumb, for example, if more than 5% of significant tests indicate that there is a difference, then we say the matched comparison group is not very similar to the participants.

The current method that we use is to calculate:

1. Mean differences between participants and comparison group for continuous variables.
2. Chi-square goodness-of-fit tests comparing observed counts in the comparison group against expected counts (calculated by multiplying the proportions in the participant group by the size of the comparison group). If the test is significant, this is followed by a post-hoc comparison of proportions at each level of the categorical variable. The purpose of this is to only do additional tests where warranted instead of doing a comparison of proportions for each level for all categorical variables.
3. Adjust the p-values for doing multiple testing by controlling the false discovery rate using the Bonferroni-Holm method. Currently, we work with accepting that 5% of rejected null hypotheses will be mistakes.

We can also check the overlap in propensity scores between participants and the potential comparison group. If both groups are similar, then the distribution of the propensity scores for the participants should substantially overlap with the distribution of the scores for the non-participant group. This overlap is referred to as common support, if common support is poor then the participants are quite unlike the population of non-participants and it is less likely that we can match a suitable comparison (ie one that meets the balance test).

Another way is to look at the impact in each month before the start of the programme for the different outcomes that we are measuring. For each outcome, there should ideally be no differences between the participant and comparison group before starting on the programme.

Non-standard propensity-matched analysis

Interventions covered are Youth Service (NEETS), Youth Transitions Service.

In addition to the standard MSD propensity matching process, we have included propensity score matching impact estimates for the two services targeted at young people transitioning from school to employment or education and training. This analysis was undertaken by the New Zealand Treasury (Dixon & Crichton, 2016).

Propensity score matched historical comparison group

Interventions covered are Youth Service (YP), Youth Service (YPP).

For two EA interventions (Youth Service Youth Payment, Youth Service Young Parent Payment) there was no contemporary non-participant population. Instead, the analysis constructed a propensity-matched comparison group based on a similar population who received the Independent Youth Benefit in the past. The comparison group resembled the Youth Service participants on average but were exposed to different policies and labour market conditions. The analysis was also undertaken by The Treasury (McLeod, Dixon, & Crichton, 2016).

Natural experiments

Interventions covered are Jobseeker Work Ready 52-week benefit reapplication, WRK4U.

Natural experiments are instances where an EA intervention is introduced in such a way that we have a natural comparison group. The key assumption of natural experiments is that the introduction of an EA intervention is unrelated to differences in future outcomes between participants and comparisons in the absence of the intervention or, if any differences do exist, they can be controlled for. For example, in the current EA report, we used a natural experiment to evaluate the impact of the 52-week reapplication process on exits from benefit and how soon affected people returned to benefit. We used information on the behaviour of jobseekers in the years before the introduction of the 52-week reapplication process to provide a baseline comparison for those affected by the new policy. Because the policy was introduced nationally, we had to include labour market measures into the analysis to help control for changes in labour market conditions before and after the introduction of the 52-week reapplication process (MSD, 2013).

Likewise, we evaluated the impact of the WRK4U seminar by comparing the behaviour of jobseekers in three trial sites before and after the intervention as well as the behaviour of jobseekers in non-trial sites before and after the intervention (de Boer, 2003).

Other considerations

Making multiple statistical inferences

When presenting summative statements about the effects of many EA interventions on these outcomes on different subgroups of participants, we are making hundreds of statistical inferences at a time. There is a chance that some of these inferences are incorrect. Specifically, we are worried about claiming that an impact exists when there is none.

For example, imagine an EA intervention that had no impacts on any outcomes in the general population. If we took a sample of participants and comparison group members and analysed 100 impact results on different outcomes for that intervention, we would expect some of these to be statistically significant due to sampling variability. This is known in the statistical literature as the multiple comparisons problem.

We have considered two approaches to adjust for this. The first is to control the familywise-error rate, and the second is to control the false discovery rate. We favour controlling the false discovery rate because we think the additional power that it gives is worth the trade-off in making a proportion of inferences that are incorrect. Controlling the familywise-error rate (by using the Bonferroni method, for example) may lead to us incorrectly rejecting many null hypotheses when we should not be doing so. The other reason is that we observe a relatively large proportion of significant tests. Therefore, the challenge is to screen a large number of signals to determine if we should look more closely at the impacts of a particular programme.

