MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

Cost-effectiveness of intensive case management services (from October 2012 to July 2017)

Technical report

June 2019

Author

Marc de Boer, Principal Analyst, Ministry of Social Development (MSD)

Acknowledgements

We would like to thank Steven Johnston for his advice on the analysis, and to Michelle Bly, Jared Forbes, Oscar Casswell-Laird, Bryan Ku, Daniel Stoner, Jayne Russell, Antony Harvey, Evan Thompson, Bryan Ku, Hugh Miller and Tim Maloney for their comments and feedback on earlier versions of this report.

Disclaimer

The views and interpretations in this report are those of the researchers and are not the official position of MSD.

Citation

Suggested citation: MSD (2019) *Cost-effectiveness of intensive case management services (from October 2012 to July 2017): Technical report*, Ministry of Social Development, Wellington.

Published

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

Published in June 2019

ISBN

Online 978-1-98-854184-6

Contents

Acronyms

DPB	Domestic Purposes Benefit (main benefit, ceased in July 2013)
EB	Emergency Benefit
EMA	Emergency Maintenance Allowance (main benefit)
GCM	General Case Management
IA	Investment Approach
IR	Inland Revenue
ISR	Intensive Service Response
JS	Jobseeker Support (main benefit)
JS HCD	Jobseeker Support Health Condition or Disability (main benefit)
JS-Student	Jobseeker Support – Student Hardship (main benefit)
LLTBR	Likelihood of Long-Term Benefit Receipt (statistical model)
LTC	Low-trust Client
MSD	Ministry of Social Development
MSM	Multi-state Models
NSDM	New Service Delivery Model
NQF	National Qualifications Framework
RCT	Randomised Control Trial
RSS	Residential Support Subsidy
SDM	Service Delivery Model
SPL	Supported Living Payment (main benefit)
SPS	Sole Parent Support (main benefit)
WFCM	Work Focused Case Management
WFCM General	Work Focused Case Management General
WFCM Health	Work Focused Case Management Health Condition or Disability
WFCM Health (YSLP)	Work Focused Case Management Health Condition or Disability (Young Supported Living Payment)
WFCM Integrated	Work Focused Case Management Integrated Service
WSS	Work Search Support
YP	Youth Payment
YPP	Young Parent Payment

Glossary

Actual service	A service a person is assigned to each week. People may change services for unobserved reasons such as becoming ineligible or exiting main benefit.
Assigned service	A service a person is explicitly assigned to by meeting set eligibility criteria for the service that are observed by the researchers. It is the assigned service allocation that is evaluated in this analysis.
Common support	Based on the idea that the participant and comparison group should have the same profile characteristics at the assignment to a service.
Control group	As part of the evaluation, eligible people for a service can be randomly allocated to a control group. The assignment to a service control group means they cannot be assigned to the service for a set period.
Counterfactual design	A technique to estimate the impact of an intervention by comparing the outcomes of people who participate in an intervention to an equivalent group who do not participate.
Holdout period	The period a person assigned to a service control group cannot be assigned to that service. The period starts from when they are assigned to the service control group. The duration of the period varies by service and can last from one to three years.
Impact	A quantitative estimate of the change in an outcome because of an intervention (causal effect). The impact is normally estimated using a counterfactual design.
Income support	The tax-based programme of financial transfer payments to people who are not working. Income support is divided into three levels: main benefits, supplementary assistance and ad hoc payments.
Intensive case management	Services where case managers are assigned a lower caseload than for the baseline service General Case Management (GCM). The lower caseload is designed to allow case managers to spend more time with people on their caseload than would otherwise be the case.
Liability Estimator Tool	Liability Estimator Tool (LET) is a statistical risk model that estimates a person's lifetime income support costs to the age of 65. Everyone who has been on a main benefit since 1993 is scored on a weekly basis. The LET tool at the time of this report was not being used for operational decision making.
Likelihood of Long-Term Benefit Receipt	The Likelihood of Long-Term Benefit Receipt (LLTBR) is a statistical risk model that scores everyone on a working age main benefit on the probability they will remain on benefit for the next

	two years. The LLTBR is updated daily and has been used for operational decision making since 2010.
Low Trust Client	People who have a history of fraud or overpayments of income support.
Main benefit	Income support is divided into three tiers. Main benefit is tier one assistance and covers the bulk of income needs for families not in paid employment.
Non-participant effects	Interventions can have impacts on non-participants. For example, increased assistance to participants to help them move into employment can mean non-participants take longer to find employment.
National Qualifications Framework	In New Zealand qualifications are ranked on the National Qualifications Framework (NQF). The NQF is from 1 (lowest) to 9 (highest), with people gaining credits at each level. Once they have sufficient credits in the right combination of subjects they gain a qualification at that level.
Outcome	A state that a person is in on a dimension of wellbeing. Normally, outcomes refer to states that an intervention is trying to change (see impact).
Partner	Income support entitlement is based on family status. For families with more than one adult, each adult is either the primary or partner.
Primary client	Income support entitlement is based on family status. For families with more than one adult, each adult is either the primary or partner. Historically, the emphasis has been on assisting the primary client.
Service	General term to refer to a case management service.
Service Delivery Model	The system introduced in July 2013 to allocate people receiving income support to specific case management services.
Site	A local office responsible for case managing people living in the immediate geographical area.
Work test	People with work obligations can be required to undertake activities, such as meeting with a case manager or attending a seminar. These activities constitute a test of whether a person is meeting their work obligations. Failure to undertake required activities can start a sanction process that can result in a temporary reduction in payments or even cancellation of their income support entitlement.
Work test exemption	People on work-obligated benefits can get short-term exemptions from these obligations. For example, people on Jobseeker Support can get a medical certificate that gives them the status of Health Condition or Disability.

Young Parent Payment	A main benefit paid to teen parents up to the age of 19.
Youth Payment	A main benefit paid to young people aged under 18 who no longer receive the support of their caregivers.
Youth Service	A programme in which all people receiving a Young Parent Payment or Youth Payment must participate. Contracted providers who run the Youth Service programme help participants gain educational qualifications and move into employment.

Overview

This report provides the technical information on estimating the welfare ROI of intensive case management services delivered by the Ministry of Social Development.

The report is in seven sections.

1: Overview of the Service Delivery Model and evaluation: before going into detail we start by providing a broad overview of the Service Delivery Model and the evaluation approach taken to evaluate its cost-effectiveness.

2: Service Delivery Model: provides an outline of how intensive case management services operated and the process by which people were allocated to different services.

3: Representation of service status in the evaluation: to enable the evaluation of the effect of intensive case management services, we needed to create a derived service history. The purpose of the derived service history was to enable the evaluation of the impact of each service on the time people spent off main benefit and to avoid introducing selection bias from uncontrolled service exits into the analysis.

4: Control group: outline of how the service allocation process allocated potential service participants to a randomised control group. Here we look at:

- ethical considerations in using a control group
- implications of the control group design for the impact evaluation
- re-weighting of control group members.

5: Participant and control group profile: comparison of the observed profile of the participants and control group between original weighting and after adjustment to the weighting.

6: Estimation of impacts: outlines how we estimated the impact of each service on the time participants spent on main benefit. We achieved this by estimating the time to benefit exit while on the service and the time to return to benefit if they exited, and what income support rate they received in each benefit state.

7: Estimation of welfare cost-benefit: based on the estimated time off main benefit, income support costs and service costs, this section outlines how we calculated the relative cost of each service and the welfare benefits (ie income support payments) from the service.

Limitations of the analysis

Readers should consider the following limitations of our analysis.

- Overall effectiveness of the Service Delivery Model: The emphasis in this evaluation was on the difference intensive case management made over and above GCM (the lowest intensity service available). What we did not test was whether the overall case management model was more effective than the case management practice in place before July 2013.
- Time on main benefit: We confined our analysis to the impact of WFCM and Work Search Support services on the time spent off main benefit (excluding temporary suspensions to benefit entitlement). We do not know whether changes in off benefit

outcomes were because of increased employment or driven by other, less desirable, outcomes (ie choosing to move off income support without having an alternative source of income).

- Non-participant effects: We were unable to examine whether the impact of intensive case management on moving participants off income support came at the expense of other groups. We do not know whether equivalent people in GCM spent longer on benefit because of the additional assistance given to people in WFCM and Work Search Support services (ie crowding out). If these non-participant effects did occur, then they would offset the participant impacts reported here.

Summary

Both the Service Delivery Model and the evaluation are complex in their design and construction. For this reason, it is useful to provide a high-level summary to help the reader understand how the Service Delivery Model operated and the challenges this posed for the evaluation, before describing in detail each aspect of the SDM and evaluation.

Case management services

From July 2013, everyone receiving a main benefit was allocated a case management service. Each service varied both in the intensity (the ratio of participants to case managers) and diversity (how similar participants are to each other, such as being on the same type of main benefit). Case management services were ranked in priority, so that a person could not be assigned to a lower priority service. The lowest ranked service was General Case Management (GCM), which was the default service people were in if they had not been assigned to any other case management service.

Service Delivery Model was about allocating people to case management services

At its most basic level, the Service Delivery Model (SDM) was an allocation tool that ensured each case management service was filled up to its capacity with people eligible for the service. This allocation process operated on a weekly basis and involved the following steps.

1 For people currently allocated to a case management service, the SDM identified who was no longer eligible (either because they had exited benefit, or their circumstances had changed so they no longer met the service's eligibility criteria). These people were removed from the service's allocation.

2 For each service, the SDM then identified people who were eligible and were currently allocated to a lower priority service. If there were spaces available (ie the service was below capacity) then the SDM assigned these eligible people to the service to fill available spaces.

3 After everyone was allocated to a service, IT systems were updated to show what service each person was in and frontline offices received updated lists of people assigned to each case management service. Case managers within each service were responsible for allocating participants to their individual caseloads.

4 The important points to understand about the allocation process are:

- The process was entirely based on administrative data and algorithms. There were no staff judgments made about who would go to what service.[1] However, in some instances, staff could make recommendations, but these had to be recorded in the IT system to be included in the allocation process.

[1] There was an exception, when the Canterbury region opted out of the SDM process and allocated people to services manually. This lasted for around nine months before the region opted back in to the SDM, largely because of the high volumes of allocation decisions that were difficult to manually process.

- Over the evaluation period, there were no formal exits from a service. In other words, people left a service either because they were assigned to a higher priority service or they became ineligible (eg their circumstances changed to make them ineligible or they exited main benefit).
- Once case managers received the list of people assigned to their service, they had discretion (within some operational limits) to work with each individual as they saw fit.

Evaluation strategy

Based on this brief description of the SDM, we now turn to examine the evaluation objectives and design.

Evaluation objective: estimate the impact of being assigned to a case management service

Within the SDM, the only active decision was to assign a person to a service. The question was whether those assigned to each service benefited (ie had better outcomes) and whether these improved outcomes outweighed the higher cost of the service.

Based on this question, the purpose of the evaluation was to estimate the impact of assigning a person to a given service (as described in step 2 above). We refer to this as the assigned service start date. The follow-on question is determining when a person's assigned service spell ends. For this evaluation, a person's assigned service spell ends when they are assigned to another service (ie a repeat of step 2 of the SDM).

What happened if a person left a service before being assigned to another?

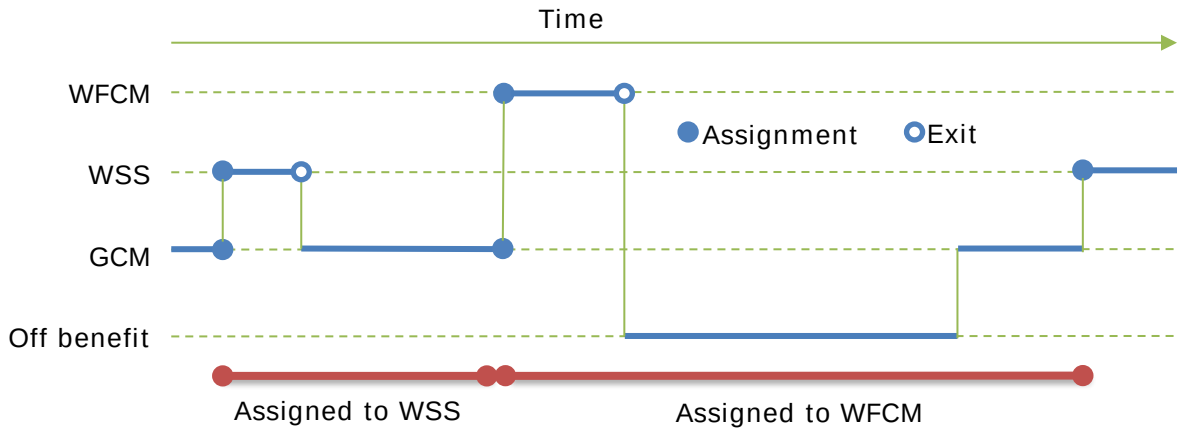
Within the assigned service spell period, a person may well have changed services by either exiting main benefit or becoming ineligible for the service. These events were all treated as part of the experience of being assigned to the service. The reasoning here is that both exits from benefit or becoming ineligible could well have occurred in response to being assigned to a service. In the case of benefit exits, the objective of more intensive case management services was to move people into employment and therefore movement off main benefit indicates progress towards this outcome. But, it can also be argued that more intensive case management could have motivated participants to change their circumstances to make them ineligible for the service.[2]

To help illustrate these features, Figure 1 shows a stylised history of an individual being assigned to different case management services within SDM. In this case, the individual starts in GCM before being assigned to WSS, this is the start of their assigned WSS spell. After a period on WSS, the individual becomes ineligible and is re-assigned to GCM (the default service). After a period, the person is assigned to WFCM. At this point, their assigned WSS spell ends and their assigned WFCM spell starts. While on WFCM the individual exits from main benefit and, therefore, is not in any service. On returning to

[2] To illustrate this point. People over the age of 55 on Unemployment Benefit had no work obligations. In 2003, this age threshold was increased to 60. The immediate response was for approximately one-quarter of newly work-obligated individuals to shift to a Sickness Benefit as this had no work obligations. The reason these people had not been on Sickness Benefit before was because of the additional cost of obtaining a medical certificate to maintain eligibility for Sickness Benefit.

main benefit they are initially allocated to GCM until they are assigned back to the WSS service. It is at this second assignment to WSS that the individual's WFCM assigned spell ends.

Figure 1: A stylised example of a service assignment history for an individual



Estimating the impact of individual services

Given that we wanted to estimate the impact of being assigned to a given service, the next step was to determine an estimation strategy. The challenge for evaluating case-management services within SDM was identifying a suitable comparison group who were not assigned to the service being evaluated. Because everyone who was eligible could be assigned to a service (within capacity) there was a very real risk that all eligible individuals would be assigned to the service over any period. Under this scenario, we would not have enough comparable people experiencing different case management services. The issue of a lack of a comparison group is referred to as the common support problem.

To overcome this challenge, MSD agreed to include a randomised control trial as part of the SDM allocation process. The randomisation occurred at step 2 in the allocation process described above. Step 2 was split into two stages:

2.1 For most of the case management services, the allocation process waited until two criteria were met. The first was that there were 10 places to fill in the service. The second criteria required that there were 11 eligible people to fill these 10 places.

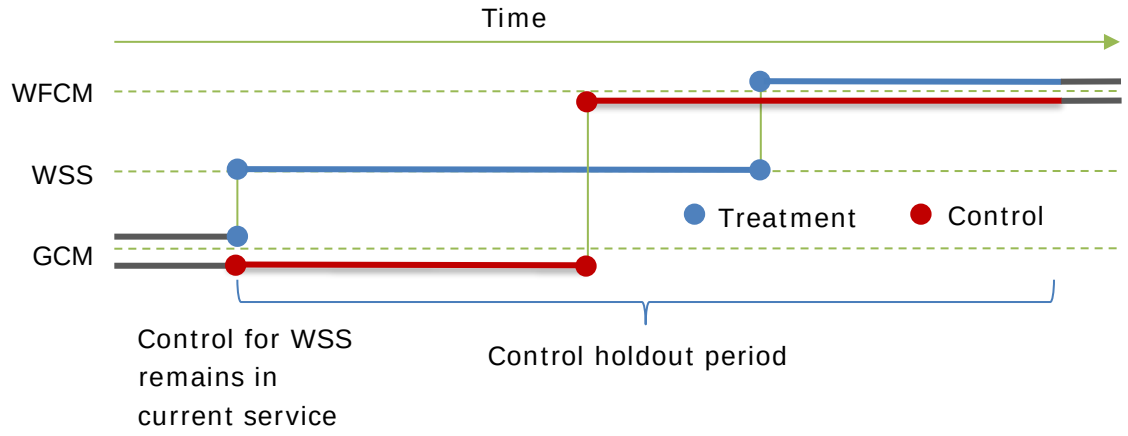
2.2 Of the 11 eligible people selected, one was randomly selected to be in the control group for the service, while the remaining 10 were assigned to the service and became the treatment group.

The control group member remained in their current service. The condition for being in the control group was that they could not be assigned to the service for a set period (this ranged from one-to-three years) and is referred to as the holdout period in this report. The 10 to 1 ratio was not based on any power calculations but was a level that was thought could be tolerated. The control group holdout period was based on an assessment of how long it would take for services to show an impact on participant's outcomes.

The challenge for the evaluation was that while a control group member was prevented from being assigned to the same service as the treatment group, they could still be assigned to other services. Therefore, the average experience of the control group was a mix of different case management services, including services that were as, or even more, intensive than the treatment group. Likewise, people in the treatment group were not restricted to remaining in the service they were assigned to and could be assigned to more intensive case management services over time.

To illustrate these issues, Figure 2 shows a stylised example of the experience of two people assigned to a control and treatment for WSS. At assignment both are in GCM and the treatment person is assigned to WSS while the control remains in GCM. During the holdout period, the control cannot be assigned to WSS; however, they can be assigned to other services (in this case WFCM). Similarly, the treatment is not prevented from being assigned to other services, so that before the end of the control hold period both are eventually assigned to WFCM.

Figure 2: A stylised example of the relative service assignment between a control and treatment group member for WSS



To be able to compare the cost-effectiveness of all 12 case management services against a common baseline, namely GCM, we took the following approach.

Estimate the impact of service assignment on welfare outcomes

For each service, we estimated the impact of assigned service on the:

- duration to exit main benefit
- duration to being assigned to another service
- income support payment rate while on main benefit
- case management and related costs when assigned to different services
- if a person exited from main benefit, the duration off main benefit
- income support payment rate while off main benefit.

In these models, we included the assigned service status of individuals at the start of each period as well as information on the eligibility criteria used for each service assignment. We did not explicitly include the treatment control assignment status into the models. Instead, the creation of the control group was to provide sufficient common support for each service being evaluated (ie equivalent individuals not assigned to the

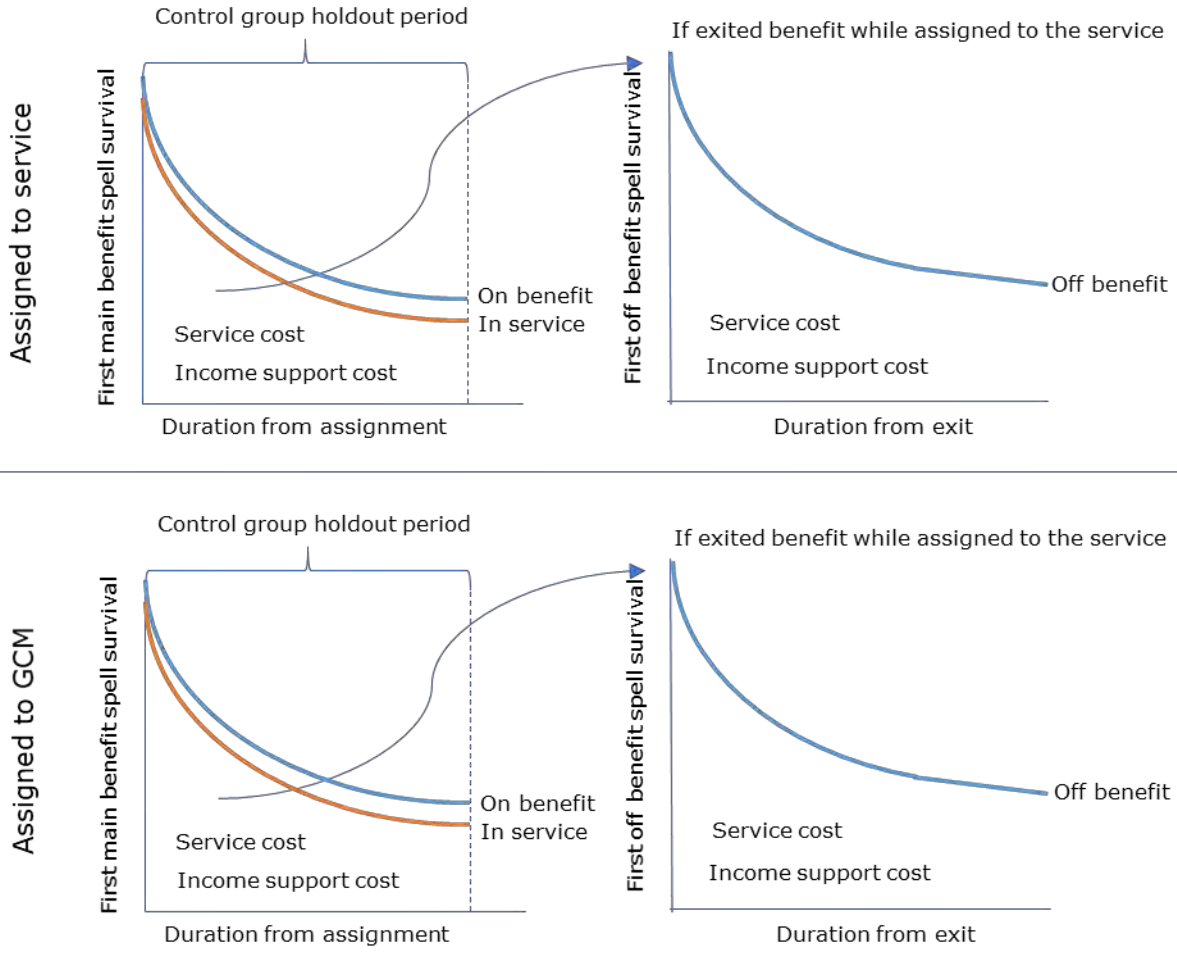
service). Because service allocation was deterministic based on administrative information, we do not have to be as concerned with unobserved selection effects as would be the case for interventions where referral decisions are made by staff and participants.

Simulate the difference in outcomes between a service and GCM

From the resulting model estimates, we ran Monte Carlo simulations based on two scenarios: (i) being assigned to the evaluated service, (ii) being assigned to GCM instead. To illustrate how the simulation process worked, Figure 3 provides a stylised illustration of how we estimated the cost and benefit of being assigned to a service relative to GCM.

From Figure 3 we track service participants from when they are initially assigned to the service. Here we estimate their duration to exit from main benefit and the duration before they are assigned to another service. The latter reflects instances where people are still on their first main benefit spell but are assigned out of the service. Once assigned out of the service, then these records are treated as right-censored (ie any subsequent costs or outcomes are excluded from that simulation run).

Figure 3: Simulation of the cost-effectiveness of each case management service to the GCM baseline



While on main benefit and in the service, we estimated the level of income support payments people receive and the level of case management and employment assistance

expenditure (service cost in Figure 3). The on benefit duration model was limited to the control group holdout period.

For those who exited benefit while assigned to the service we undertook a separate estimate of how long they would remain off main benefit. While off main benefit, we estimated the income support and any case management and employment assistance costs.

In each simulation we ran two scenarios, the first assuming the treatment group were assigned to the service, while the second assumed they had been assigned to GCM instead. The difference in experience (case management EA costs and income support payments) was the estimated impact of being assigned to the service compared with being in GCM.

Calculating the return on investment

Based on the simulation results, the return on investment was calculated as follows. From Table 1, the investment is the gross cost of the service. Gross cost is made up of service costs while on main benefit and in the service up to the end of the control group holdout period and any service costs of those who exited benefit while on the service. The return was the income support costs while in the service, minus the equivalent income support costs under the GCM scenario. The Return also includes the service cost under the GCM scenario. In this report, the ROI was calculated as a ratio of the Return divided by the Investment so that a value above one indicates the returns exceed the gross investment cost.

Table 1: Return on Investment cost-benefit calculation

Scenario	Investment	Return	
Service	Service cost	Income Support costs	
GCM		Income Support costs	Service Cost

The reason for using gross investment rather than net investment was to ensure the ROI had an ascending scale. In other words, the higher the ROI ratio the better the performance of the intervention is. It is possible for a service to have a lower overall cost than GCM. In this scenario, a net-investment/net-return ratio would be negative. While symmetrical and valid, a net-investment ratio is more difficult to explain to a non-technical audience, especially if a negative ROI can represent either a poor result (ie a negative net-return) or a positive result (ie a negative net-investment).

Table 2 illustrates the above points with a stylised example. The first row is the case where the service has both positive net-investment and net return. From a net-benefit (ie benefits minus costs) of $1,700, we have a gross ROI of 2.70 compared to a net ROI of 6.67. The second and third rows illustrate two divergent instances that have the same net-ROI, while net-benefit and gross ROI clearly show the difference between the two scenarios. It is the ambiguity of the net-ROI that is the reason we did not use it in this analysis.

Table 2: Alternative ways to calculate ROI net-investment compared with gross investment

Gross Investment	Counterfactual investment	Net Investment	Net return	Net-return plus counterfactual investment	Net-Benefit	ROI	
						Gross investment	Net-invest-ment
$1,000	$700	$300	$2,000	$2,700	$1,700	2.70	6.67
$700	$1,000	-$300	$2,000	$3,000	$2,300	4.29	-6.67
$1,000	$700	$300	-$2,000	-$1,300	-$2,300	-1.30	-6.67

Service Delivery Model

This section describes the operation of the Service Delivery Model in its pilot and national forms. We do not go into detail on the operation of the services within the SDM, readers interested in these aspects can refer to previous evaluations (Bandookwala, Kemp, Anderson, & Bly, 2014; Crane & Kemp, 2013).

Service Delivery Model pilot (October 2012 to July 2013)

In October 2012, 24 selected sites commenced the pilot of the New Service Delivery Model (NSDM). The NSDM involved targeting various levels of case management based on people's circumstances. In the pilot sites, case management was split into three services:

- Work Focused Case Management (WFCM pilot): one-to-one case management, with caseloads capped at 108 participants[3] for every case manager. WFCM case managers provided both income support and employment assistance.
- 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, Work Search Support involved a structured sequence of job search seminars. In addition to job search assistance, Work Search Support participants also participated in employment programmes and services. Participants on main benefit for less than seven weeks were left to manage their own job search.
- General Case Management (GCM): all remaining people on income support were assigned to GCM. For GCM there was no caseload restriction. As a result, the caseloads were much higher in GCM than for WFCM and Work Search Support. GCM case managers provided both income support and employment assistance.

Pilot sites

The selection of sites into the NSDM pilot was based on two pragmatic criteria. The first was to have sites of sufficient size to have at least two WFCM and Work Search Support case managers in each site. The second was that each of the 12 MSD operational regions in New Zealand had at least one pilot site. Since the site selection was not based on ensuring a representative sample of all sites we cannot be sure that the results of the pilot would necessarily be the same as for non-pilot sites. Table 3 shows the 24 sites selected for the pilot period, with the average number of participants in each service and the caseload for each site. Of interest is the large variation in caseload for GCM between sites. In certain sites, we found the GCM caseload was lower than for Work Search Support.

[3] The 108 number excludes people defined as partners for income support assistance. Partners were assigned to the same service as their partner who was defined as the primary beneficiary. This can result in the number of adults on a case manager's caseload exceeding the stated caps.

Table 3: NSDM pilot sites with an average of client and caseloads in Work Search Support, WFCM and GCM for each site

Site	Participants			Caseload size		
	WFCM	WSS	GCM	WFCM	WSS	GCM
Avondale	344	649	2,949	115	216	227
Dinsdale	331	616	3,556	110	205	187
Dunedin Central	766	1,156	4,595	109	165	306
Five Cross Roads	333	617	3,673	111	206	184
Glenview	117	213	1,728	117	213	192
Hamilton	126	490	2,454	126	245	245
Hamilton East	111	414	2,136	111	207	194
Hastings	400	654	4,053	100	164	289
Mangere	678	1,249	7,047	113	208	294
Manurewa	793	1,263	6,153	113	210	293
Naenae	334	657	3,307	111	219	301
Nelson	220	623	2,301	110	208	192
New Brighton	317	463	2,718	106	154	340
New Plymouth	213	427	3,812	106	213	224
Papakura	548	1,142	4,961	110	228	276
Porirua	561	1,146	4,842	112	229	269
Riccarton	573	535	3,357	96	178	305
Rotorua	883	1,467	6,562	110	210	328
Tamaki	454	683	3,093	114	171	281
Timaru	219	453	3,110	109	151	239
Wairarapa	334	614	2,960	111	205	296
Waitakere	777	1,505	8,269	111	215	276
Whangarei Central	789	1,112	4,470	113	185	406
Total	10,219	18,147	92,109	110	199	266

Participants include both main beneficiary and partners. For this reason, the average number of participants in the table can be higher than the maximum caseload size stated in the text. For example, WFCM has a maximum caseload of 108 primary beneficiaries, but when partners are included, the total caseload can exceed this cap.

National roll out of SDM (July 2013 onward)

From July 2013, MSD rolled out the NSDM to all sites and re-named it the Service Delivery Model (SDM). The SDM differed from the NSDM in several respects, including the number of services offered, the caseloads within each service and the eligible population.

- Work Focused Case Management General (WFCM General): the caseload cap increased from 108 to 121 participants (excluding partners) for each case manager. WFCM General participants were mainly made up of sole parents, jobseekers and a small proportion of jobseekers with a health condition or disability (with part-time work obligations).

- WFCM Health: caseload of 100 participants who were exclusively jobseekers with a health condition or disability and had deferred work obligations (predominantly work preparation obligations).
- WFCM Integrated Support: integrated service 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 clients 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.
- WFCM Integrated Support (ex-YP/YPP): people who came onto a main benefit from Youth Payment or Young Parent Payment were prioritised for the WFCM Integrated Support service but may have received the WFCM General service in some cases (ie if WFCM IS was not available in their area).
- Work Search Support: involved a structured sequence of job search seminars with participants. Caseloads were capped at 217.
- General Case Management (GCM): all remaining people receiving income support in the site were assigned to GCM. For GCM there was no caseload restriction. As a result, the caseloads were much higher than for other services. GCM case managers provided both income support and employment assistance. However, because of the high caseloads, case managers had only a limited amount of time to work with individual clients.

Investment Approach trials

In addition to the above services, MSD also trialled several service designs that are incorporated into the SDM. Because they are part of the SDM we also included them in the analysis.

- Sole Parent Employment Service trial: The aim of the Sole Parent Service was to provide employment support to people on Jobseeker Support who had sole parent responsibilities and were returning to full-time work. Case management providers were contracted to achieve this aim through employment-related case management and assistance in overcoming barriers to full-time work, including employment placement and post-placement support. Providers were to tailor their case management activities to the individual needs of participants. Provider payments were linked to the outcomes achieved by participants (MSD, 2016).
- Mental Health Employment Service trial: The aim of the Mental Health Service trial was to support people with common mental health conditions to gain work and achieve sustainable employment. Contracted case management providers were to provide employment-related case management, placement and post-placement support, integrated with the participant's clinical support. The target group for the service was jobseekers registered with Work and Income who were willing to undertake full-time employment but were limited in their capacity to look for or be available to work because of common mental health issues such as anxiety, stress or depression. Similar to the Sole Parent Service, providers were to tailor their case management activities to the individual needs of clients. Provider payments were linked to the outcomes achieved by participants (MSD, 2016).
- WFCM Intensive Client Support trial: The WFCM Intensive Client Support (ICS) trial focused on improving outcomes for people who first entered the welfare system

as teenagers, and who were currently aged between 18 and 39. The trial was developed in response to welfare valuations that identified this group as having a significant share of the welfare liability and facing multiple barriers to employment. The trial started in March 2015 and was scheduled to run for three years. A key feature of the ICS trial was the use of specialised case managers (ICSMs) to provide a more intensive service than traditional case management, via a 1:40 caseload (Hall, Herdina, & Henshaw, 2016).

Allocation of clients to services

The evaluation team was responsible for the allocation of people receiving income support to each service based on eligibility and priority rules developed in conjunction with our client-facing operational team. There were no unobserved selection effects as people could not select themselves into any service, likewise, case managers had no direct control over what service individuals were allocated to other than formally lodging recommendations into the administrative systems.

Service allocation was primarily based on observed characteristics of individuals, such as what type of main benefit they were on, their age, their statistical risk of remaining on benefit long-term and so on. There was one exception, where case managers could formally nominate people for a Work Focused Case Management Integrated. Here we still observed which individuals were nominated and subsequently allocated to the service on this basis. In this analysis, these individuals were excluded from the analysis.

Service Delivery Model

The MSD evaluation team controlled the allocation of individuals to each service. Each week the Service Delivery Model allocation process would assess the status and eligibility of individuals. Based on available places in each service by site and group, people would be assigned based on a pre-determined set of eligibility and priority rules. The Service Delivery Model allocation process also managed the randomisation of people into control groups for each service.

The output from the Service Delivery Model allocation process included:

- people's spells in each service
- profile of individuals when assigned to the service or the control group
- the weighting of the control group (based on the probability of being selected into the control group from the potential participant pool).

From the allocation datasets, we could identify when people moved between services, the exact client characteristics used in the allocation of people to each service, and the correct weight for the control group members for the analysis.

Eligibility rules for the SDM national programme

Allocation of individuals between services

The national rollout of the SDM increased the complexity of the allocation of clients to services. To manage this complexity, the allocation rules first ranked people by order of priority.

The priority order was:

1. IA Trial: Mental Health Employment Trial
2. IA Trial: Sole Parent Employment Trial
3. WFCM Intensive Client Support trial
4. WFCM Integrated Support / WFCM General Ex-YP/YPP
5. WFCM Health
6. WFCM General
7. Work Search Support
8. GCM.

The allocation process was designed on the principle that people moved up the priority order of services, but not down (other than through exiting the SDM entirely). Conversely, when the allocation process was run, it began from the highest priority service to the lowest (with people assigned to higher priority services being ineligible for lower priority services). This ordering was designed to ensure the caseloads of higher priority services were filled first and any under-capacity allocations would occur for lower priority services.

Allocation of individuals within a service

Before outlining the allocation process, readers need to understand how the capacity of each service was defined. The overriding objective of the allocation process was to ensure that the number of participants assigned to each service at a site did not exceed the capacity of the site. In other words, the goal was to maintain caseloads at or below the set maximums for the service in each site. A secondary objective was to have a mix of people in each service. To achieve this goal, the capacity for a service was partitioned into blocks for different groups (eg Jobseeker or Sole Parent Support). Finally, with each site/client block, the allocation process assigned a priority ranking based on a set of characteristics. The intention was that higher priority individuals would be more likely to be selected for the service.

To summarise, the service allocation process involved the following concepts.

- Service: the specific case management service or trial that eligible individuals were assigned to.
- Site: a single or group of sites that had a set capacity for a service. For example, a site might have had three WFCM General case managers and, therefore, would have had a WFCM General capacity of 363 places.

- Block: a site's service capacity was partitioned into blocks that were reserved for specific groups. In the case of WFCM General, a site may have had 30 percent of its capacity for Jobseekers, 60 percent for part-time work tested sole parents and 10 percent Jobseekers with health conditions, injury or disability without a deferral to work obligations.
- Priority rank: eligible clients were given a priority rank based on rules. These rules vary according to service and are outlined for each service in Appendix 1.
- Eligible candidates: people who are selected to be assigned to either the treatment group or the control group.

To allocate people to each service, the following steps were followed:

- Identify eligible individuals: for each service site and block identify all individuals eligible for the service.
- Remove ineligible participants from the service: from the people currently in the service, remove any that are no longer eligible to be on the service.
- Determine spare capacity: calculate the number of spare places available for each site, both overall and at each client block.
- Determine if there were enough capacity and eligible individuals: for the allocation of the control and participant groups there needed to be sufficient places available as well as enough eligible individuals in each site and block to allocate at least one person to the control group. For most sites, this was set at a 10 to one ratio.[4] Therefore, for an allocation to proceed there needed to be at least 10 spare places in a block and at least 11 people who were eligible. For smaller sites, where this allocation design at individual site level was not feasible, the random allocation into control and treatment group occurred within blocks based only on Selection date and Group with no stratification by site. See Control Group Allocation page 27 for more detail.
- Selection of eligible candidates: allocation to a service occurred if two conditions were fulfilled. The first was that there were sufficient places available in a site block and the second was that there were enough eligible people to have a treatment and control group assignment. If these conditions were not met, then the allocation process would not occur for that site/block and the available spaces for the service would remain unfilled until the following week.

  It is important to stress here, that people were not yet assigned to treatment or control at this point. Instead, this stage was only concerned with selecting enough eligible candidates to allocate to the control and treatment groups.

  The selection of eligible candidates was based on a weighted random sample. The weighting was based on each person's priority rank, so that higher priority individuals were more likely to be selected into the eligible candidate group than

[4] This ratio was not based on any power calculations, but was agreed as providing a reasonable sized control group to identify meaningful impacts on the time individuals spent off main benefit under different services. For the main services (WSS and WFCM General) the absolute numbers (eg in the tens of thousands of participants each year) would ensure there would be sufficient absolute number of controls selected at a 10 to 1 ratio. For smaller services and the IA trials, the ratio was set at 2:1.

lower priority people were. The motivation of a weighted random sample was to ensure that each service had at least some low priority individuals allocated to the service. In doing so, we could test whether the prioritisation rules either increased or decreased the overall effectiveness of the service.[5]

- Assignment to the control or treatment group: after selecting eligible candidates, a set number were randomly allocated to the control group. For the WFCM and Work Search Support services, the ratio was usually 10 participants and 1 control. The people assigned to the control group remain in the service they are currently in. From an evaluation perspective, it would have been ideal to assign controls to GCM to enable a direct comparison of service assignment to being assigned to GCM. However, it was considered this imposed an excessive cost for the control group.

  Note that before August 2014 a coding error was made that meant the randomisation variable for the priority selection (see Selection of eligible candidates above) was reused for the selection of the control group. As a result of this re-use of the randomisation variable, the control group differed in profile from the comparison group. We outline on page 30 how we addressed this error in the current analysis. From 27 July 2014, the coding error was fixed with a new random variable created for the allocation of eligible candidates into the treatment and control group.

- Update service history: the service history for individuals was updated, removing them from the eligible group for lower priority services.

Exiting from a service (NSDM and SDM)

Within the SDM/NSDM, participants could exit from a service in the following ways:

- exited from main benefit[6]
- transferred to a more intensive case management service (from GCM to Work Search Support to WFCM)[7]
- selected for a higher priority trial involving case management services (SDM only)
- their circumstances changed, making them ineligible for a service
- for NSDM and services operating only in selected sites (ie WFCM Integrated Support, WFCM Health, WFCM pilot and Work Search Support pilot) participants could exit by moving to a non-participating site.

There were no formal voluntary exits from the in-house case management service by case managers or participants.

People exiting from services may introduce unobserved selection effects. For example, people may respond to more intensive case management by either seeking a work test

[5] If we had not included a random weighting and instead taken a simple rank from highest to lowest priority individuals, then the number of low priority participants in a given service could be too low to evaluate the impact of the service for these groups.

[6] If a person returns to benefit within two weeks and is still eligible, then they are returned to the service.

[7] Note this does not apply to the controls for a service. Here, being a member of a service's control group makes the individual ineligible to be assigned to the service. The duration in the control group ranged from one year for WSS through to three years for WFCM-related services.

exemption or applying for a benefit that makes them ineligible for a service. We explicitly control for this possibility in the evaluation using assigned service spells (refer to page 19).

Service participants excluded from the evaluation

We excluded a small number of participants in the SDM from the evaluation. These participants were selected onto services in such a way that we could not estimate the difference the service made to their outcomes.

- Nominated WFCM Integrated Support participants: case managers could nominate people for WFCM Integrated Support. Nominating people introduces unobserved selection effects that we cannot account for in the current analysis. We could have included nominated clients by randomising the allocation of nominated individuals into WFCM Integrated Support or into the control group for WFCM Integrated Support (ie they would have remained in their current service). However, MSD decided not to take up this option, and chose instead, to allocate all nominated clients to WFCM Integrated Support.
- WFCM Low Trust Clients: Low Trust Clients (LTC) were people with a record of benefit fraud or frequent overpayments. From October 2013, MSD required all LTC to be in the WFCM General service. The aim was to minimise the opportunity for fraud or overpayment through regular face-to-face contact with a designated case manager. The allocation of LTC was automatic and was not included in the RCT process for service assignment.

Alongside these people, there were a number of participants that were treated differently from the general population of participants.

Ex-YP/YPP clients

People who had been on YP or YPP benefit after December 2013 were always assigned to a one-to-one case management service. For sites where WFCM Integrated Support was available, people who had received YP or YPP were randomly assigned to either WFCM Integrated Support or to WFCM General. Ex-YPP/YPP who were in sites with no WFCM Integrated Support service were assigned to WFCM General as a matter of course. These mandatorily assigned participants were outside of the evaluation.

Table 4 summarises the number of starts within each service over the reporting period (each start represents a unique participant, but individual participants can be included in more than one service). Of these starts, we identify the proportion of people who were excluded from the evaluation. Relative to the total number of service participants, WFCM Integrated Support-nominated participants represented the largest group of participants outside of the evaluation. By comparison, the other excluded groups made up a relatively small proportion of participants for each service.

Table 4: Number of service starts and non-evaluated client groups

Service	Period	Total starts	ex YP non IS	LTC group	SLP opt-in	WFCM Integrated Support (Nominated)
WFCM General	30JUN2013-03SEP2017	292,668	1.0%	1.9%	-	-
WFCM Health	30JUN2013-03SEP2017	58,153	-	0.3%	0.8%	-
WFCM Intensive Client Support	22MAR2015-03SEP2017	1,076	-	0%	-	-
WFCM Integrated Support	30JUN2013-03SEP2017	21,903	-	0.2%	-	15.6%
WSS	30JUN2013-03SEP2017	207,416		0%	-	-

Total starts by service count individual participants only once (ie we ignore subsequent starts on the same service). People can be in more than one service over the coverage period.

Control group

Central to the evaluation of intensive case management services was the creation of a control group for each service. In this section, we discuss what function the control group played and the implications of the design of the control group for the evaluation.