Interpretation of EA impacts in the context of multiple interventions

As the analysis makes clear, Service Delivery runs many different types of employment assistance interventions. Moreover, any one individual may receive one or more interventions over time. Therefore, it is important to understand what an impact estimate for an individual EA intervention is telling you.

We are estimating the impact of participating in an EA intervention at a point in time

Participations in EA interventions occur over time, a person may participate in only one or a series of either the same or different interventions. When we estimate the impact of an EA intervention, we are looking at the single event, namely the date a person starts an EA intervention. We are comparing this to a similar person who did not participate in the intervention on that date. Anything that happens after this date is regarded as an outcome, including subsequent participation in EA interventions. For example, an EA intervention may well increase the probability of participants receiving additional assistance relative to the comparison group. This is interpreted as an impact of the initial EA intervention. But this also means that the impact on longer-term outcomes, such as employment or income is a combination of the initial EA intervention as well as subsequent assistance. Currently, we do not have reliable techniques to try and disentangle these effects.

A further point to make about the comparison group, other than the participation selection period,⁵ is that we do not exclude any comparison group member who subsequently goes onto to receive the EA intervention being evaluated. More generally, the comparison group will also receive other types of EA assistance over the outcome period. Therefore, when we report an impact, it is the marginal effect of the EA intervention relative to the average level of assistance received by the comparison group. This is an important point to keep in mind, as in some instances, a specific

⁵ This is usually a calendar year. So we identify all participants in an EA intervention in a given year (say 2018) and define everyone who did not participate in the intervention in 2018 as non-participants.

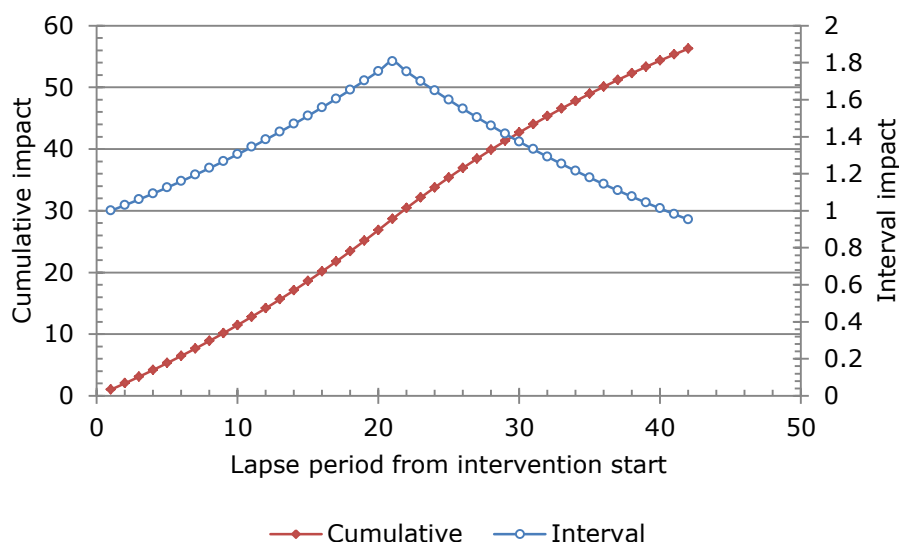
intervention may appear to have no impact because the comparison group is receiving similar levels of assistance or near proxies. These issues point to the need to carefully examine the relative experience of the comparison group against the participants to properly interpret the observed impacts.

Estimating future impact from observed impact

In general, the period that we can observe outcomes over is shorter than the period that an intervention has an impact on participants' outcomes. In addition, EA interventions often have negative short-term impacts, such as lock-in effects,⁶ while positive impacts tend to occur over the medium- to long-term. Taken together, if we judge EA intervention effectiveness over a too short follow-up period, we are more likely to rate the intervention as ineffective by including short-term negative impacts and failing to include potential long-term positive impacts.

Figure 4 gives a stylised example of this problem. For the hypothetical EA intervention's impact on time off main benefit, Figure 4 shows the interval impact (which is defined as the impact **within** a particular lapse period) steadily increasing until month 21 after the start of the intervention before it begins to fall. For example, at month 21, the difference in time off benefit between the participant and control group is 1.75 days. The cumulative impact, on the other hand, is the difference in the outcome **since** participation start (this measure is a cumulative sum from participation start up to a given lapse period). To continue the above example, the difference in cumulative time off benefit at month 21 is 30.43 days (ie the sum of all the interval impacts up to and including month 21).