Purpose of the control group

Readers will recall from the outline of the service allocation process in the previous section (page 19) that:

- service participation was compulsory
- selection into the service was based on a combination of observed characteristics and randomisation
- updates to service eligibility, service status and allocation occurred on a weekly basis
- the selection process was centrally controlled by the evaluation team.

For these reasons, we are not concerned about unobserved selection bias in our estimates of the impact of being allocated to each service (selection bias through service exits was dealt with through the creation of assigned service spells, which will be discussed in the following section). However, because the services involved a sizeable proportion of the eligible population we ran the risk of having no suitable comparison (referred to as common support). In other words, if all people eligible also received the service, then we have no point of comparison. For this evaluation, the key role of the control group was to ensure that enough of the eligible population received some alternative set of services. In so doing, we could compare the outcomes for a similar group of people receiving a range of alternative services.

Ethical considerations

In developing the control group design, we looked at the ethics of imposing a control group for the evaluation within the Service Delivery Model. Here we set out the reasoning for having a control group over alternatives.

The net social benefit of SDM was uncertain

The first consideration was whether we are justified in evaluating services known to have net social benefit. If we already know that – for most service participants – the social benefits of the service outweigh the social costs, there is no value in maintaining a control group. In the case of the SDM, we had some evidence that the service was cost-effective (from a service delivery perspective) for Work Search Support and was likely to be cost-effective for WFCM (MSD, 2014).

However, we did not consider this was sufficient evidence to assume that the SDM would have the same impacts. We give the following reasons for this conclusion:

- The pilot analysis indicated these services were ineffective for some groups included in the SDM.
- Some newer SDM services differed in some respects from those in the pilot sites (ie WFCM for people with a health condition or disability and WFCM Integrated Service).
- The SDM included groups not included in the pilot of WFCM and Work Search Support.

- The pilots were a non-representative and non-random sample of MSD sites and for this reason, the results observed during the pilot could not be generalised to the whole country.
- The effects of social services are sensitive to changes in context. For example, service impacts can change in response to social or economic factors as well as through regulatory and changes in other service settings. Therefore, we cannot assume that results from earlier studies will necessarily continue.[8]

A further point to consider here is that while the pilot services showed increased time off main benefit for their participants, we do not currently know the broader social benefits of this type of intervention. Specifically, time off main benefit is being used as a proxy for increased time in employment. It is increased time in employment that would provide evidence that the intervention provides benefits to participants. We do not address this critique of the evaluation in the current analysis. But, we do plan to extend the evaluation to include employment outcomes in future work.

SDM is compulsory

The compulsory nature of SDM services greatly increased the need to know whether these services have a net social benefit. Under current law, people with work obligations are required to participate in SDM services with the intention to help them move into employment. It is therefore incumbent on government to show that these services do improve participant's chances of gaining employment, since people have no choice in the decision to participate other than to leave income support.

Note that while assignment to services is compulsory, once a person is in a service their obligations to undertake any activities is based on what main benefit they are on and if they have work exemptions. For example, people receiving Jobseeker Support Health Condition or Disability and are assigned to WFCM Health have the same obligations as those assigned to GCM.

No feasible robust alternatives

We examined alternatives to randomised control trial designs. The main challenge for this evaluation was ensuring similar people received distinct types of case management service. Because of the size of the services within SDM, it was very likely that all eligible people, especially for those in the high priority groups, would be assigned to a service. In such cases, it would not be possible to reliably estimate the impact of the service, since there are no comparable individuals who did not receive the service to compare against. The alternative, non-experimental designs were either not feasible or lacked robustness. For example, while a matched comparison group design would be robust; such a method is not feasible since there would be very few non-participants to select a suitable comparison from. On the other hand, pre-post designs would not be sufficiently robust given the confounding effects of economic and policy differences before and after the introduction of the SDM.

[8] By way of contrast, we normally accept the continued relevance of experimental evidence for pharmaceutical drugs such as statins (ie it is safe to assume the human physiology will respond in the same way over many generations). However, even in the biomedical field we consider context when assessing effectiveness. For example, we have seen a reduced effectiveness of antibiotics through the emergence of resistant bacteria (ie the combination of selective pressure and rapid regeneration of bacteria means the evolution of drug resistant bacteria is highly relevant to the ongoing effectiveness of antibacterial drugs).

Minimising the potential cost of control group membership

Having decided that a control group design would provide the most robust results possible, we also considered how to minimise the cost of being assigned to a control group.

Restricted from one service only

When a person was assigned to the control group for a service, they were only prevented from receiving that service. At assignment to the control group, the individual remained in their current service and could be assigned to services that were more intensive if they were eligible. Therefore, a person in GCM assigned to the control group for Work Search Support could still be assigned to WFCM General if they were eligible. One consequence of this design was that control and treatment group members may end up in similar services over time. This aspect of the control group greatly complicated our analysis of the effectiveness of intensive case management services. We outline in the impact estimation section below on how we dealt with this issue.

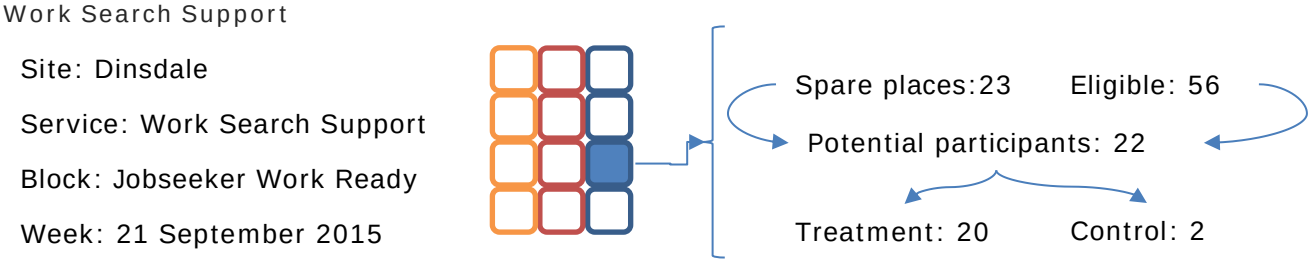
Control group duration

Control group status lasts for a defined period only, call the holdout period. For Work Search Support, people remained in the control group for 52 weeks, after which they can be assigned to the service they are in the control group for (if still eligible). For WFCM services, we extended the control group period to three years. The reason for the extension was to allow more time for WFCM services to have an impact on clients facing greater barriers to employment.

Control group allocation

The protocol for allocating clients into the control and treatment groups varies across services as well as sites. For most services, we allocated clients in a block design, based on selection date, selection site and selection group.

Figure 4: A stylised example of assignment of people to the control and treatment group for Work Search Support



For each service, selection date, selection site and selection group, the allocation protocol involved the following steps:

- Spare places: identify if the block had at least 10 places available.
- Eligible candidates: identify if there were at least 11 people eligible for the block.

If these conditions were met, then the allocation procedure randomly selected 11 eligible people for the block. Figure 4 shows a stylised example, here for the Work Search Support service in Dinsdale. In the Jobseeker Support Work Ready block there are 23 places available and 56 people are eligible. Under these conditions, 22 people are randomly selected from the 56 eligible. Of the 22 selected, 20 are randomly allocated to the treatment and are assigned to Work Search Support. The two people allocated to the control group remain in their current service. The 34 eligible people not selected into the treatment or control group remain in their current service and can either be assigned to another service in the current allocation run or be selected for WSS in subsequent weeks.

If the eligible and spare space criteria are not met (eg less than 10 or a remainder after a multiple of 10) then the allocation process would not fill these open places. Instead, the allocation process left these places unfilled until the next allocation date (normally run weekly apart from the Christmas/New Year break).

- The randomisation of eligible candidates: once people were selected into the eligible candidate group they are given a random number and ordered from lowest to highest random value.
- Selection of the control group: based on the number of available places (in multiples of 10), the top 10 eligible candidates were selected into the participant group, whilst those with the highest random number were assigned to the control group. For example, if there were 20 places to fill then the first 20 of the 22 randomly ordered people were selected into the participant group (treatment), with the last two assigned to the control group.

Aggregation of sites

The implication of the randomisation process for the operation of the Service Delivery Model was that caseloads in each service block needed to fall by at least 10 before new participants were allocated to the service. In addition, if a service had more than one client block, then it was possible that the caseload could fall to 10 places times the number of blocks before an allocation was made.

For the larger sites, the total caseload was sufficiently large relative to the number of open places waiting to be filled. However, for smaller sites, the requirement to have 10 places would noticeably reduce the allocated caseload in these sites.

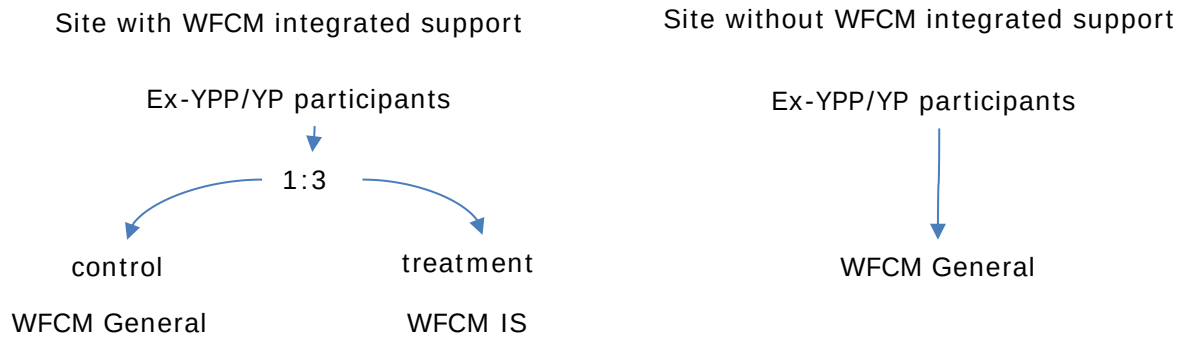
In addition, for WFCM Integrated Support, one case manager could manage participants from multiple sites. Here, the block design was across several sites rather than just a single site. However, these sites are geographically proximate, so we retained some power through the block design by having participant and controls in the same general geographic location, albeit with a wider spread than a single site block.

For the larger services (WFCM General and Work Search Support) all smaller sites were aggregated into a single site. At this point, we lost the advantage of geographical blocking, because within each selection date, site and client group, individual participants and controls can be from very different geographic locations.

WFCM Integrated Support (ex-YP/YPP)

People who had been on YP or YPP main benefits after December 2013 were required to be in an intensive one-to-one case management service. If possible, we assigned them to WFCM Integrated Support, and if not then we assigned them to WFCM General (Figure 5). Within the evaluation, we could randomly assign these people between WFCM General and WFCM Integrated Support, in those sites where WFCM Integrated Support was available.

Figure 5: Allocation of people who have been on YP or YPP to the WFCM Integrated Support control and treatment group



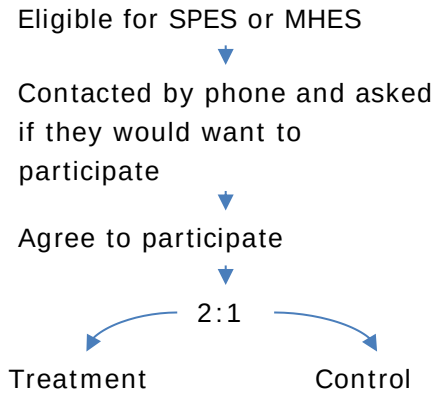
The randomisation protocol differed from the other SDM services outlined above. Instead of 10 places, the allocation of ex-YP/YPP to WFCM Integrated Support was based on multiples of four, where three people were assigned to WFCM Integrated Support (treatment) and one to WFCM General (control). In addition, and in a further departure from the previous protocol, if there were fewer than three places, the allocation procedure would allocate people based on the number of places (one or two) and have at least two people eligible. For example, if within a block, there were two places in WFCM Integrated Support and there were two eligible people, then one was randomly allocated to WFCM Integrated Support and the other to WFCM General. In those instances, where there was only one eligible person then they were assigned to WFCM General and fell outside of the RCT design.

This divergence from the main protocol reflected concerns over leaving people who had been on YP/YPP in GCM or Work Search Support for any period. If we had a 10 treatment to 1 control, the caseload would have needed to have at least 10 places available before allocation could occur. Until these spaces become available, eligible YP/YPP people would have remained in GCM or Work Search Support.

SPES and MHES trials

The two contracted-out case management trials were voluntary. The process for allocating clients to the control group differed from that used in the compulsory intensive case management service discussed above (see Figure 6). Eligible people where identified based on their characteristics on the administrative systems. Those eligible were contacted by phone and asked whether they would like to participate in the service. The call centre staff member recorded the contact with each person and whether they agreed to participate. From those who agreed to participate in the trial, one out of every two was randomly allocated to the control group and remained in their current case management service.

Figure 6: Allocation of eligible and voluntary participants to control and treatment groups for SPES and MHES trials



Control and treatment groups

Table 5 summarises the calendar period the RCT operated over for each service and the total number of individuals assigned to either the control or treatment group.

Table 5: Control group period and number of controls and treatments by service

Service	Start	END	Treatment	Controls
WFCM General	June 2013	July 2016	300,627	30,266
WFCM Health	June 2013	July 2017	54,720	5,490
WFCM Integrated	June 2013	July 2017	12,887	1,319
WFCM Integrated (Youth)	March 2014	July 2017	4,351	2,576
Work Search Support	June 2013	July 2016	188,675	19,009
Mental Health Service	September 2013	June 2014	1,785	880
Sole Parent Service	September 2013	May 2017	2,273	1,121
WFCM Intensive (Early)	March 2015	July 2017	370	454
WFCM Intensive (Entrenched)	March 2015	July 2017	564	720
WFCM pilot	September 2012	June 2013	21,199	4,854
Work Search Support (pilot)	September 2012	June 2013	42,636	5,512

Control group re-weighting

In the implementation of the SDM allocation process, there was an error in the allocation of eligible participants into control and treatment groups.

Randomisation error

The error involved the re-use of a randomisation variable for both the selection of people into the potential candidate pool and to then split this pool into the treatment and control groups. We identified this error in July 2014 and corrected for all allocations after 27 July 2014. Note that this error did not affect the pilot version of the NSDM. However, the bulk of clients in national SDM analysis would have been selected before 27 July 2014 and are affected by this error.

The re-use of the randomisation variable meant the profiles of the treatment and control groups differed based on the priority rules. Here the participants were more likely to be in the high priority groups whilst the controls were more likely to be in the lower priority groups. The reason this occurred was that the selection into the potential pool of participants involved a combination of randomisation and prioritisation.

The selection process first assigned a randomly drawn number to all eligible participants. This random value was then multiplied by priority rank (from 1, highest, to 9, lowest) for everyone. The selection of potential participants proceeded from the person with the lowest priority weighted random number to the highest. What this meant was that those with a higher priority were more likely to be selected into the potential pool of participants than those with a lower priority. The aim was to ensure that the potential participant group had a mix of low and high priority individuals. The other, unintended, consequence was that higher priority individuals selected into the potential participant pool had lower values of the random number than the lower priority individuals did.

The problem arose when the same random value was used to split the potential participant group into the participant and control group. Here the potential participants were ranked by the same random variable as used to select the potential pool of participants. From the randomly ranked list, participants were selected first, with those unselected assigned to the control group. Hence, the participants tended to have higher priority rankings on average than the control group.

Table 6 summarises the impact of the randomisation error on the distribution of participants by block and priority group. For each block the proportion of control and participants was the same, the difference occurs within each combination of block and priority group. Without the randomisation error, the proportion of control and treatment should be the same. However, this was not the case for several services. For example, for Work Search Support block priority group 1:8 contains 37.1 percent of participants, but 44.0 percent of the controls, a difference of 6.9 percentage points. Of all the services in the evaluation, the largest difference between control and treatment group occurred for WFCM Integrated Support, where the difference was 44.0 percentage points between the control and participant groups for each of the two priority groups.

Table 6: Differences in the allocation of participants and controls by client block and priority group

Service	Block: Priority	Participants	Control	Difference
WFCM General	1:1	20.2%	23.0%	*2.8 ppt
	1:2	8.4%	7.0%	*1.4 ppt
	1:3	2.2%	2.1%	0.1 ppt
	1:4	2.3%	2.5%	0.2 ppt
	1:6	1.7%	1.5%	0.2 ppt
	1:7	4.4%	4.9%	*0.4 ppt
	1:9	5.7%	3.7%	*1.9 ppt
	2:4	9.0%	9.4%	0.5 ppt
	2:9	0.7%	0.2%	*0.5 ppt
	3:1	5.1%	5.0%	0.1 ppt
	3:2	7.3%	10.2%	*2.9 ppt
	3:3	4.9%	4.9%	0.0 ppt
	3:6	8.8%	8.5%	0.3 ppt
	3:7	11.1%	9.6%	*1.5 ppt
	3:9	8.2%	7.2%	*1.1 ppt
WFCM Health	1:1	7.8%	10.3%	*2.5 ppt
	1:3	10.8%	10.8%	0.0 ppt
	1:4	8.5%	9.1%	0.6 ppt
	1:5	23.2%	20.8%	*2.5 ppt
	1:6	10.6%	10.9%	0.4 ppt
	1:7	6.4%	5.7%	0.6 ppt
	1:8	25.3%	25.0%	0.3 ppt
WFCM Integrated	1:4	30.7%	74.4%	*43.7 ppt
	1:9	62.6%	18.6%	*44.0 ppt
Work Search Support	1:1	16.7%	17.0%	0.3 ppt
	1:6	2.0%	0.9%	*1.1 ppt
	1:7	1.5%	0.7%	*0.8 ppt
	1:8	37.1%	44.0%	*6.9 ppt
	1:9	32.8%	27.4%	*5.4 ppt
	2:1	6.5%	7.1%	*0.5 ppt
	2:7	1.2%	1.1%	0.1 ppt
	2:8	2.2%	1.8%	*0.4 ppt

*: the difference is significant at the 5% confidence interval.

Implications for the evaluation

The key result of the randomisation error was a decrease in the common support between the participants and the control group. In other words, the control and participant group profiles differ in systematic ways according to the nature of the prioritisation rules. On the other hand, the randomisation error did not introduce any unobserved selection effects into the analysis since we retain all information used to allocate people to the treatment and control groups.

Addressing the randomisation error within the analysis

We took two steps to reduce the impact of the randomisation error in our analysis. The first was to re-weight the affected control group members to more closely match the profile of the participants (selected between 30 June 2013 and 20 July 2014). The second step was to include all variables used to prioritise potential participants in the regression models to account for any underlying differences between the participant and control group.

We re-weighted the control group through a generalised regression weighting algorithm called GREGWT developed by the Australian Bureau of Statistics. The purpose of the algorithm is to adjust selection weights for individual units so that their weighted values along specified dimensions match specified benchmarks. In our case, the benchmarks are the proportion of participants in each of the following dimensions:

- Selection date: the proportion of people selected for a service each week
- Site: the location that people are assigned to a service
- Client block priority rank: the people block and priority rank within each service caseload (see Table 6).

We look at the profile of the participants and controls at assignment in the following section. However, Table 7 compares the profile of participants and control by the client priority group levels before and after the adjustment to the weighting. Overall, we can see the adjusted weighting helps to reduce the differences in priority selection groups between the participant and control groups.

Implications of re-weighting for the evaluation analysis

The re-weighting of the control group decreased the power of the analysis. In effect, the re-weighting increased the weight of control group members where they were under-represented. To take a simple example, if control group members had an initial weight of 10, but they were in a block where they represented 5 percent of the participant group their weights needed to be increased to 20. Conversely, control group members had their weights reduced if they were in block/priority segments where controls were over-represented. The impact of this re-weighting was largest for WFCM Integrated Support and is reflected in the standard errors for the impact estimates for this service.