Figure 4: A stylised example of the relationship between interval and cumulative impact on time off benefit



⁶ Lock-in refers to the phenomenon that, while on the intervention, participants are less likely to move into employment than the comparison group. As a result, when participants finish an intervention, their average time on benefit is longer than that of the comparison group. Therefore, if the intervention increases their employment prospects at completion it still takes time after completion before the intervention has a cumulative positive impact.

Turning our attention to the last data point in Figure 4 (month 41), we can see that the interval impact is greater than zero (impact: 0.95 additional days off benefit in month 41). What this tells us is that we have not seen the full impact of the intervention on time spent off main benefit. This will occur when the interval impact converges to zero.

The challenge in this analysis is to estimate the unobserved interval impact to be able to get an estimate of the full cumulative impact on participants' outcomes. We do this using a three-step process:

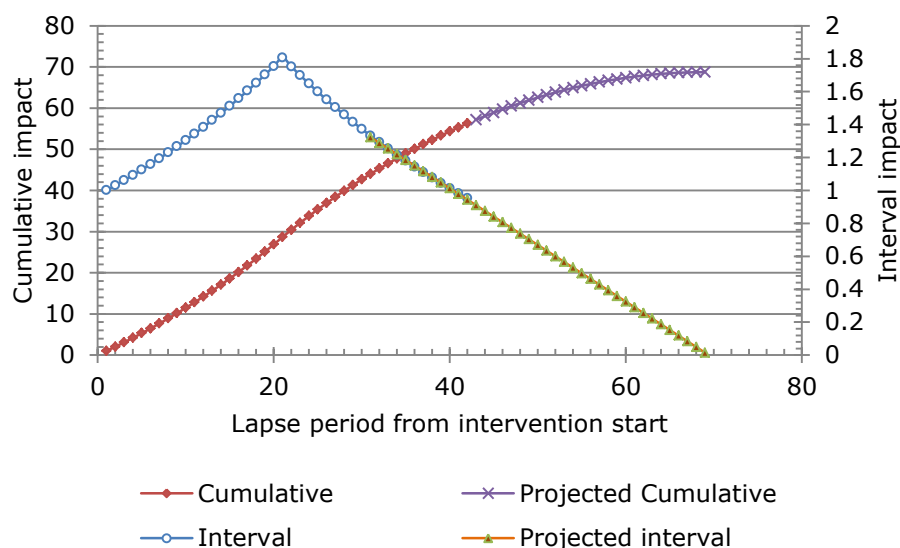
1. Based on the entire participant group, we project the interval impact until it converges on zero. If the natural trend is away from zero, we force the trend towards zero.
2. Using the projected interval impact we calculate the projected cumulative impact (ie add up each projected impact over successive lapse periods).
3. Using the projected cumulative impact results from step 2, we add the trend in cumulative impact to the observed impact with appropriate scaling if required.

Below is a more detailed outline of each of the above steps.

Step 1: Estimate the expected interval impact

The first step is to estimate the trend in the interval impact (Projected interval line in Figure 5). We use the last 12 observed impact intervals and take a least squares regression estimate of the interval impact by interval duration. We run the regression model estimates through to unobserved lapse periods until the interval impact reaches zero. We discuss below how we handle instances where the interval impact is trending away from zero.

Figure 5: A stylised example of projecting the interval impact



Step 2: Calculate the projected cumulative impact

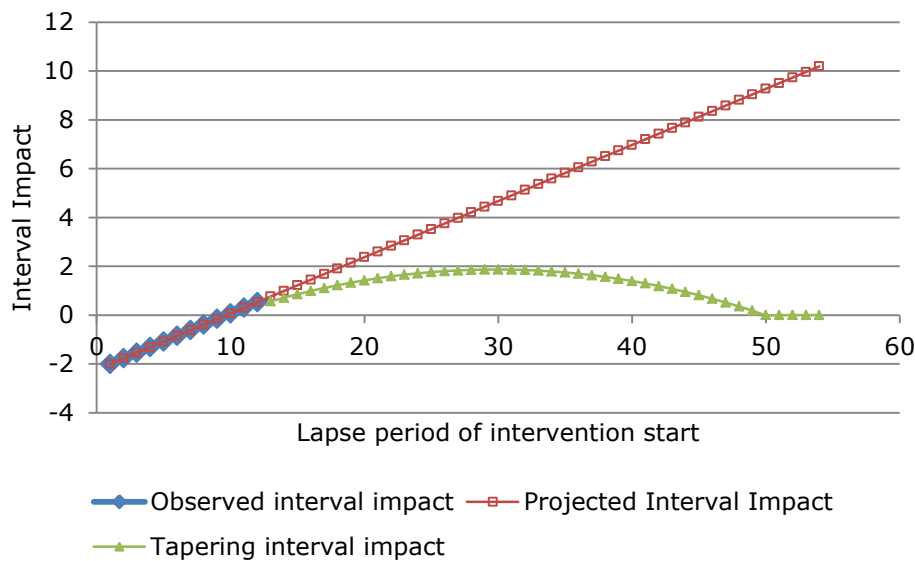
The second step as shown in Figure 5 is to take the last observed cumulative impact and add the projected future interval impacts to construct the projected cumulative line shown in the graph. We stop adding the projected interval impacts when the last interval

impact is zero (this occurs at lapse period 68 on the graph). We have estimated the expected full impact of the intervention once this occurs. In this example, we estimate the full impact is likely to be observed after 68 months; at this point, the full impact of the intervention is estimated to be 68.73 days.

Interval impacts that do not trend towards zero

In practice, we find a number of instances where the projected impact either has a trend that is away from zero (resulting in infinitely large impacts) or is constant over time (this result is more plausible). In both these instances, we have chosen to force the interval impacts to zero. Our main motivation for this decision is to ensure that the resulting estimates are plausible and to limit the influence of projected impacts on the analysis. Our method for forcing projected interval impacts to zero is by applying a proportional decrease in the interval impact from the first projection interval. In other words, the interval impact is reduced by a set proportion, with this proportion increasing as the projected period increases (so that the reduction eventually reaches 100%). Figure 6 illustrates how the forced taper would apply to an increasing projected interval impact. As the projection period increases the proportional reduction increases forcing the projected interval impact to eventually decrease to zero. In the current analysis, the proportional reduction increases at a linear rate of 0.05% for each day of the projection period.

Figure 6: Forced taper in the projected impact of an intervention



Step 3: Project cumulative impact from observed cumulative impact

The final stage in estimating the projected impact for an EA intervention is to take the last observed cumulative impact and then include the projected cumulative impact. Here we face two issues that need to be addressed:

- scaling the interval impact to the cumulative impact for each EA intervention participant group
- estimating the confidence interval for the projected impact.

Scaling interval impacts

For each EA intervention group, we compare the last 12 observed interval impacts to the series of projected impacts and calculate the ratio between the two. For example, if a particular EA intervention group is showing higher observed impacts than projected then the ratio would be greater than one. From these last 12 intervals, we calculate the average ratio and then scale projected interval impacts by this ratio. Once scaled we can then add each projected interval impact to the last observed cumulative impact to arrive at the total cumulative projected impact.

The confidence interval for projected impact

The second issue is to provide an estimate of the confidence interval for the projected cumulative impact. There are two sources of uncertainty for the projected impact:

- the observed impact has a given intrinsic level of uncertainty
- the projected interval impact is itself also an estimate with its own level of uncertainty.

In the current analysis, we only include the uncertainty from the first source. We plan to look at including the uncertainty introduced through the projection process itself in later updates. Therefore, the confidence intervals for the projected impact understate the true uncertainty for these estimates.

To reflect the confidence intervals for the observed impact in the projected impact we used Monte Carlo simulations by taking random draws from the observed cumulative impact distribution and running the projected impact calculation for each draw. We repeated these simulations 1,000 times and took the 2.5 and 97.5 percentiles as the 95th confidence intervals for the projected cumulative impact result.

Rating the effectiveness of interventions

This section outlines how we systematically rate the effectiveness of interventions based on their impacts on outcomes. The aim of providing a rating is to qualitatively summarise the effectiveness of an EA intervention from its quantitative impacts. The goal here is to ensure that all EA interventions are rated in the same way and that the rating process is transparent.

Rating by outcome domain

For each EA intervention, we have one outcome measure grouped under each broad outcome domain. In the current effectiveness report, we focus on five outcome domains: income, employment, justice, educational qualifications and independence from welfare.