Table 7: Differences in the allocation of participants and controls by client block and priority group original and adjusted

		Original			Adjusted		
Service	Block: Priority	Participants	Control	Difference	Participants	Control	Difference
WFCM General	1:1	20.2%	23.0%	*2.8 ppt	23.0%	22.7%	0.3 ppt
	1:2	8.4%	7.0%	*1.4 ppt	7.0%	7.0%	0.1 ppt
	1:3	2.2%	2.1%	0.1 ppt	2.2%	2.2%	0.0 ppt
	1:4	2.3%	2.5%	0.2 ppt	2.7%	2.8%	0.2 ppt
	1:6	1.7%	1.5%	0.2 ppt	1.6%	1.5%	0.1 ppt
	1:7	4.4%	4.9%	*0.4 ppt	4.7%	4.7%	0.0 ppt
	1:9	5.7%	3.7%	*1.9 ppt	3.8%	3.9%	0.1 ppt
	2:4	9.0%	9.4%	0.5 ppt	9.3%	9.4%	0.1 ppt
	2:9	0.7%	0.2%	*0.5 ppt	0.2%	0.3%	0.0 ppt
	3:1	5.1%	5.0%	0.1 ppt	5.2%	5.3%	0.1 ppt
	3:2	7.3%	10.2%	*2.9 ppt	10.2%	10.3%	0.2 ppt
	3:3	4.9%	4.9%	0.0 ppt	5.0%	5.1%	0.1 ppt
	3:6	8.8%	8.5%	0.3 ppt	8.2%	8.2%	0.0 ppt
	3:7	11.1%	9.6%	*1.5 ppt	9.6%	9.5%	0.1 ppt
	3:9	8.2%	7.2%	*1.1 ppt	7.2%	7.2%	0.1 ppt
WFCM Health	1:1	7.8%	10.3%	*2.5 ppt	7.8%	10.1%	*2.4 ppt
	1:3	10.8%	10.8%	0.0 ppt	10.8%	10.6%	0.2 ppt
	1:4	8.5%	9.1%	0.6 ppt	8.5%	8.9%	0.4 ppt
	1:5	23.2%	20.8%	*2.5 ppt	23.2%	20.6%	*2.6 ppt
	1:6	10.6%	10.9%	0.4 ppt	10.6%	11.3%	0.8 ppt
	1:7	6.4%	5.7%	0.6 ppt	6.4%	5.6%	*0.8 ppt
	1:8	25.3%	25.0%	0.3 ppt	25.3%	25.4%	0.0 ppt
WFCM Integrated	1:4	30.7%	74.4%	*43.7 ppt	73.6%	74.4%	0.8 ppt
	1:9	62.6%	18.6%	*44.0 ppt	19.5%	18.6%	0.9 ppt
Work Search Support	1:1	16.7%	17.0%	0.3 ppt	17.0%	17.0%	0.0 ppt
	1:6	2.0%	0.9%	*1.1 ppt	0.9%	0.8%	0.1 ppt
	1:7	1.5%	0.7%	*0.8 ppt	0.8%	0.8%	0.0 ppt
	1:8	37.1%	44.0%	*6.9 ppt	44.2%	43.8%	0.4 ppt
	1:9	32.8%	27.4%	*5.4 ppt	27.1%	27.6%	0.5 ppt
	2:1	6.5%	7.1%	*0.5 ppt	7.0%	7.0%	0.0 ppt
	2:7	1.2%	1.1%	0.1 ppt	1.0%	1.1%	0.1 ppt
	2:8	2.2%	1.8%	*0.4 ppt	1.9%	1.9%	0.1 ppt

Table 8 shows the difference in the time off main benefit at 52 weeks after assignment to the control and treatment group. The table compares the average and standard deviation in time spent on benefit based on the original and the adjusted weights. As expected, the adjusted weighting altered the average outcomes of the control group for some of the services. The largest change occurred for WFCM Integrated Support with the average outcome for the control group decreasing by 10 days. The implication for the subsequent impact results would be that using the original weights would have resulted in larger effects than using the adjusted weights.

On the other hand, for Work Search Support the effect is reversed, with the average outcomes of the controls with adjusted weights being slightly higher than with the original weighting.

Table 8: Time off main benefit in the 52 weeks after assignment to the participant and control group in each service

Service	Treatment		Control	
	Original	Adjusted	Original	Adjusted
WFCM General	266.3 (123.8)	266.3 (123.8)	272.6 (378.1)	272.0 (382.3)
WFCM Health	294.0 (109.5)	294.0 (109.5)	295.3 (335.7)	295.3 (335.7)
WFCM Integrated	315.7 (95.3)	315.7 (95.3)	333.7 (240.4)	322.3 (271.6)
WFCM Integrated (Youth)	209.8 (124.0)	209.8 (124.0)	220.5 (372.2)	213.8 (421.4)
Work Search Support	222.6 (130.2)	222.6 (130.2)	232.7 (398.9)	235.0 (399.5)

Numbers in the parentheses represent the standard deviation.

The second observation is the increase in the standard error, whereby with the original weighting the standard error for the control group was smaller than when we use the adjusted weights. The services most affected were WFCM Integrated (240 to 271) and WFCM Integrated (Youth) (372 to 421).

Control group integrity

One critical issue with control groups is their integrity. In this context, we mean to what extent do control groups end up receiving the same service as the participant group? The closer the actual experience of the treatment and control groups are, the more difficult it becomes to identify the difference a service makes.

We can look at this problem at two levels. The first is ensuring that control group members are not assigned to the very service that they are the control group member for (eg a person in the control group for Work Search Support being assigned to Work Search Support). The second, and more important, is the extent to which frontline staff might override control group membership. We discuss each in turn.

Maintenance of control group service status

The first level of integrity is at the service allocation level. As noted previously, the evaluation team within MSD ran the process of allocating people to services. For this reason, we had good control over the assignment of people to services to avoid mistakenly allocating control group members to the wrong service.

Table 9 summarises the number of control group members in each service and we identify the proportion of control group members who have been assigned to their control service within the control group period. These control group members were excluded from the analysis.

Table 9: Control group violation by service

Service control group	Control group N	Violated N	Weeks after assignment
MHES Trial	880		
SPES Trial	1,126	12	15
WFCM GEN	30,361	603	43
WFCM HCD	5,491	23	36
WFCM ICS	2,354	12	87
WFCM IS	1,324	20	16
WFCM IS YS	2,576		
WFCM pilot	4,854		
WSS	19,011	93	31
WSS pilot	5,512		
Total	73,489	763	

Mandatory assignment to services

One source of control group violations involves the mandatory assignment of clients to specific services. Within the SDM population, there are two such groups:

- Low Trust Clients (see page 23 for more details). If a person is in a control group for a WFCM service and was subsequently identified as a Low Trust Client, then they are automatically assigned to a WFCM service.
- Ex-YP / YPP: people exiting YP or YPP benefits after December 2013 were also required to be in a WFCM service. This policy was implemented on 30 March 2014, with all ex-YP/YPP clients in WFCM control groups assigned to either WFCM Integrated Support or WFCM General.

Participants moving in and out of partner benefit status

Partners were not treated as individual participants; instead, they were treated as an attribute of the primary beneficiary.[9] Accordingly, within the service allocation process, partners were assigned to the same service as their primary. Therefore, if the primary was assigned to the control group, so was the partner. Later, if the couple split and the partner became a beneficiary in their own right, they were no longer identified in the allocation process as a control group member. For this reason, ex-partners can end up in the service that they were initially assigned to the control group as a partner.

[9] This practice arises from the assessment of income support eligibility based on family units, and the outdated view in the legislation that families have a single income earner. Whilst the legislation has been updated, with increased work obligations for partners, how MSD works with partners is taking longer to change.

Case management of control group members

The second area of control group violation is more difficult to identify. Here we are discussing the way that case managers treated people in the control group.

We did not tell staff who was in the control group, but it may be inferred

The first point to make is that control group membership was not made available to case managers. However, they may infer membership from which service a person is in. For example, if most people on Jobseeker Support are in either Work Search Support or WFCM, then any people on Jobseeker Support in GCM are likely to be in the control group. On this basis, case managers might have treated, such people differently from other people in GCM (ie give them more assistance). We believe this risk was greatest for control group members in GCM, control group members in other services would have been harder to identify.

Quantifying case manager effort and assistance

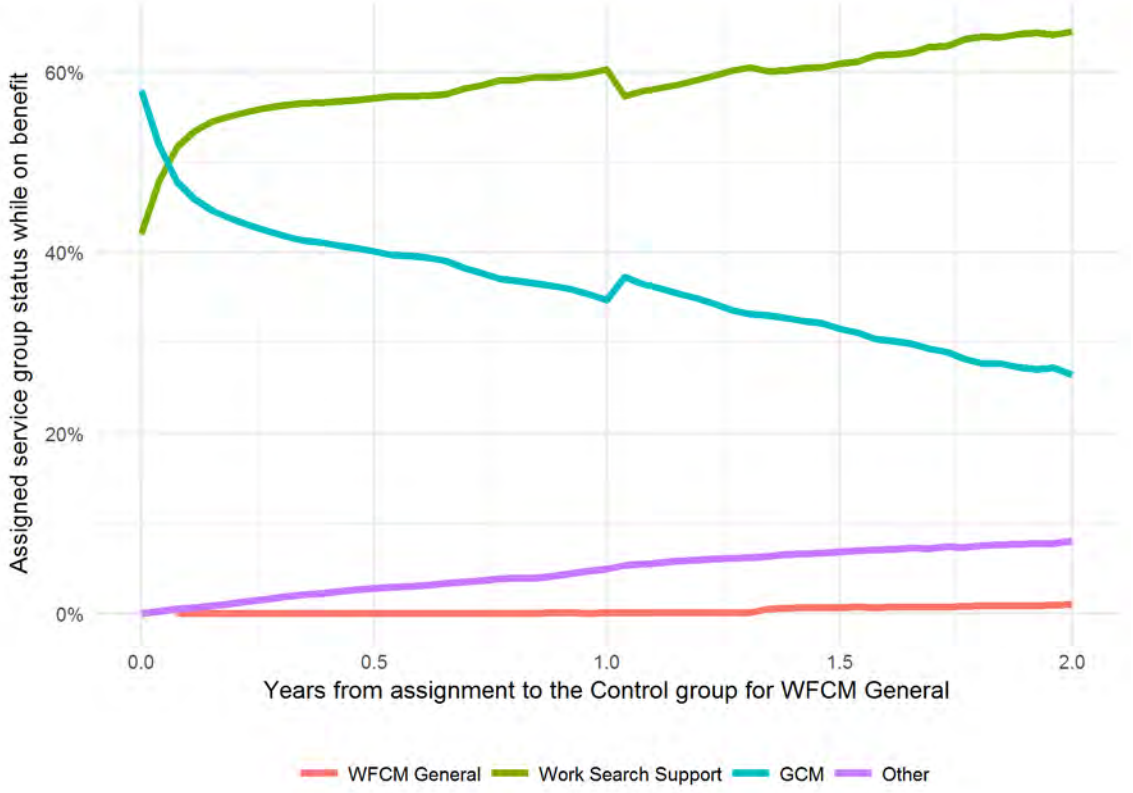
In the current evaluation, we explicitly looked at the amount of staff time spent with individuals. An important finding from this evaluation was that case managers did not follow the service allocations as expected. Instead, the primary source of variation in case manager time and investment was by benefit type, age group and risk of long-term benefit receipt. In other words, the level of assistance was highest for young jobseekers and those with low expected risk of long-term benefit receipt, irrespective of what service they were on.

Service status of control groups

In the previous section on assigned service spells, we focused largely on people assigned to a specific service. Here we examine the service status of people assigned to the control group for each service. Figure 7 tracks the assigned service status of people assigned to the control group for WFCM General over the duration of their first spell on main benefit.

The services received by the control group represent the expected service mix in the absence of WFCM General. From assignment, we see that most clients in the control group are either in Work Search Support or GCM. Over the first year after assignment, the proportion of clients in Work Search Support increases relative to GCM. This illustrates the challenge for the evaluation since we want to estimate the impact of being assigned to WSS compared with the baseline service of GCM.

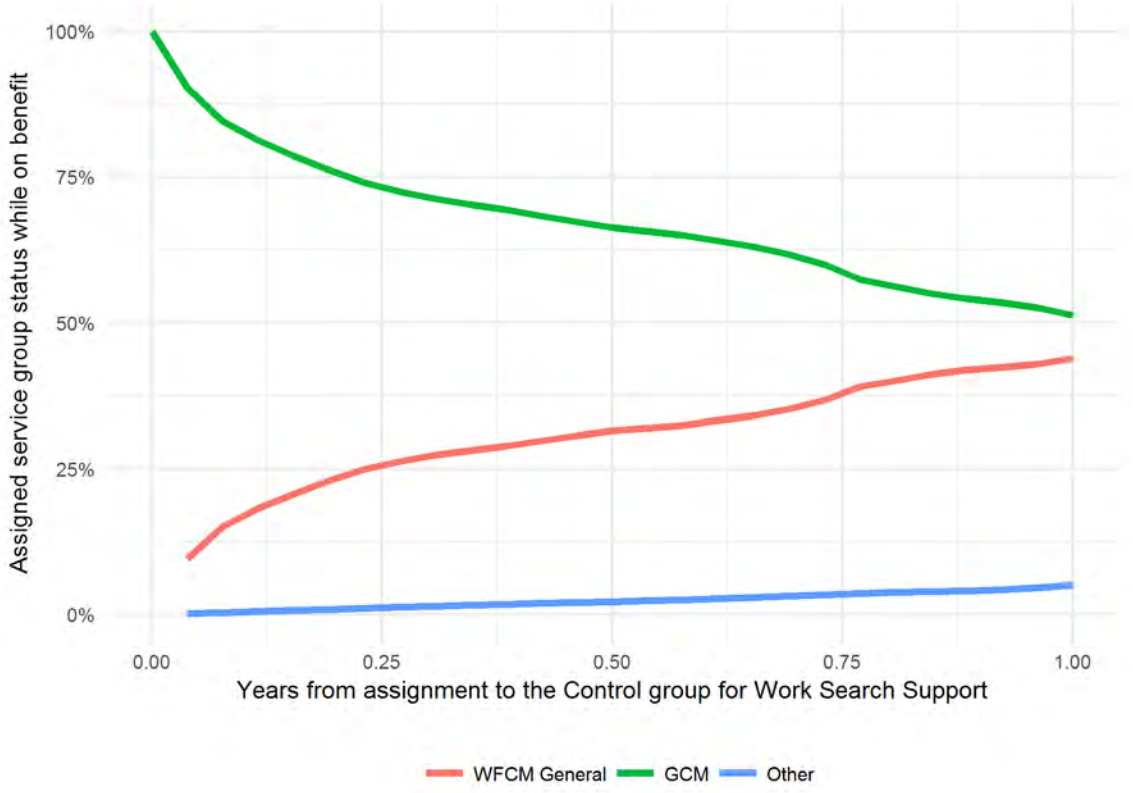
Figure 7: Assigned service status for people in the WFCM General control group in their first spell on main benefit



First spell: This is from assignment to control group through to the first exit from main benefit.

Figure 8 shows the assigned service status of people allocated to the Work Search Support control group. For this service, all control group members are in GCM at assignment. However, over the following year this proportion steadily falls as people commence WFCM General. This trend illustrates the challenge of control groups being allowed to be assigned to other services. We will compare the experience of treatment and controls for each service in later sections of this report.

Figure 8: Assigned service status for people in the Work Search Support control group in their first benefit spell



First spell: This is from assignment to control group through to the first exit from main benefit.

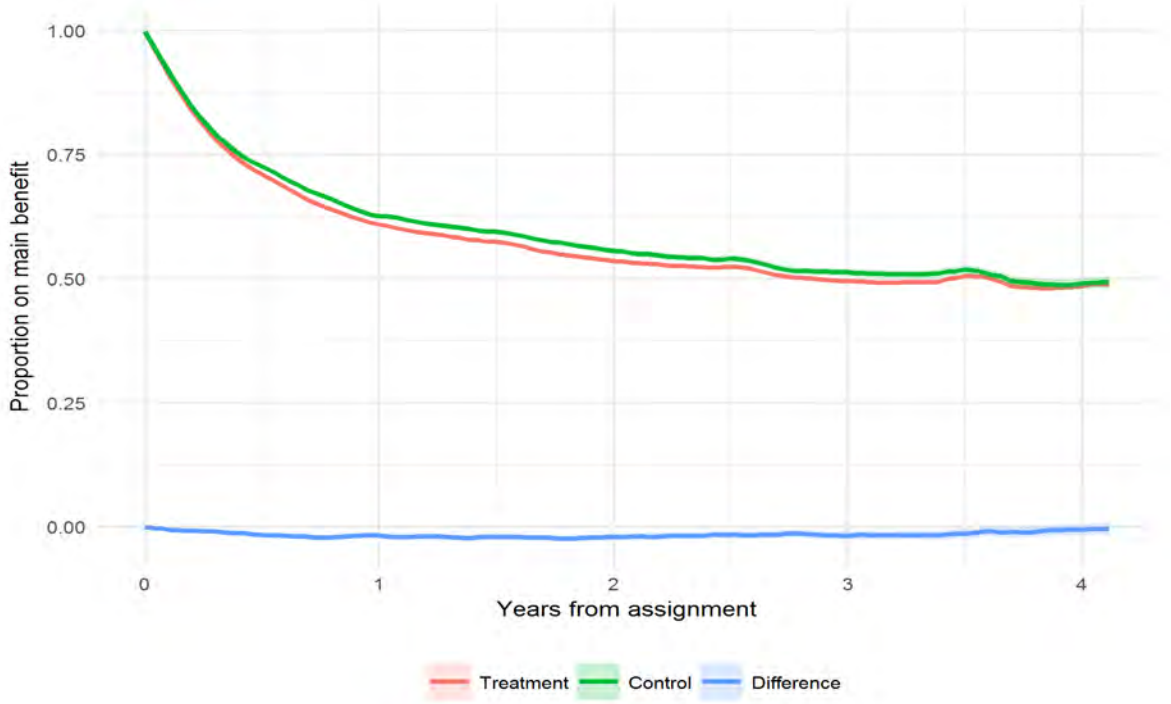
Analysis of impacts beyond the control group period

As discussed at the start of this section, the main reason for the service control group was to ensure that similar individuals received a range of alternative services. Once the control group period ends, those control group members who are still eligible can be assigned to the service.

Differences in outcomes between participants and control persist

We expected the removal of the control group holdout period would result in a convergence in main benefit outcomes between the participant and control group. This is exactly what happened. For Work Search Support (Figure 10) we see the off benefit outcomes of the control and treatment converge at around one-and-a-half years after assignment (ie six months after the end of the control group hold out period). For WFCM General (Figure 9) it took nearly twice as long, with treatment and control group off benefit outcomes converging at four years (one year after the end of the control group holdout period).

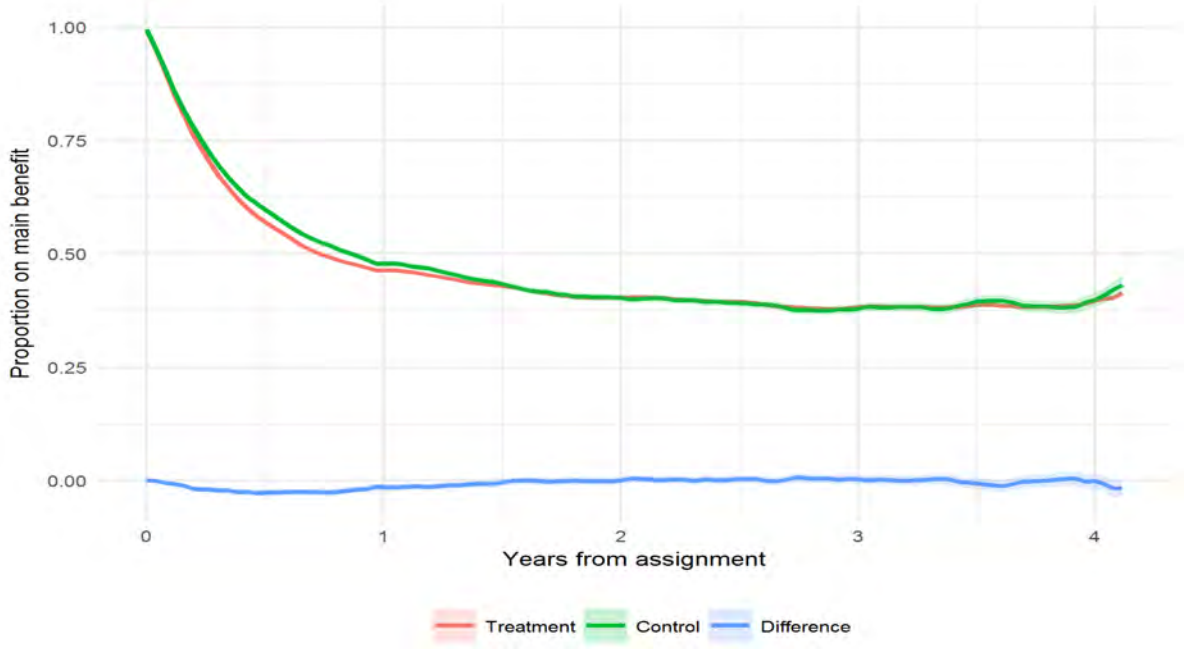
Figure 9: Difference in proportion of participants and control on main benefit in each interval after assignment for the WFCM General



On main benefit: Receiving a main benefit such as Jobseeker Support or Sole Parent Support, and can also be receiving supplementary assistance such as Accommodation Supplement.

Interval: Is the average proportion of each interval that the group is on main benefit (ie the total time on benefit in the interval divided by the total number of clients observed in the interval).

Figure 10: Difference between participants and control on main benefit outcomes in each interval after assignment for the Work Search Support



On main benefit: Receiving a main benefit such as Jobseeker Support or Sole Parent Support, but and can also be receiving supplementary assistance such as Accommodation Supplement.

Interval: Is the average proportion of each interval that the group is on main benefit (ie the total time on benefit in the interval divided by the total number of clients observed).

Table 10 summarises the difference in outcomes between participants and controls for each service at selected intervals after starting the service. Three services show longer lasting differences in off benefit outcomes between control and treatment groups. These are WFCM General, WFCM Integrated and WFCM Intensive Client Support (Entrenched).