At present, we select one outcome measure to provide the summative assessment for the impact of each EA intervention on that domain. In the current analysis:

- income effectiveness is based on the EA intervention's impact on net income from all sources
- employment effectiveness is based on the impact on any time in employment
- justice is the time that participants spend in correctional services

- qualifications: increase in average NQF level
- independence from welfare assistance is based on time spent independent from Work and Income Assistance (ie not on main benefit or participating in EA interventions).

Translating impact to an effectiveness rating

For each outcome, we examine the observed and projected cumulative impact and categorise intervention effectiveness as shown in Table 6. In our analysis, we start with an initial assessment based on the observed impact and then adjust this assessment based on the projected impact. The higher weight given to the observed period is because it has an empirical basis, while the projected impact is sensitive to the most recent trend in the observed impact (see page 31). The projected impact serves to moderate the observed impact in those instances where the two differ (ie in the off-diagonal cells in Table 6). For example, if an intervention has a significant negative observed impact and a significant positive projected impact, we only increase the rating from negative to likely negative, rather than to promising. In practice, the majority of observed and projected impacts are consistent with each other in terms of sign (ie they are either both positive, or both negative).

Table 6: Rating of outcome domain by the impact on outcomes

		Projected impact		
		Significant positive	Zero	Significant negative
Observed impact	Significant positive	Effective	Effective	Promising
	Zero	Promising	No difference	Likely negative
	Significant negative	Likely negative	Negative	Negative

Rating the overall effectiveness of an intervention

Once we have an effectiveness rating for each outcome domain we then combine these ratings to arrive at an overall rating of a programme. Because we are combining five outcome domains, the number of combinations of results becomes much greater. We use the following steps to determine an intervention's overall effectiveness.

1. Convert outcome domain impacts into numerical values: positive = 1, likely positive = 0.5, no difference = 0, likely negative = -0.5 and negative = -1.
2. Taking the average of these, values are then categorised as follows: Effective over 0.66, Promising between 0.3 and 0.66, No difference is between -0.3 and 0.3, Likely negative between -0.3 and -0.66, and Negative less than -0.66.
3. If there is a significant positive and negative impact then the rating is Mixed.
4. If the rating is less than 0.3 (No difference) or Mixed and the outcome period is less than two years then the rating is Too soon to rate.

Based on these rules, the definition of each of the effectiveness ratings is as follows.

Effective: EA interventions are rated effective only if they are effective against the majority of outcome domains and they show no sign of having a negative impact on any

other outcome domain. We do not wait two years before rating a programme as effective.

Promising: promising interventions are those that are effective or likely effective for at least one outcome and show no negative effects.

Mixed: mixed covers interventions that show both positive and negative effects across outcome domains. We wait until we have two years of outcome data before rating a programme as mixed.

Makes no difference: includes all EA interventions that have no effect on any outcome domain. We wait until we have two years of outcome data before rating an intervention as making no difference.

Likely negative: interventions are in this group because either a minority of outcome domains are rated as negative with the remainder having no impact. Or, the majority are negative, with a minority having the possibility of being positive. We wait until we have two years of outcome data before rating an intervention as likely negative.

Negative: interventions where the majority of outcome domains are rated as negative. We wait until we have two years of outcome data before rating an intervention negatively.

Too soon to rate: with the exception of interventions rated as effective or promising, interventions with less than two years of observed impacts are rated as too soon to rate. The reason for waiting at least two years is that the majority of EA interventions have negative effects in the short-term (eg lock-in effects) and it is necessary to wait some time after commencement before positive effects are potentially observed.

Not feasible: a number of interventions have been identified as where it is not currently feasible to estimate their effectiveness.

Cost-benefit analysis

We are at the start of calculating the cost-benefit of EA interventions covered by our analysis. For this year, we have calculated the welfare Return on Investment, but in future updates, we will extend this analysis to include more outcome domains.

What does Welfare Return on Investment consist of?

In the 2018 version of EA cost-effectiveness results, we only report on Welfare Return on Investment (wROI). This is a ratio of savings in terms of benefit payments and other administrative costs to programme investment.

The return on investment is:

$$\frac{\text{Net Benefit Payments} + \text{Net Other}}{\text{Total Programme}}$$

Where:

Net Benefit Payments: is the difference in benefit payments between the participant and comparison groups. This includes the difference in estimated future benefits costs at the end of the observation period.

Net Other: is the difference between the treatment and comparison groups in the utilisation of other programmes and case management services.

Total Programme: is the total cost of the programme (excluding overheads) for participants.

Where the ratio is greater than one, then we can conclude that the net return of the intervention exceeds its cost (ie it is cost-effective). However, because of uncertainty in the estimates and the desire to see a substantive net return above estimate, we use 1.5 as the cut off for an intervention being cost-effective. Interventions in the range between 1 and 1.5 are said to be breaking even.

Financial outcomes

Outcomes are measured by the financial savings made when MSD no longer needs to pay a person a main benefit, supplementary benefit, or a one-off payment. There are also savings in administrative costs, as MSD no longer needs to administer income support or provide intervention programmes.

People who are helped into employment may not require a benefit for many years – so much of a programme’s success may lie in the future. It would be impractical to wait for these savings to be realised before assessing the impact of a programme, so the outcomes also include a prediction of future savings as well as observed savings.

Savings and costs do not include a share of indirect costs or overheads. The cost of the employment programmes is relatively small in the context of total welfare costs and is unlikely to materially influence indirect costs and overheads.

These cost estimates come from three different sources, which are:

1. Amount of income support paid to participants and matched comparison groups
2. Amount of income support that we estimate will be paid in the future
3. MSD administrative costs and expenditure involved in administering EA interventions.

Estimates of savings come from all three sources, while estimates of investment in specific EA interventions come from the third source only. Savings are estimated by taking the difference in the cumulative average cost for the participants from the cumulative average cost for the matched comparison group. We can then estimate the wROI for an intervention by dividing the cumulative savings by the cumulative investment at the end of the observation period.

The estimated amount of income support paid in the future is sourced from the Liability Estimator Tool (LET)

MSD has developed a model called the Liability Estimator Tool (LET) which predicts the amount of income support that will be paid to individual clients in the next four years. An estimate of income support that will be paid until age 65 years can also be obtained by scaling the four-year estimate. For a high-level overview of the model, please refer to the summary report LET v5 (MSD, 2016).

MSD administrative costs and EA intervention costs are sourced from the individual Cost Allocation Model

Besides the actual income support paid to clients in the form of benefits and grants, expenditure on MSD clients also includes administrative costs associated with paying those benefits and costs involved in case management. These costs are sourced from an

individual cost allocation model maintained by Insights MSD. For a short description of the model, see the Estimating the cost of EA interventions section (page 12).

Welfare ROI for case management services

Welfare ROI for case management services is calculated differently. Welfare ROI for case management services was based on General Case Management as the baseline service. The returns consist of savings in income support costs and savings in avoided case management costs. The investments consist of costs of staff time and contract and subsidy payments for employment assistance. For these services, we did not include a projection of likely future income support savings. Refer to MSD (2018a) for more detail on how we calculated the wROI for case management services

References

- Crichton, S. (2013). *The Impact of Further Education on the Employment Outcomes of Beneficiaries*. Wellington: Ministry of Business, Innovation and Employment.
- de Boer, M. (2003). *Impact of Jump Start seminar on the uptake of the unemployment benefit*. Wellington: Ministry of Social Development (EDRMS id: A177933).
- Dixon, S., & Crichton, S. (2016). *Evaluation of the Impact of the Youth Service: NEET programme* (The Treasury working paper No. 16/08). Wellington.
- McLeod, K., Dixon, S., & Crichton, S. (2016). *Evaluation of the Impact of the Youth Service: Youth Payment and Young Parent Payment* (The Treasury working paper No. 16/07). Wellington.
- MSD. (2013). *Impact of the 52-week Unemployment benefit reapplication process Update 2: Technical Report*. Wellington: Ministry of Social Development (EDRMS id: A6516526).
- MSD. (2016). *LET v5 model information*. Wellington: Ministry of Social Development (EDRMS id: A9188745).
- MSD. (2017). *Service Delivery Cost Allocation Model for Individual Outputs: 2017 version*. Wellington: Ministry of Social Development (EDRMS id: A9317887).
- MSD. (2018a). *Cost-effectiveness of intensive case management services*. Wellington: Ministry of Social Development (EDRMS id: A9467849).
- MSD. (2018b). *Propensity matching MSD interventions: technical notes*. Wellington: Ministry of Social Development (EDRMS id: A6927997).