Table 10: Difference in on benefit outcomes at selected lapse intervals between participants and controls (treatment – controls = impact)

Service	Years from assignment to control or treatment group							
	0.5	1	1.5	2	2.5	3	3.5	4
MHES Trial	0.3	-0.1	0.9	2.2	0.3	1.3	-1.0	
SPES Trial	-3.0	-1.2	0.9	-1.5	0.6	-1.9	0.7	
WFCM General	*-1.6	*-1.8	*-2.1	*-2.0	*-1.6	*-1.8	*-1.3	
WFCM Health	-0.7	-0.7	-0.5	-1.0	0.8	-0.7	1.0	
WFCM Integrated	*-9.5	*-6.8	*-8.5	*-7.3				
WFCM Integrated (Early)	-6.9	-5.4	-1.9	0.5				
WFCM Integrated (Entrenched)	*-11.3	*-7.6	*-12.6	*-11.3				
WFCM Intensive	*-6.8	*-7.4	*-9.6	*-9.7	*-7.6	*-9.4	*-8.8	
WFCM Intensive (Youth)	0.1	-0.1	-1.3	1.1	2.8	-1.8		
WFCM pilot	*-3.7	*-4.4	*-2.2	*-2.0	-1.2	-0.2	0.6	0.5
Work Search Support	*-2.7	*-1.4	-0.4	0.1	0.4	-0.3	-0.8	
Work Search Support (pilot)	*-2.6	-1.1	-0.7	-0.7	-0.3	0.2	0.0	0.8

ppt: percentage point difference (eg 70% - 60% = 10 ppt).
*: difference significant from zero at the 95% confidence interval.

Control and participant status after assignment

One reason for the reduction in the difference in outcomes between WFCM pilot participants and controls after 52 weeks is the increase of controls in WFCM General.[10] To explore this further, Figure 11 summarises the assigned service status in the period after participants and control were assigned in the WFCM pilot. Of interest here is the experience of control after the control period ends (at 52 weeks). After this point, control group members can be assigned to WFCM General if still eligible. Indeed, at 54 weeks we saw a large jump in the proportion of all controls participating in WFCM General.

[10] The control group period for WFCM pilot ended after the national roll out of the SDM in July 2013. After July 2013 WFCM pilot control group members were prevented from being assigned to WFCM General.

Figure 11: Participants and controls participating in WFCM pilot after assignment while on their first on-benefit spell

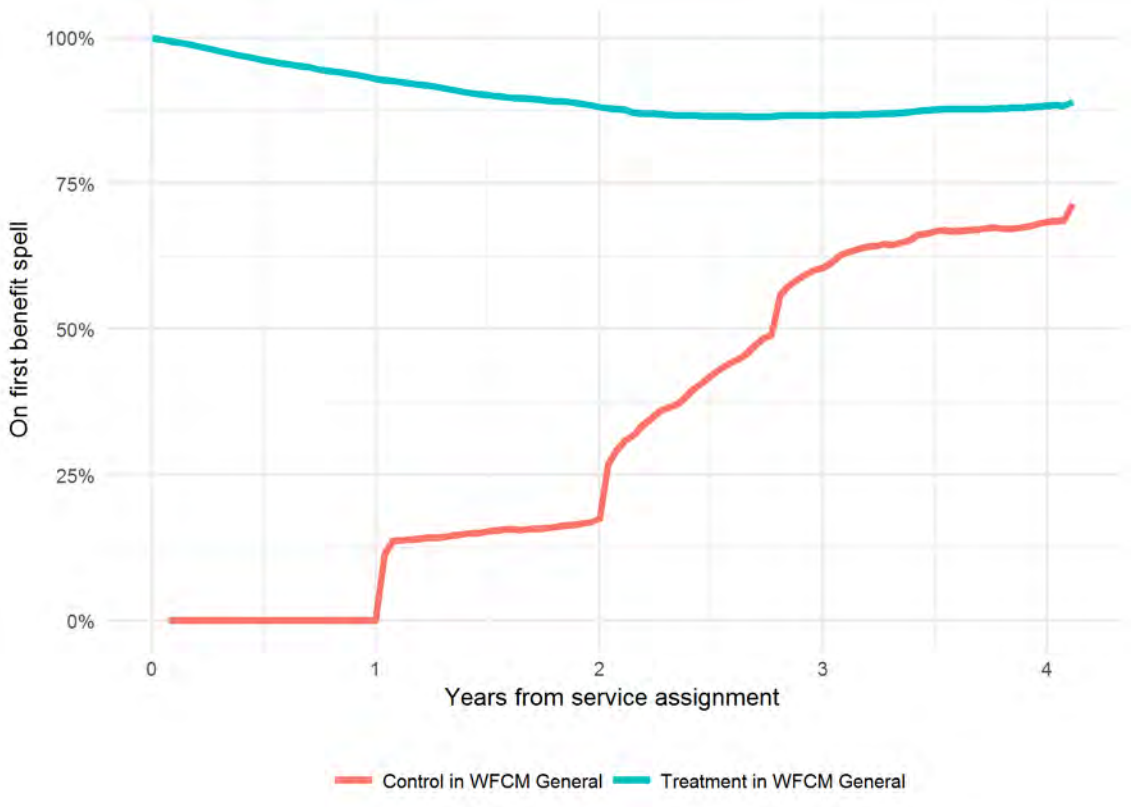


Table 11 summarises the proportion of the controls who are receiving the service they were assigned to the control group for. Across the services, the proportion ranges from zero through to 67 percent.

Table 11: Proportion of control in service at selected periods after assignment

	Years after assignment to the control group							
	0.5	1	1.5	2	2.5	3	3.5	4
MHES Trial								
SPES Trial	0.1%	0.4%	0.3%					
WFCM General	0.0%	0.1%	16.1%	19.1%	42.9%	60.8%	67.5%	
WFCM Health	0.1%	0.1%	1.7%	3.0%	7.0%	7.5%	19.2%	
WFCM Intensive (Early)								
WFCM Intensive (Entrenched)				1.1%				
WFCM Integrated Support	0.6%	0.8%	3.7%	3.3%	4.2%	12.1%	30.4%	
WFCM pilot	0.8%	0.8%	32.1%	37.4%	39.4%	37.4%	44.5%	46.8%
Work Search Support	0.0%	0.1%	25.5%	23.8%	19.1%	15.8%	15.2%	

Participant and control group profiles

In this section, we examine the profile of participants and controls for each service at the point of assignment. Statistical theory tells us that if the two groups are properly randomised in the allocation to participant and control groups, we should see no substantial differences in the profile of the two groups. However, as discussed in the previous section, an error in the randomisation process had systematically affected the allocation into control and participant groups. Here we examine how well the re-weighting of the control group has reduced the impact of this error.

Calculation of differences in means

Calculating the standard errors for variable means and proportions as well as the differences between the participant and control group was complicated by the:

- block design underpinning the allocation of clients into the participant and control group
- large variation in the weights assigned to individual observations.

For the calculation of standard errors, we used SAS SurveyMeans procedure. Alongside the control weights (participants have a weight of 1), we included Selection date, geographic location (Geo block) and client group (Group block) as stratification variables. For estimating the difference between participant and control group we used the SAS SurveyReg procedure. Tests for differences between participant and control group mean were based on the following equation:

$$\boldsymbol{Y} = \boldsymbol{\alpha} + \boldsymbol{\beta X_{pc}} + \varepsilon$$

Where: Y is the variable whose mean or proportion we want to compare

α intercept term

β for the X_{pc} term indicating membership in the control or participant group

ε: error term.

Our test of a significant difference was based on the standard error of β. Here we have not made an adjustment for multiple tests (ie Bonferroni or False Discovery Rate adjustments).

Figure 12 summarises the difference in mean or proportions for profile variables at assignment to control and treatment groups for WFCM General. Differences were standardised by dividing the difference between treatment and control, by the mean or proportion of the treatment.

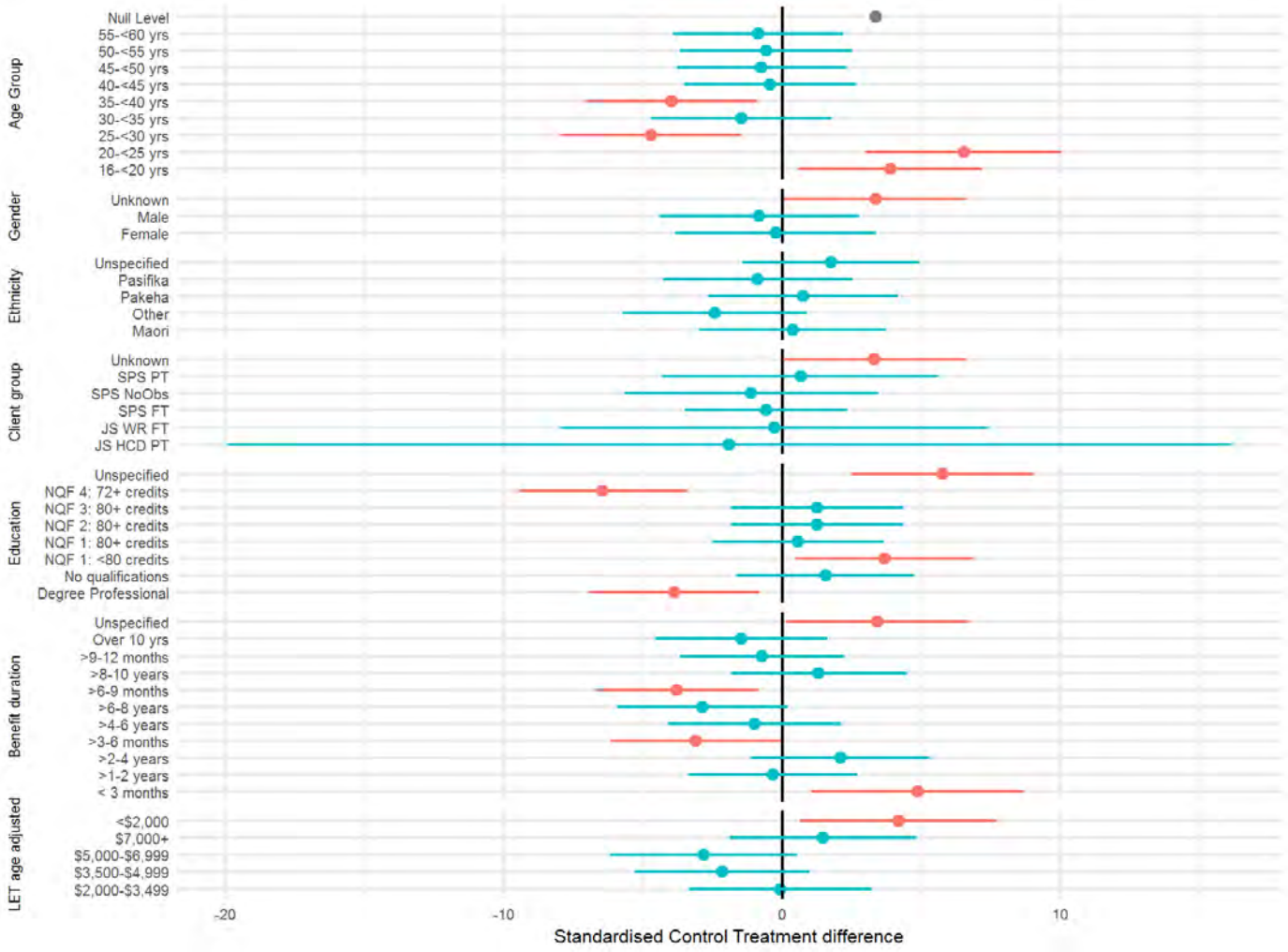
$$\frac{(\mu_t - \mu_c)}{\mu_t}$$

Where: μ = mean or proportion for the treatment (t) or control (c) group

Therefore, a positive value indicates that the mean or proportion of the treatment group is larger than that of the control group. If the standard error bars in Figure 12 do not cross zero then the difference is significant at the 95 percent confidence interval.

Overall, a minority of the differences between the treatment and control means are not significantly different from zero.

Figure 12: Standardised difference in profile between the treatment and control group for WFCM General



Error bars represent the 95% confidence interval for the difference in means or proportions of the profile variables between the treatment and control group.

Profile of participants and control group

Here we compare the demographic profiles of the participant and control group for each service. If the randomisation was successful we would have expected to see few differences between the two groups. However, as discussed in the previous section on the Control Group Allocation, that introduced an observed bias into the allocation mechanism (see page 30).Table 12 below we show the number of tests of significance and the number that were significantly different at the five percent alpha level. If the profiles of the two groups have equivalent profiles, we would have expected no more than five percent of the tests to have a p-value below 0.05.

Table 12: Proportion of significant differences between participants and controls for original and adjusted weighting

Service	Original weighting	Adjusted weighting
MHES Trial	17%	17%
SPES Trial	7%	7%
WFCM General	9%	7%
WFCM Health	5%	5%
WFCM Intensive (Early)	1%	1%
WFCM Intensive (Entrenched)	2%	2%
WFCM Integrated Support	16%	7%
WFCM Integrated Support YS	3%	3%
WFCM pilot	2%	3%
Work Search Support	5%	2%
Work Search Support pilot	4%	4%

Apart from the two IA trials (SPES and MHES), the adjustment to the weighting has helped to reduce the number of significant differences between the participant and control groups. However, despite this improvement, we can still see that for a number of services the proportion of significant differences is over five percent. Despite these differences, we have not attempted to control for related variables (ie if there is a difference for males then there will be a difference for females). In addition, on examining these significant differences, many occurred for characteristics involving a small number of participants (ie less than one percent of the participants had the characteristic in question). Because there are still observed differences between the participant and control, we will also include the relevant prioritisation variables in the regression model to further control for the impact these differences might have on participant's outcomes.

MHES was not affected by the randomisation error, but as Table 12 shows, this intervention shows the largest number of significant differences. We have traced this to the inclusion of non-Jobseeker HCD participants in the treatment arm, but the control arm was made up almost entirely of Jobseeker HCD. Given the number of observations (over 2,000 in each arm), this is very unlikely to have occurred by chance. As a consequence, the treatment group are at a lower risk of long-term benefit receipt than the control group.

We have not been able to identify the reason why this systematic difference occurred in the assignment of the control and treatment groups for Mental Health Employment Service Trial. For this reason, we are more cautious about the results of this service.

Defining service spell duration

The evaluation centred on the time that individuals spent in each service. As will become clear in later sections, our estimate of the impact of each service was based on the time that people spent in each service. The assumption we made was that the longer the exposure to a service, the greater its influence on the individuals' time off main benefit.

However, we could not use the observed duration in each service, instead, we needed to use the assigned service history (detailed below). There are two important motivations for creating an assigned service history.

- Service spells end with ineligibility: as discussed previously, we fully observe the service allocation process, and for this reason can be confident there are no unobserved selection effects in people commencing a service. However, we cannot say the same for exits. Here we only knew that a person had become ineligible based on their observed characteristics, but we did not necessarily know the underlying reasons for them becoming ineligible. For example, people on Jobseeker Support may choose to gain a medical exemption for work obligations to drop out of a service.

  Without controlling for such behaviour, we would overstate the effectiveness of the service. An upward bias would occur because people who respond by making themselves ineligible are assumed to be also less likely to exit from benefit. Accordingly, those who remain in the service are more likely to exit benefit relative to those who were assigned to less intensive case management service (independent of receiving more intensive case management).
- Service spells end at benefit exit: exit from benefit was an important reason that participants became ineligible for a service. In other words, actual service spell does not extend to the off-benefit period. In these instances, we needed to associate the last service a person was on to the time they subsequently spent off benefit. If we did not, it would appear that service participation only relates to time on main benefit and, therefore, conclude what service a person was on had no impact on duration off main benefit.

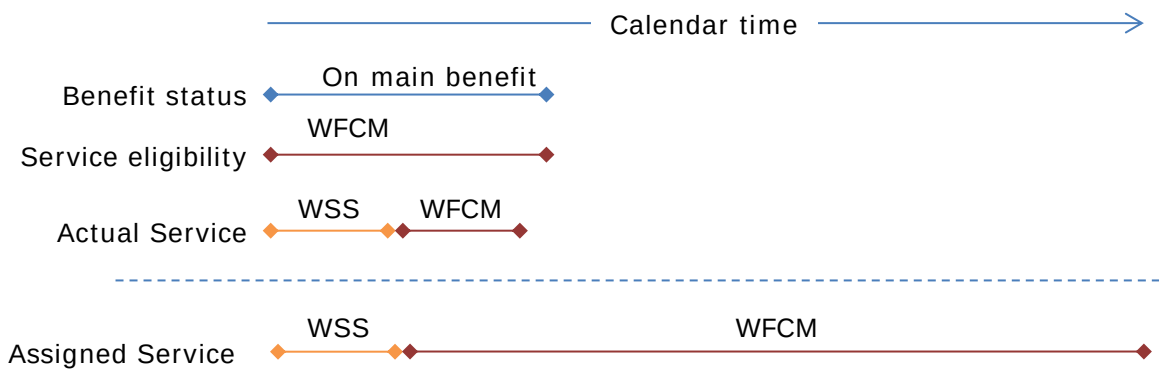
Creating an assigned service history

The creation of an assigned service status history can be difficult to conceptualise. The easiest way to think of assigned history is that it centres on the assignment of people to services, with their spell on a service only ending if they are assigned to a different service. To help understand how this works, we use a number of stylised scenarios to illustrate how we derived service history for different scenarios.

Scenario A: Exit from main benefit

The simplest example is where a person exits from main benefit after commencing a service, in this case, WFCM. In Figure 13 we show the actual benefit status of the person, their eligibility for WFCM and what service they are on. In the scenario, the individual starts on Work Search Support and is allocated to WFCM, after which they exited benefit. Their exit from benefit made them ineligible for WFCM and their actual service spell on WFCM ends. Our assigned service spell history extends the individual's WFCM spell until the end of the observed period.

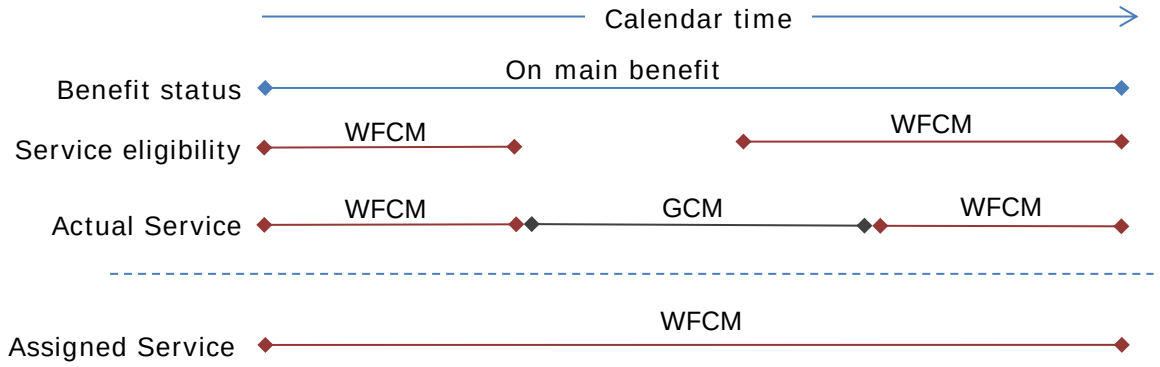
Figure 13: Scenario A client exits benefit while on WFCM



Scenario B: A participant becomes ineligible for WFCM for a period

The second scenario involves the case where a person becomes ineligible for a service for a period. In Figure 14 we see a person on WFCM who becomes ineligible for reasons such as a work test exemption, and is moved to GCM. At a later point, they again become eligible for WFCM and are subsequently assigned back to the service.[11] For the evaluation, we treat this period on GCM as part of the WFCM assignment spell. The reasoning here is that we do not observe why a person becomes ineligible and do not want to introduce these unobserved selection effects of exiting from a service into the analysis.[12]

Figure 14: Scenario B: A person becomes ineligible for WFCM for a period



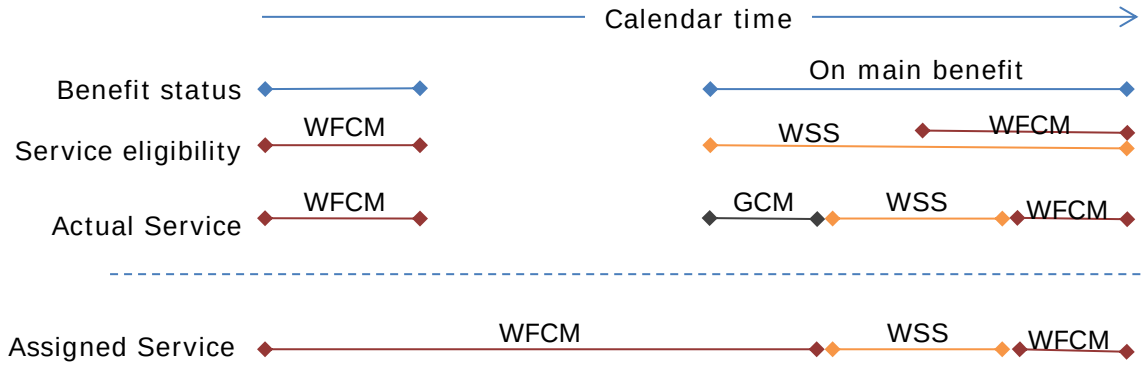
Scenario C: A person exits and returns to main benefit

Scenario C is for a person who exits benefit and subsequently returns to main benefit. In Figure 15, the individual exits from benefit while on WFCM. They subsequently return to main benefit, although eligible for Work Search Support, there are not enough places, and so the individual is assigned to GCM. At a later point, they move from GCM into Work Search Support and ultimately back to WFCM. For the assigned service spell in scenario C, we first ignore both the benefit exit and time on GCM and continue the individual's spell on WFCM until they are assigned to Work Search Support. Again, the reasoning is that we only end derived service spells when people are explicitly allocated to another service. This is not to say that in the

[11] People are not automatically assigned to a service if they become eligible. Besides eligibility, there needs to be sufficient spaces available on the service before eligible individuals can be assigned to it.
[12] For the purpose of this evaluation we are not looking at the underlying reasons why people become ineligible. Instead, we are simply ensuring that these behavioural responses do not confound our estimate of the impact of being assigned to the service on the time spent off main benefit.

analysis of the service's impact on time off benefit, we ignore the time off benefit in this example. Instead, the time spent off main benefit does not change the assigned service status of the individual.

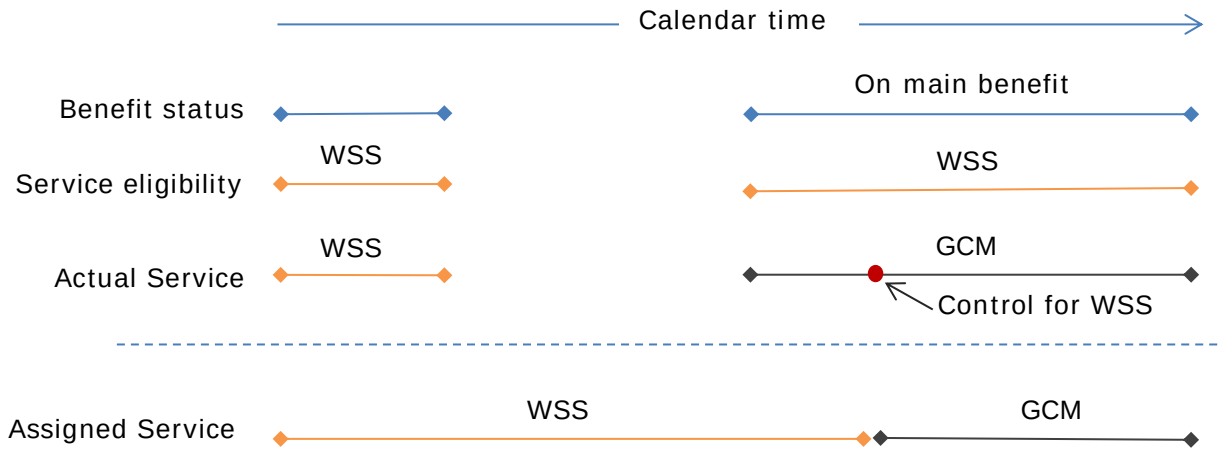
Figure 15: Scenario C: A person exits and returns to main benefit



Scenario D: A person is assigned to the control group of a service

The last scenario involves the case where a person is assigned to a control group for a service they have been on previously. Figure 16 shows an individual who was on Work Search Support, exited and returned to benefit on GCM. On returning to benefit, the individual is selected into the control group for Work Search Support and remains in GCM. In this instance, we treat the assignment to a control group in the same way as a selection onto a service. Instead of the assigned service being the service a person moves to, the assigned service is the service that they are currently on. In the example case, the assigned service at the assignment to the control group is GCM. Analysis of outcomes and costs for this individual would start from the point when they were assigned to the control group.

Figure 16: Scenario D: client assigned to a service control group



Analysis of service spell duration

In this section, we illustrate the difference between actual and assigned service status for two services (WFCM General and Work Search Support).

Actual and assigned service status of people starting WFCM General

Figure 17 shows the broad service groups people are in after being first assigned to WFCM General. At assignment, nearly all individuals are in WFCM, as expected. However, the proportion of WFCM drops over the following year, with less than half still in the service. One important reason for this decrease is the proportion of people who have exited main benefit and are therefore not in any service. The remaining individuals have become ineligible for WFCM and are in GCM. It is the last group that poses the greatest challenge for the evaluation, since it may well be that becoming ineligible for WFCM General is a behavioural response to being assigned to WFCM. Because assignment to case management service was mandatory, the only way to opt out of the service was to become ineligible. People could, for example, seek to get a medical certificate to gain a deferral of their work obligations. If we do not account for these ineligible individuals, we could well overstate any impact of WFCM on participant's outcomes.[13]

Figure 17: Actual service status of clients starting WFCM General

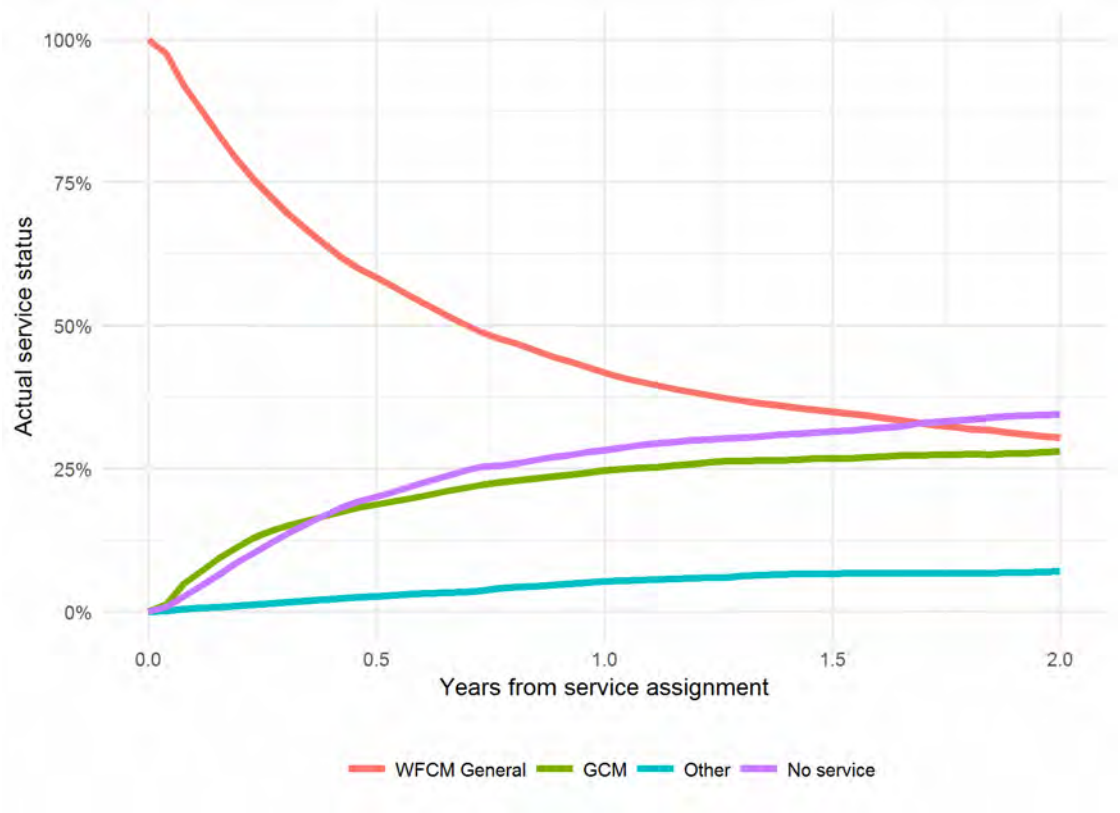
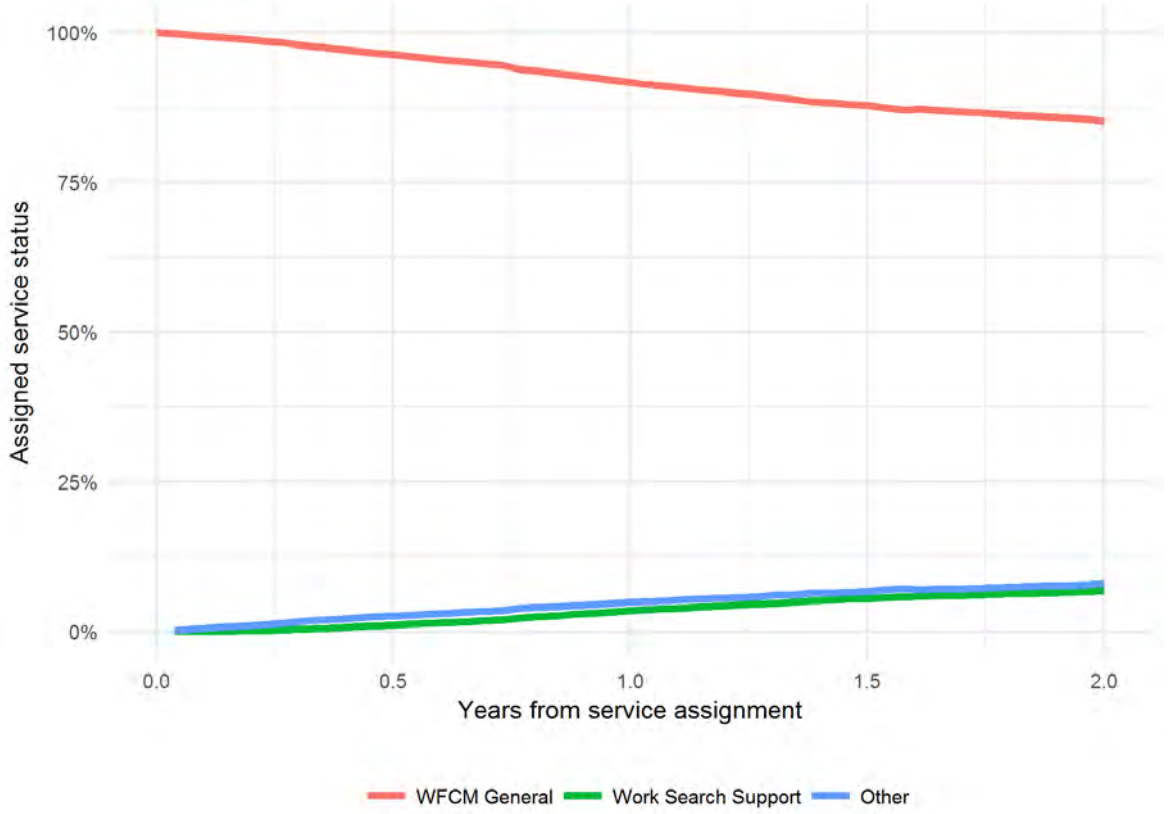


Figure 18 shows the assigned service status for the same group starting on WFCM General in Figure 17. The application of the rules outlined in the above scenarios can be seen in the assigned service. To quickly restate, people could not voluntarily opt out of the service,

[13] The first reason is that the denominator decreases, thereby increasing apparent number off main benefit. The second reason is more speculative, in that we assume people who seek to become ineligible for active case management would also have a lower probability of exiting main benefit. Therefore, if this assumption is true, by excluding these people we would, on balance, increase the probability of service participants exiting the benefit.

therefore, the only means of leaving a service was to become ineligible (including exiting benefit) or being assigned to another service. Because the assigned service ignores exits from benefit and becoming ineligible for a service, the proportion on WFCM remains at close to 90 percent after one year. Of the remaining 10 percent, the second largest group are those on Work Search Support. These people would have exited from main benefit, returned to benefit and been subsequently allocated to Work Search Support. The other group includes assignment to higher priority services such as the IA trials, WFCM Intensive Support or WFCM Integrated Client Support.

Figure 18: Assigned service status of clients starting WFCM General



Actual and assigned service status of people starting Work Search Support

Work Search Support service status is perhaps the most complicated of all the services in the SDM. The reason is that it sat between GCM and WFCM services in terms of allocation priority. As such, people were allocated into and out of Work Search Support on a regular basis. Figure 19 shows the service status of people commencing Work Search Support. Like WFCM, the proportion of people remaining in Work Search Support falls sharply, such that by one year less than 20 percent remained on the service. Again, like WFCM, the two drivers are people leaving benefit and are not on a service (the 'No service' group) as well as becoming ineligible for intensive case management and moving to GCM. However, unlike WFCM, we also see a large proportion of Work Search Support participants moving to more intensive services (eg WFCM General).

Figure 19: Actual service status of people starting Work Search Support

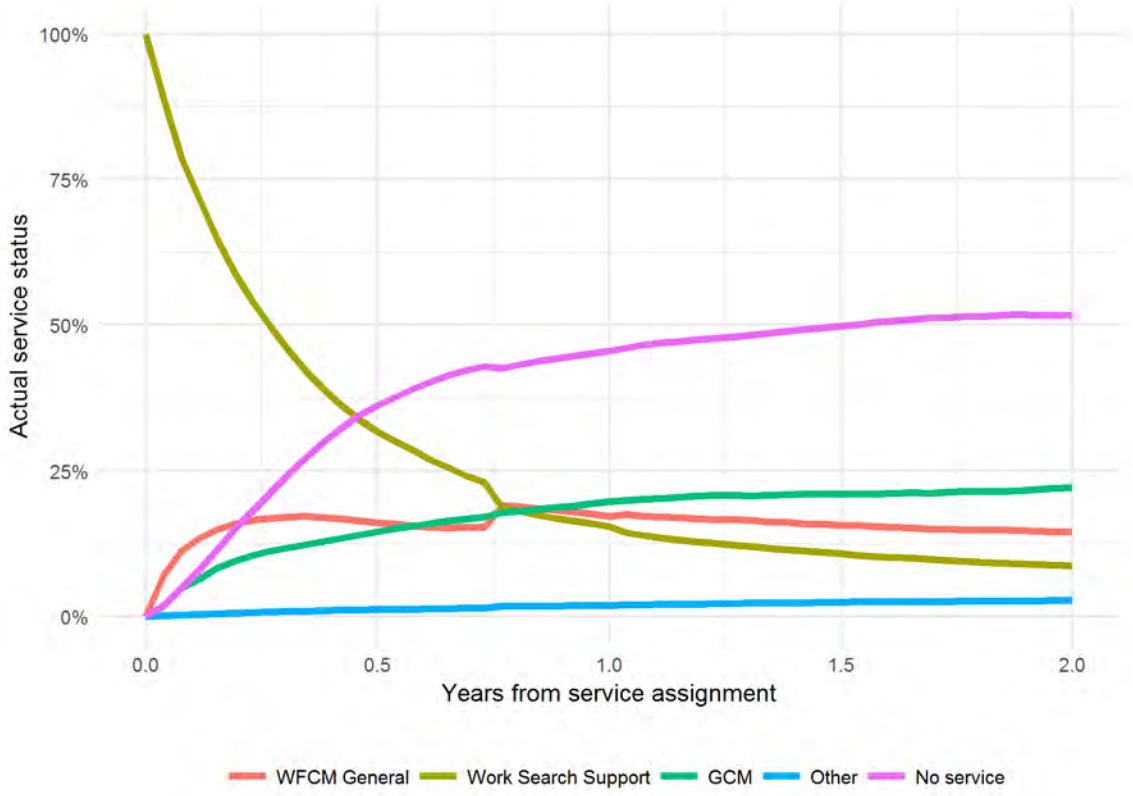
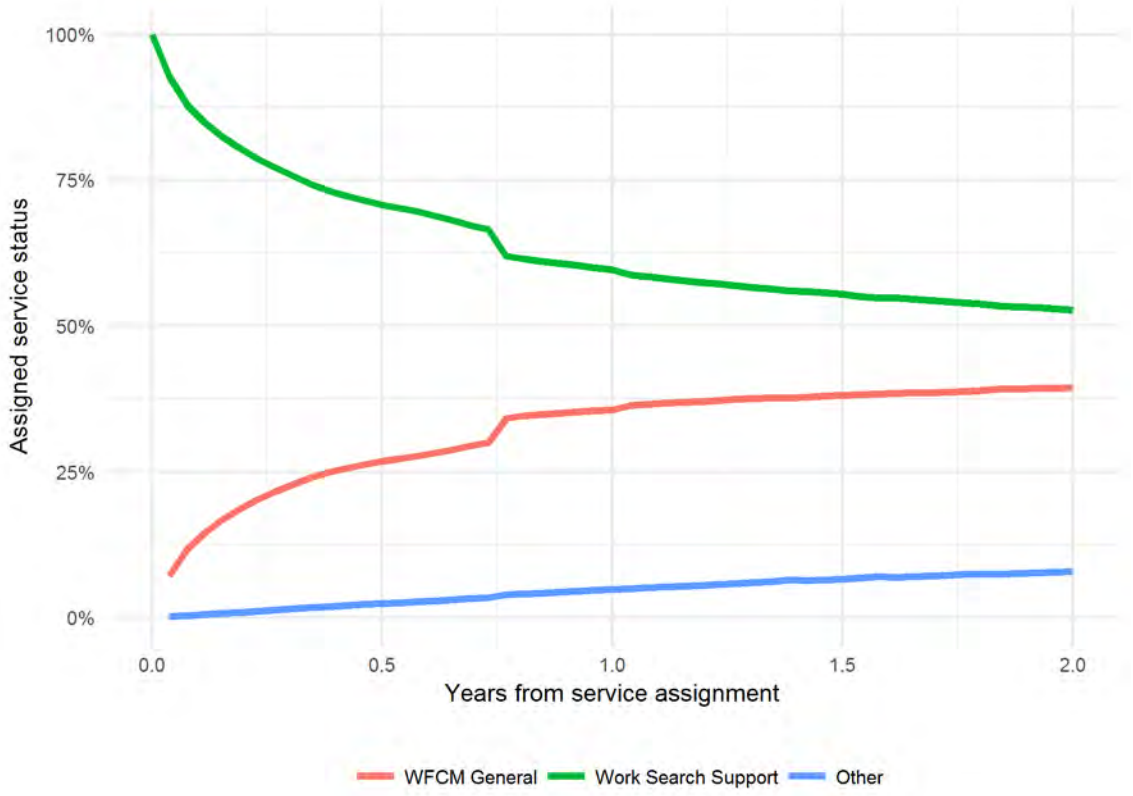


Figure 20 presents the assigned service status for people starting Work Search Support. The proportion of people who remain in Work Search Support is much higher when we apply the assigned service rules. At one year, more than half are still assigned to Work Search Support and primarily include those who left main benefit or became ineligible and shifted to GCM. Nearly all of the remaining people are in WFCM General. The logic here is that assigned service allows for allocation to another intensive service. It also follows that after being allocated to WFCM, we ignored any subsequent exits from benefit or becoming ineligible. For this reason, the proportion of WFCM using the assigned rules is much higher than the actual service status as shown in Figure 19.

Figure 20: Assigned service status for people starting Work Search Support



Changes in service status after assignment

The results from Figure 20 also point to a challenge in evaluating the impact of services such as Work Search Support, since the outcomes of people assigned to Work Search Support is a combination of the effects of both Work Search Support and WFCM General. One of the goals of the subsequent impact analysis is to try to separate the effects of each service on people's outcomes (see next section).

Estimating impact of SDM services

Estimating the impact of each SDM service presents several challenges. Although we have a controlled process to allocate people to different services as well as incorporating a control group, estimating the impact of any one service was not straightforward. In this section, we describe our conceptual framework for estimating the impact of services before going into detail on the estimation method used and the results from our analysis.

Outcome measures

Before looking at the impact of each service, we first discuss the outcome measures used in the evaluation and how they are represented in the impact analysis.

Time off main benefit

While the evaluation was focused on estimating welfare ROI, we based this calculation primarily on the impact that services have on the time people spend on and off main benefit. The definition of main benefit includes the time people are entitled to receive any main benefit. The most typical main benefits include Jobseeker Support, Sole Parent Support and Supported Living Payment. Conversely, we did not count time where a person is receiving supplementary assistance only (eg Accommodation Supplement). Readers should also note that entitlement does not always mean a person is receiving income support payments. Within a benefit entitlement period, the amount of income support received may vary by other income earned. For example, for people doing part-time work, the amount of income support paid to them is abated according to their earnings. In addition, people may have their entitlement suspended for short periods within which they receive no income support payments from their main benefit. Such suspensions can occur if the individual has a brief period of employment (usually less than eight weeks).

The measure of the time off main benefit was an 'as at' measure. What this means is that we measured the benefit status as it was on the day and ignored any retrospective actions to benefit entitlements. Retrospective actions occur because changes in entitlement to main benefit are not immediately reflected in the administrative data. For example, it might take several days for a benefit application to be assessed before it is granted. Conversely, a person may be fraudulently receiving a benefit and these entitlements are subsequently removed from the administrative systems.

The reason for using an as at rather than current view of entitlement was to ensure that the analysis units remain stable over time and are not potentially biased through changes in case management that in turn result in differences in how benefit entitlement is managed. For example, more frequent contact under an intensive service may result in the detection of overpayment and result in a retrospective adjustment to benefit entitlement spells that would not have occurred with lower intensity case management.

Income Support Expenditure

We calculate total income support payments based on current payment rates for each individual included in the analysis. The reason for using current rates was because, at the time of the analysis, we did not have access to 'as at' income support payments. Income support payments come in three forms:

Primary or first tier: these are payments for people entitled to a main benefit such as Jobseeker Support or Sole Parent Support. Primary payments make up the bulk of income support payments received while on main benefit. Primary benefits are taxed and therefore we used payments net of tax.

Secondary payments: secondary payments are regular top-up payments to cover issues such as accommodation costs, costs with disability and childcare. Second tier payments are made either while a person is on a main benefit or they can be paid to low-income working families. Secondary payments are not taxed, and we included all secondary payments including family tax credit paid by MSD on IR's behalf.

Third tier/hardship payments: the final level of assistance are ad hoc payments for one-off events. Hardship payments can be made to both families on main benefits as well as low-income families off main benefit. Like secondary assistance, hardship payments are not taxed. However, hardship payments can be recoverable. In other words, they have to be paid back (interest-free) over a set period of time. In this analysis, we included both recoverable and non-recoverable assistance.

Staff time

An important aspect of this evaluation was to understand how staff allocated their time to working with individuals on their caseload. Staff time was a key driver of the overall cost of each service and is covered in the next section. But in addition, it is informative on its own to see how frontline staff responded to the service allocation process.

In 2016, MSD developed a measure of staff time based on timestamps from administrative systems used by staff. In each instance, when a staff member searches for information on a person or updates the person's record, the MSD data warehouse records each event with a timestamp. From these timestamps, person identifiers, and by linking staff identities across administrative systems we were able to create a digital signature of each staff member's day. Through this process, we estimated the amount of time a staff member spent with individuals as well as a broad indication of what they were doing with them. For more detail on the staff time allocation model, refer to the technical notes (MSD, 2016).

Cost of case management activities

Following on from the staff time allocation, we also used the MSD's individual Cost Allocation Model (iCAM) to identify the cost of case management activities undertaken for individual clients. The iCAM covers costs such as:

- income support administration
- case management
- employment assistance interventions.

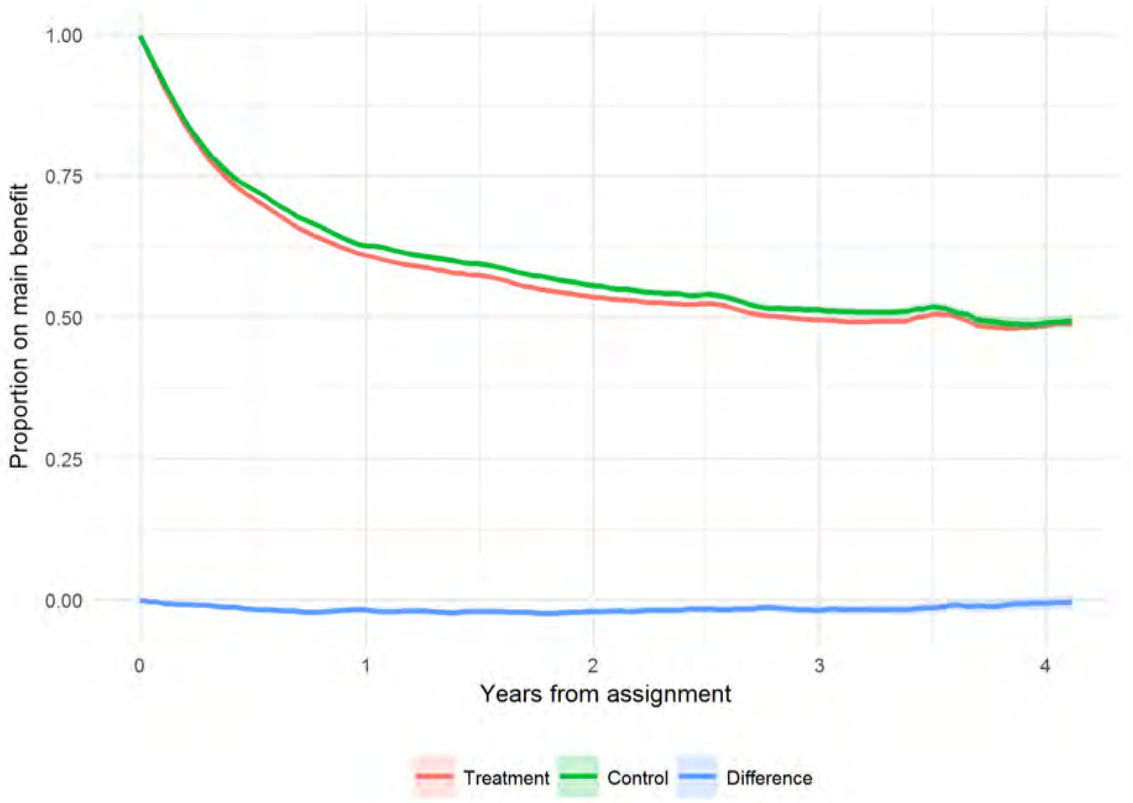
Like the staff time allocation model, iCAM costs are calculated at the level of the individual and include the costs specific to each event (ie these are not averaged costs). For more detail on the construction of the iCAM refer to the model's technical notes (de Boer & Ku, 2017).

Interval and cumulative measures

As already shown in the previous section, we measured the outcomes of participants and control group participants from when they are assigned to the group (lapse period 0). We can then measure outcomes over different periods before and after assignment date. In addition, we can measure outcomes over intervals or cumulatively.

By interval, we mean a discrete period (eg at 52 weeks after assignment). Figure 21 illustrates interval outcomes for WFCM General participants and controls. Each point in the chart represents the proportion of time in the interval after assignment that participants or control spent on main benefit. Reporting interval outcomes are very useful in understanding how outcomes are changing in each period after assignment. Interval outcomes enable us to see how the outcomes of the control and participants change relative to each other (eg do the outcomes converge over time). For WFCM General the convergence in outcomes appears to occur at nearly four years after assignment.

Figure 21: Difference between participants and controls off main benefit outcomes in each interval after assignment for the WFCM General

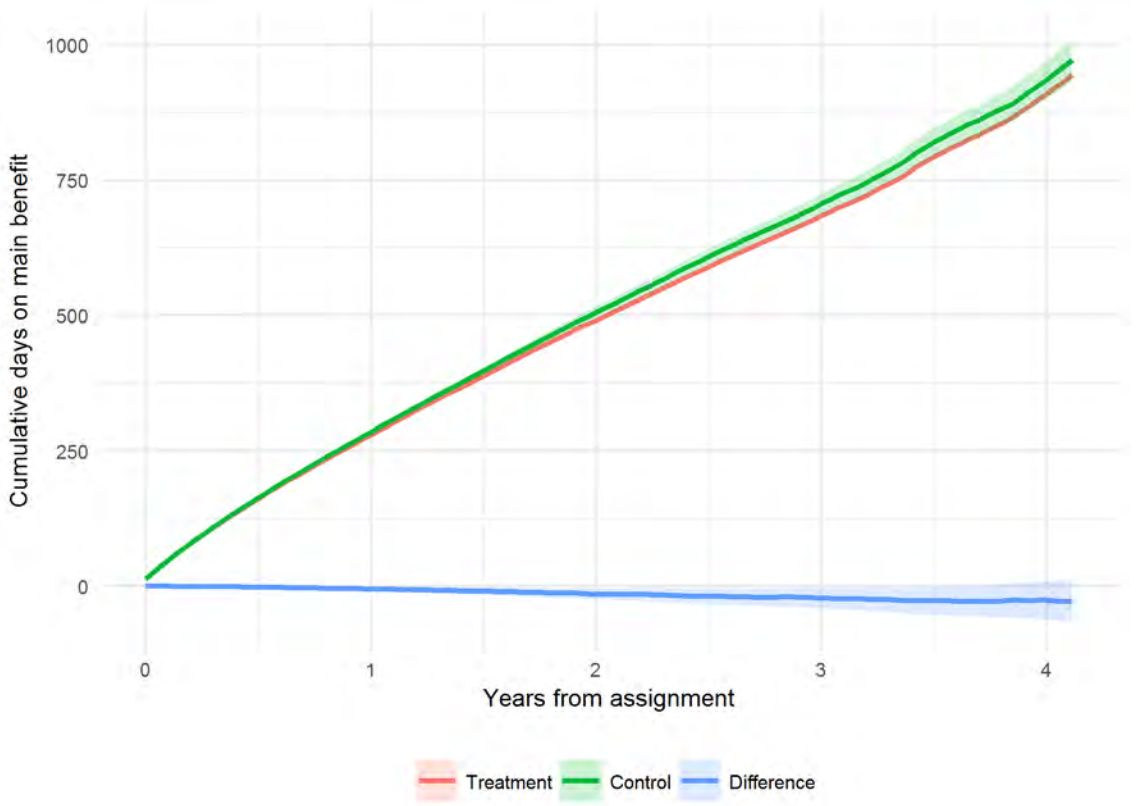


On main benefit: receiving a main benefit such as Jobseeker Support or Sole Parent Support (we do not include time receiving supplementary assistance such as Accommodation Supplement as being on benefit).
Interval: is the average proportion of each interval that the group is off main benefit (ie the total time off benefit in the interval divided by the total number of days in the month).
Weighting: control and treatment outcomes are weighted.

However, interval outcomes have the disadvantage that they do not provide a summative view of the outcomes since assignment. To answer summative questions, we need to use cumulative outcome measures. The only difference between interval and cumulative measures is that the cumulative is always measured from lapse period 0. Figure 22 shows the cumulative time off main benefit from assignment (lapse period 0) for WFCM pilot participant and control groups. As

the follow-up period increases, cumulative outcomes can only stay the same or increase (because they are cumulative sums of count data). The small differences between the treatment and controls at each lapse interval in Figure 21 are accumulated in Figure 22. But because the variance in outcomes also increases with longer follow-up periods, the width of the confidence intervals also increases (as indicated by the light shaded areas for each line in the charts).

Figure 22: Difference between participants and control off main benefit outcomes from assignment for WFCM General



On main benefit: receiving a main benefit such as Jobseeker Support or Sole Parent Support (we do not include time receiving supplementary assistance such as Accommodation Supplement as being on benefit).
Cumulative: is the average days on main benefit from assignment (lapse period 0) to the end of each interval.
Weighting: control and treatment outcomes are weighted.
Confidence Interval: shaded area indicates the 95 percent confidence interval for the number of days on main benefit.

Defining service effectiveness

Before we examine the technical issues in estimating the impact of case management services, we first consider what we are trying to estimate. Impact evaluation centres on the counterfactual question: what would have happened without the intervention? The assumption is that any difference in outcomes between the intervention and counterfactual scenarios provides evidence of the causal effect of the intervention. For this evaluation, the intervention and counterfactual scenarios are expressed as follows:

- Intervention: outcomes of participants if they remain in the service
- Counterfactual: outcomes if these same participants had spent the same period of time in an alternative service (namely GCM).

In other words, we are trying to estimate the impact of each service relative to a baseline. Further, to enable ease of comparison between services, we try to estimate the impact to the same baseline of GCM. However, we were unable to achieve this latter goal in all instances. Specifically, WFCM Integrated Support (Youth) targeted a group of people (those leaving YP/YPP benefits) where none of this group was assigned to GCM. Accordingly, we have no empirical information on the expected outcomes of ex-YP/YPP if they had been assigned to GCM. Instead, we used WFCM General as the baseline service for WFCM Integrated Support (Youth).

Challenges in achieving a robust counterfactual

The problem is that, even with the best designed and run RCTs, the experience of the treatment and control rarely fit into a neat counterfactual framework described above. The current evaluation was no exception. Here the participants represent the intervention scenario, while the control group represents the counterfactual scenario. For each we have the following issues:

- Participants: participants spent time receiving other services.
- Control: control group members received a range of different services, not just GCM.

The challenge for the evaluation is correcting for these aspects of the RCT design to arrive at an estimate of the service impact relative to the GCM baseline.

Service participation by participant and controls

Table 13 compares the average proportion of time that treatment and controls spend in assigned services, over the control group holdout period. We use the Work Search Support service to illustrate the problem of the exposure of participants and controls to different services.

Work Search Support, of all the SDM services, had participants and controls who participated in the widest range of services.

Table 13 compares the average duration that Work Search Support participants and controls spend in assigned SDM services in the 52 weeks from assignment. Participants spend 71.5 percent of the period on Work Search Support, whilst those in the control group spend an equivalent time on GCM.

The other feature of Table 13 is that for services other than GCM and Work Search Support, participants and controls spend nearly the same amount of time on each. For example, Work Search Support treatment and controls spent the same proportion of the control period on WFCM related services (27.6 percent). Such a result was a direct consequence of the randomisation into the participant and control groups that meant the two groups were expected to be eligible for the same services. Any minor differences can be explained by the increased probability of participants exiting benefit under Work Search Support compared to the controls in GCM.

Table 13: Proportion of time assigned to services by treatment and control groups over the control group holdout period

	GCM	WFCM	WSS	IA trial	None	Total
MHES Trial						
Treatment	6.6%	12.4%	0.8%	80.2%		100%
Control	69.1%	30.5%	0.4%			100%
SPES Trial						
Treatment	9.5%	25.3%	16.4%	48.8%		100%
Control	5.7%	72.6%	21.4%	0.3%		100%
WFCM General						
Treatment	0.6%	98.3%	0.7%	0.4%	0.0%	100%
Control	40.7%	2.6%	56.3%	0.5%		100%
WFCM Health						
Treatment	0.3%	98.9%	0.2%	0.6%	0.0%	100%
Control	89.2%	9.2%	0.7%	0.9%		100%
WFCM Intensive						
Treatment	0.6%	95.4%	4.0%			100%
Control	12.3%	59.0%	28.5%	0.2%		100%
WFCM Intensive (Early)						
Treatment	0.4%	95.4%	4.2%			100%
Control	9.9%	70.8%	19.2%	0.2%		100%
WFCM Intensive (Entrenched)						
Treatment	0.7%	95.3%	4.0%			100%
Control	13.5%	53.6%	32.7%	0.2%		100%
WFCM Integrated						
Treatment	0.2%	99.2%	0.5%	0.1%		100%
Control	49.8%	45.4%	4.7%	0.1%		100%
WFCM Integrated (Youth)						
Treatment		99.8%	0.1%	0.1%		100%
Control		100.0%	0.0%	0.0%	0.0%	100%
WFCM pilot						
Treatment	0.5%	96.6%	2.8%	0.2%		100%
Control	14.0%	1.5%	84.3%	0.2%		100%
Work Search Support						
Treatment	0.5%	27.6%	71.5%	0.4%	0.0%	100%
Control	72.0%	27.6%	0.0%	0.4%		100%
Work Search Support pilot						
Treatment	0.1%	12.5%	87.5%			100%
Control	86.8%	13.2%				100%

A further observation from Table 13 is that the control group is not always located in GCM. Examples include WFCM Integrated Support YS, WFCM Intensive Client Support, WFCM Pilot, SPES Trial. Here we see control groups exposed to other intensive case management services.

Estimation of service effectiveness approach

Direct estimation of assigned service impact

We have taken the following approach to resolve the challenge of estimating the impact of services. The first step in the approach was not to use the membership of treatment or control group to estimate the impact of the service. Instead, we directly modelled the changes in assigned service for individuals. The role of the control group in this approach was to provide what is referred to as 'common support'. Common support is the idea that within the analysis we have similar individuals who do and do not receive the service we are evaluating. Without the creation of a control group, we run the risk that there are systematic differences in the types of people in each service that makes it difficult to isolate the effect of the service.
The other reason we can take this approach is that the assignment of individuals to services is fully observed in the analysis. As was outlined in the service allocation process (page 19), service assignment was based on administrative data and prescribed business rules without the discretion of individual staff. Further, the specific data used to allocate individuals was saved to ensure the analysis was based on the same characteristic profile as used in the allocation process.

The fully observed allocation process also allowed us to combine services targeted to the same participation group. In doing so we could increase the level of common support by increasing the number of people similar to the service participants in GCM. We boosted the control group populations with other service treatment and controls for the following:

- WFCM General: include Work Search Support
- WFCM pilot: include Work Search Support pilot
- WFCM Intensive Client Support: include WFCM General same age first benefit and current age
- SPES Trial includes WFCM General sole parents based on the number of children and partner status.

Multi-state models

The second aspect of our approach was to estimate the impact of service on the time on benefit based on multi-state duration models. Here we modelled two consecutive states:

- The hazard of exiting benefit in the period between being assigned to the control or treatment group through to the end of the control period. Hazard rate is the number of people who exit benefit in a given period by those still on benefit at the start of the period.
- For those who exited benefit while in the above state, we estimated the hazard of returning to main benefit. The hazard of returning to benefit is the number of people who return to benefit in a period divided by the number who are off main benefit at the start of the period.

We combined these models using a Monte Carlo simulation to estimate the time to benefit exit while in the service, and for those who do exit the time to benefit return. Taken together we can estimate the overall change in the time people spend on main benefit if assigned to the service compared with being assigned to the counterfactual (which, in all but one case, was GCM).

Impact on time to benefit exit

The first outcome we examine was the time to benefit exit, using duration modelling. For this analysis, the duration to benefit exit was from when a person was first assigned to a service, rather than from when they started a main benefit.

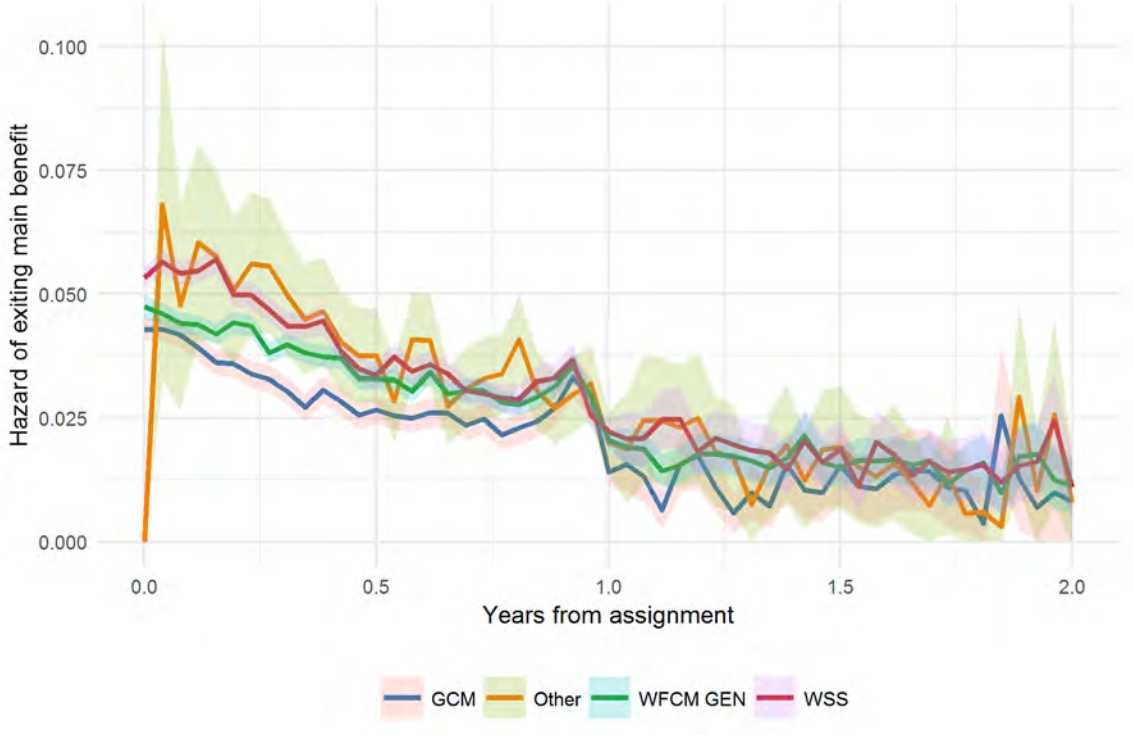
Duration on any main benefit

In this analysis, we combined any consecutive spells on any main benefit (eg Jobseeker Support, Sole Parent Support, Supported Living Payment and retirement-related[14]) into a single spell. In addition, we combined any spells separated by less than 14 days. Any change in partner status had no effect on spell duration, so a person changing from single to a partner will not end their current spell on main benefit.

Comparison of hazard by service for WFCM General analysis

Before examining the model results, we first illustrate how we estimate duration on different outcome states. Figure 23 shows the observed hazard rate for people in the WFCM General and Work Search Support treatment and control groups by assigned service status. In the WFCM GEN model, we have three main services, WFCM General, GCM, and Work Search Support, while people assigned to any other service are grouped under Other. From Figure 23 we can see the hazard of exiting from benefit falls with increased time after assignment. This downward trend in hazard of exiting benefit is a common feature of people on income support.

Figure 23: An observed hazard of exiting main benefit for WFCM General by assigned service



Hazard rate: the number of exits in the interval divided by the number who remain at the start of the interval
Confidence interval: shaded areas show the 95% confidence interval for the estimate.

[14] New Zealand Superannuation and Veteran's Benefits.

When we compare the hazard rates between the four service states, we can clearly see the lower hazard rate for those assigned to GCM, relative to the more intensive services. However, because we know there are also differences in the make-up of people assigned to each group, we cannot determine relative effectiveness from this descriptive analysis alone.

Estimation of the impact of service assignment

To control for the profile of people assigned to different services we used a logistic regression model (SAS PROC LOGISTIC) to estimate the influence of people's observed characteristics at service assignment as well as the service they were assigned to. Here we rely on the two features of the allocation process that enables us to provide an unbiased estimate of service assignment. Namely:

- service assignment is based on pre-defined criteria, and not subject to the discretion of frontline staff or participants
- the criteria for assigning people to each service are fully observed.

Alongside service status, we included all the information used to determine service eligibility and priority for selection into the model. If a person was assigned to a new service, then their profile was updated to that at assignment to the new service. Apart from service assignment, duration events and calendar time-related variables, the profile characteristics remain constant within the model.

Definition of the hazard function

We identified when a person starts and ends a benefit spell. For analytical simplicity, we converted these continuous time units into *k* discrete intervals of 14 days. Therefore, the approach we took to modelling the data is an example of 'discrete-time' survival analysis (Allison, 1984).

For each individual *i* at interval *t* we have a status y_{it} = 1 if the individual ended their benefit spell in the interval, and 0 if the individual was still in the spell.

The hazard function (h_{it}) for an individual at each interval is:

$$h_{it} = \Pr(y_{it} = 1 \mid y_{ik} = 0 \text{, for all } k \in \{1,....,t_{-1}\} \quad (1)$$

That is the probability an individual will end their spell at interval *t* given their spell has lasted for *t*-1 intervals. Related to the hazard function is the survival function S_{it}, which is the probability that the individual *i* was still on benefit at the end of interval *t*.

$$S_{it} = \Pr(y_{ik} = 0) \text{ for all } k \in \{1,....,t\}) \quad (2)$$

The survival function is the product of terms involving the hazard.

In our model, the hazard function was assumed to relate to the explanatory variables through a logit transformation:

$$\log\left(\frac{h_{it}}{1-h_{it}}\right) = \alpha(t) + \beta' X_{it} \quad (3)$$

where $\alpha(t)$ is the baseline hazard and X_{it} is a vector of covariates representing the values for an individual *i* at interval *t*. The variables for X_{it} are summarised below in the section on explanatory variables.

As the previous equation shows, each model had two parts, the baseline hazard and fixed and time-varying individual characteristics. The baseline hazard characterises the overall pattern of exits over spell duration. Individual characteristics, on the other hand, identify how an individual's probability of ending a spell varies according to their fixed characteristics (eg age or education) as well as those that vary over time (eg assigned service and time-varying characteristics at assignment date).

Estimation of model parameters

Allison (1984) shows how the parameters of the discrete-time hazard model in equation (3) can be estimated using standard logistic regression procedures, after restructuring the data so there is one record for each period that each person is at risk of experiencing the event of interest (ie constructing a 'person-period dataset').

Selecting analysis sample

Because people have multiple service spell starts, we needed to select a single spell for each person. The selection was a two-stage process. First, we split the population into the control and treatment group. All control group members are included in the analysis. If a control group member was selected more than once, then we randomly selected one of the service starts and removed the remainder. For the treatment group, we needed to take a subset to make the analysis manageable. We first randomly selected one service start for each person. From the dataset of individual treatment group members, we took a simple random sample to make a total sample of 30,000 individuals. We created a weight variable to account for the sampling of the treatment group.

All the models are based on the same sample of individuals for each service.

Accounting for overfitting

We took an initial participant sample for developing the models included in the analysis. Once the model fitting was completed, we re-selected and updated the control and treatment samples as described in above. Alongside including more recent observations into the analysis, the goal of re-selecting the treatment group sample is to reduce the problem of over-fitting the analysis to a specific sample of the analysis population.

Creating a person-period dataset

For each spell, we created a 'person-period' dataset by dividing each spell into intervals of 14 days. Each interval can have one of three statuses:

- no exit (the spell did not end within the interval and the interval end date is less than the censor date)
- exit (the spell ended during the interval and the interval end date is less than the censor date)
- censored (the spell end date is greater than the censor date).

Censoring occurs either because:

- the spell is current to the end of the analysis period (July 2017)
- the spell duration exceeded the control period for the service.

Therefore, censor dates are either the end of the analysis period or the end date of the control period. Time-varying characteristics are calculated at the start of each interval. There were two types of time-varying characteristics.

Service allocation variables

Service allocation variables describe the eligibility and priority criteria for people to be assigned to each service. In the model, these variables only change in the interval a person was assigned to a new service. During the periods within services, these variables were not changed. The reasoning here is that we are primarily interested in accounting for observed selection effects, rather than in modelling the influence of these characteristics on outcomes more generally.

Calendar time and annual reapplications

The second group of time-varying characteristics reference calendar month and annual benefit replications. These variables did vary with each interval as these have important effects on the hazard but do not influence service assignment.

Defining interval duration

The selection of a discrete interval model requires the selection of the duration of the interval. The selection of the duration of the interval in the analysis is a trade-off between identifying detailed events and computational resources. We selected a 14-day interval since it provided enough detail on patterns of exits over the duration people were on the service.

Estimating hazard of exiting main benefit

Model estimates

Figure 24 summarises the beta estimates of the model for the first benefit spell for WFCM General. A beta estimate greater than 0 indicates increased odds of exiting main benefit, while the width of the lines shows the 95 percent confidence intervals for each beta estimate.

We binned continuous variables into categorical variables with multiple levels. The reason for binning was two-fold. The first was that some continuous variables had null values that are not equivalent to 0 (ie a statistical risk score where null means no score was available, while 0 indicated a low risk). The second reason was that many continuous variables had a non-linear relationship with the dependent variable.

Prior to running the main model for a given service, each binned continuous or ordinal variable had consecutive levels combined when there was no significant difference between the beta estimates.[15] The motivation here was to reduce the number of parameters to estimate. In the model fitting, we undertook backward selection of significant variables. Non-significant variables were excluded from the final model estimation.

Consistent with previous work, the most important predictors of benefit exit were:

- current benefit duration (DurIntCat)
- statistical risk profiling tools (LLTBRcat, LETageScat)
- benefit group (ClntGrp).

[15] The binning of ordinal variables was based on a separate sample from the treatment group.

The target predictor variable of interest (service_dur) also showed a strong relationship between the type of service and duration of service and the hazard of exiting from main benefit.

The pattern of beta estimates follows what we would expect to see based on prior evidence on factors that influence the hazard of exiting main benefit.

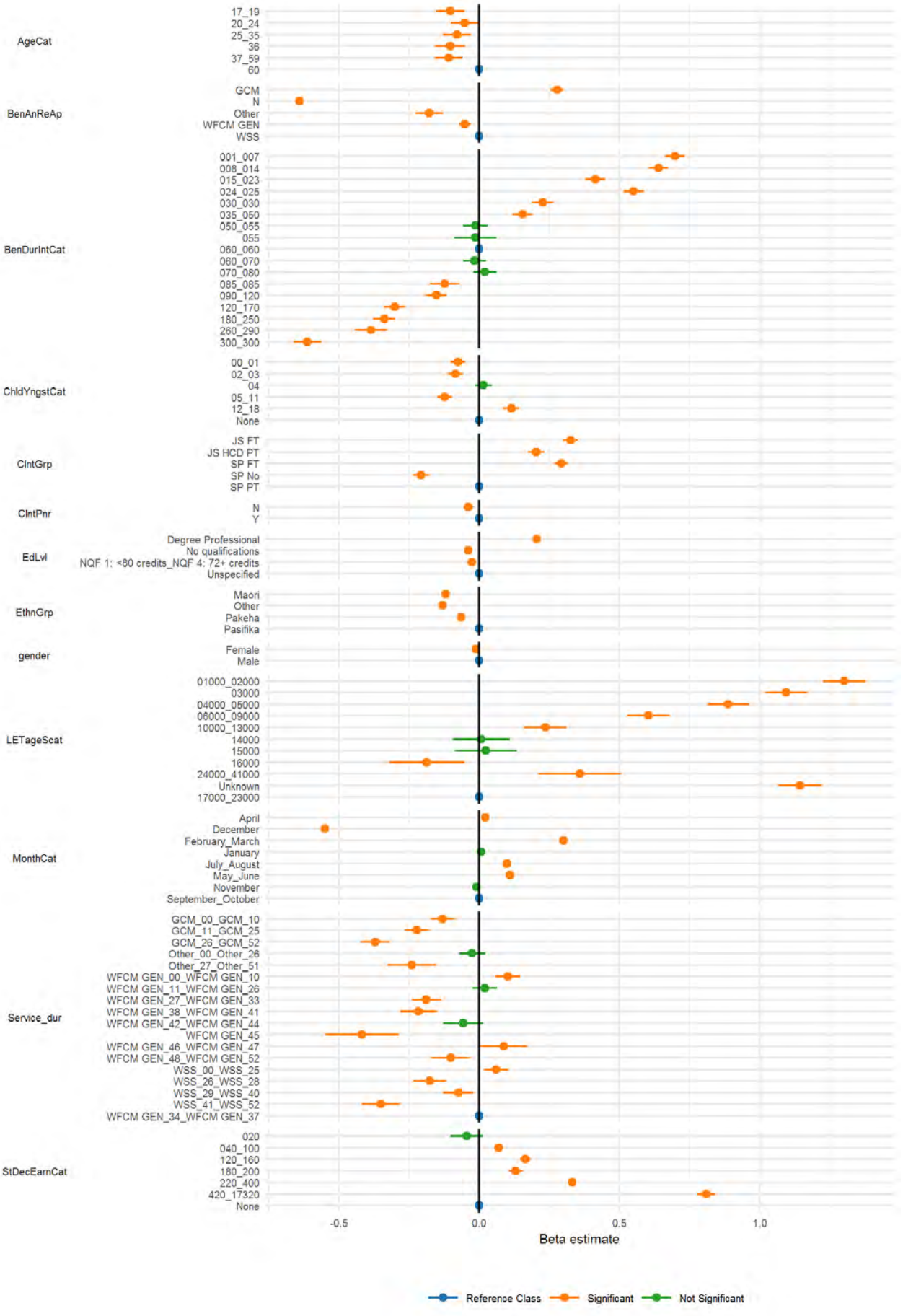
Benefit duration: the longer a person had been on benefit at assignment to a service, the lower their odds of exiting.

Client groups: sole parents with young children have lower odds of exiting than sole parents with older children.

Predicted future income support receipt: the scores from the statistical risk profiling tool (LET) shows a clear relationship to exits, with the odds of exiting decreasing with expected future income support costs.

Part-time work: people declaring higher income from part-time work show higher odds of exiting benefit.

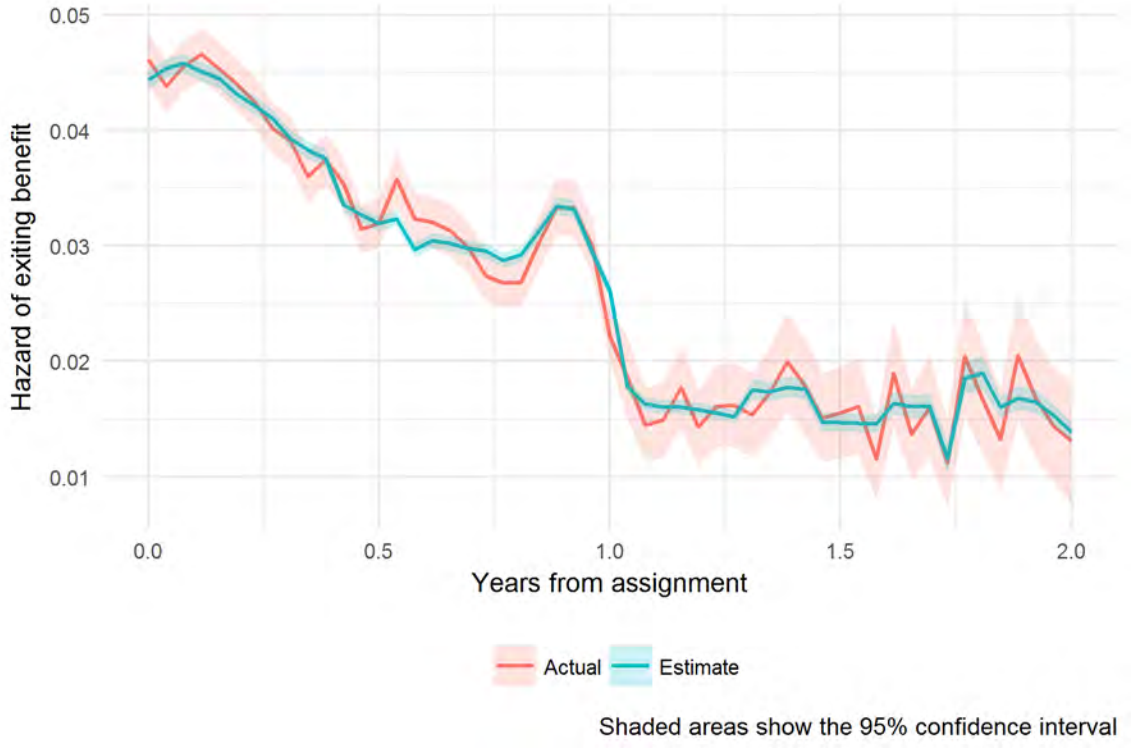
Figure 24: First main benefit spell WFCM General model beta estimates



Estimated and actual hazard

Based on the model predicted and observed hazards, we can compare how well the model is able to predict the observed hazard rate. Figure 25 compares the actual and predicted hazard of exiting main benefit while people are assigned to WFCM General. Overall, the predicted hazard rate lies within the expected range based on the observed hazard.

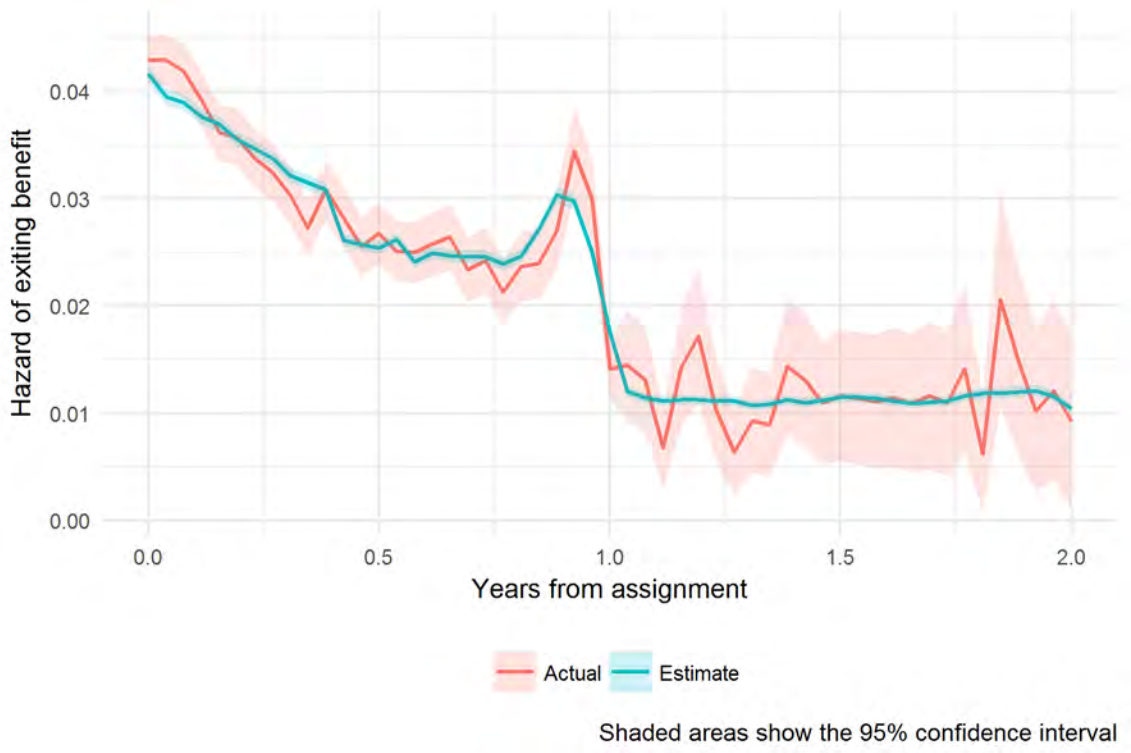
Figure 25: Observed and predicted hazard rate while assigned to the WFCM General service



The increase in the hazard rate at just before one year after assignment corresponds to the requirement to reapply for benefit every 52 weeks. If a person fails to reapply at this time, their benefit is automatically cancelled.

Figure 26 shows the same predicted and actual hazard rates for people while they are in GCM. We see the same overall pattern, with the predicted hazard falling within the expected range of the observed hazard rate. One difference between the WFCM General and GCM status is that the confidence intervals for observed hazard in GCM are larger after one year than for WFCM General. This comes about because the number of people remaining in GCM for two years is much lower than for WFCM General.

Figure 26: Observed and predicted hazard rate while in GCM service

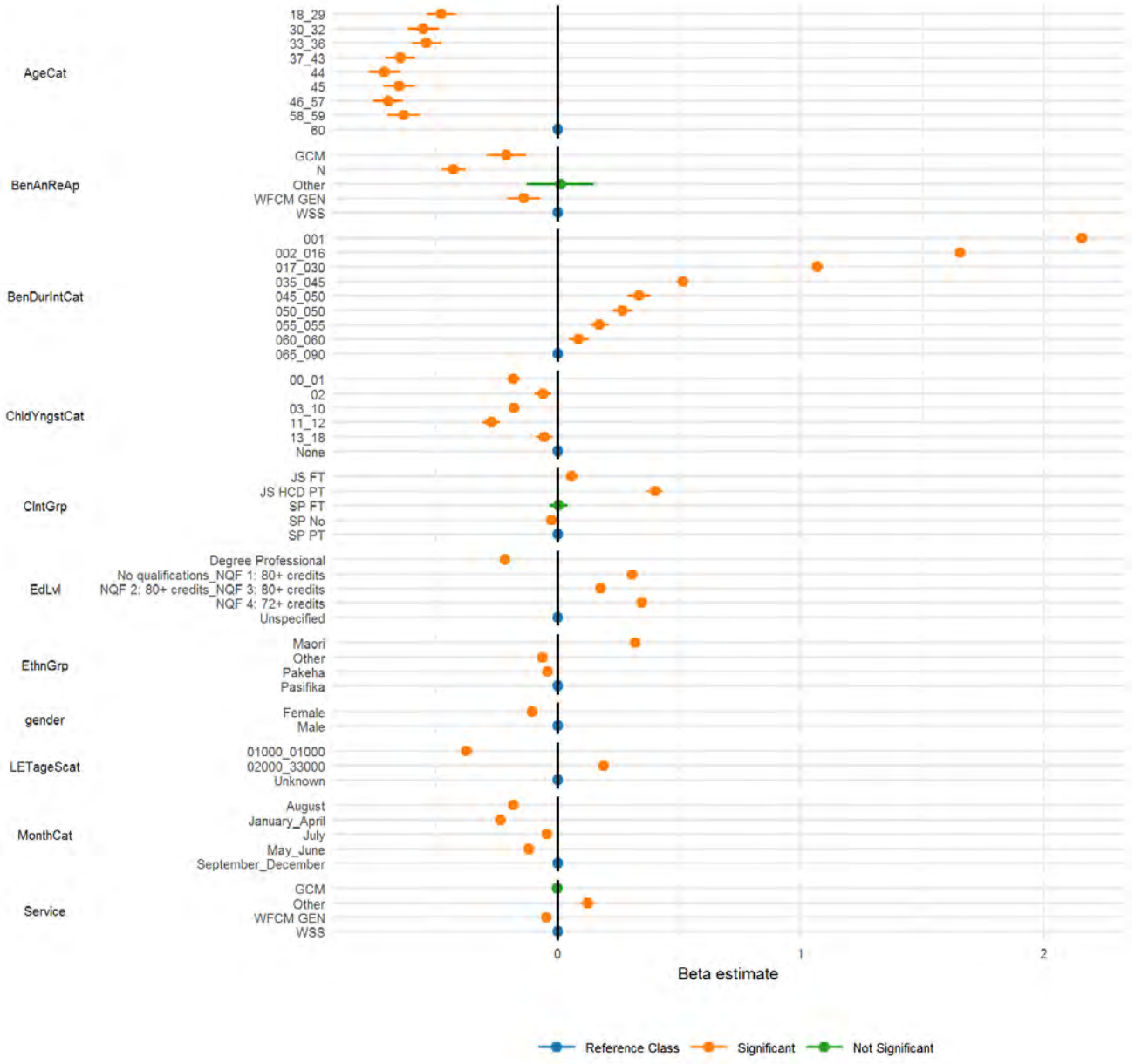


Estimating hazard of returning to main benefit

Model estimates

The models that estimate the hazard of returning to main benefit have fewer predictor variables than the ones that estimate the hazard of exiting main benefit. Figure 27 summarises the model betas for the WFCM General off benefit spell model. The hazard of returning to main benefit is strongly duration dependent, with the immediate period after exiting benefit associated with increased odds of returning (BenDurIntCat). Conversely, those with lower risk scores (LETageSCat) have lower odds of returning to benefit. What service a person was on at exit has a comparatively modest role to play, with WFCM GEN decreasing the odds of returning to benefit overall.

Figure 27: Model betas for WFCM General hazard of returning to main benefit



Estimated and actual hazard

The following two figures plot the observed and estimated hazard of returning to main benefit between whether a person exits while on WFCM General (Figure 28) or while on GCM (Figure 29). There is little difference between the two figures, and both indicate that the estimated hazard does not closely follow the observed hazard over the initial 1.5 years after assignment. The reason for this lack of precision is the relatively small number of observations in each case.

Figure 28: Actual and estimated hazard of returning to main benefit when exiting from the WFCM GEN service

Figure 29: Actual and estimated hazard of returning to main benefit when exiting from the GCM service

The hazard of transferring to other services

To provide a realistic estimate of the impact of case management services we also need to account for the transfer of participants to other services. The transfer rate to other services varies by service type. The most affected service was Work Search Support, as many people in Work Search Support are also eligible for WFCM General. Because WFCM General was a higher priority service during this analysis, we found an elevated level of attrition. To account for transfers, we ran a hazard model on transferring to another service conditional on being on main benefit. In other words, if a person exited the service for other reasons (ie exit from benefit) these spells were censored. The reason for taking this approach is so we could separately estimate the hazard of exiting benefit and hazard of transferring to other services. Figure 30 summarises the beta estimates for the hazard of transferring out of WFCM General and Figure 30 shows the hazard.

Figure 30: Model betas for WFCM General hazard of transferring out of service

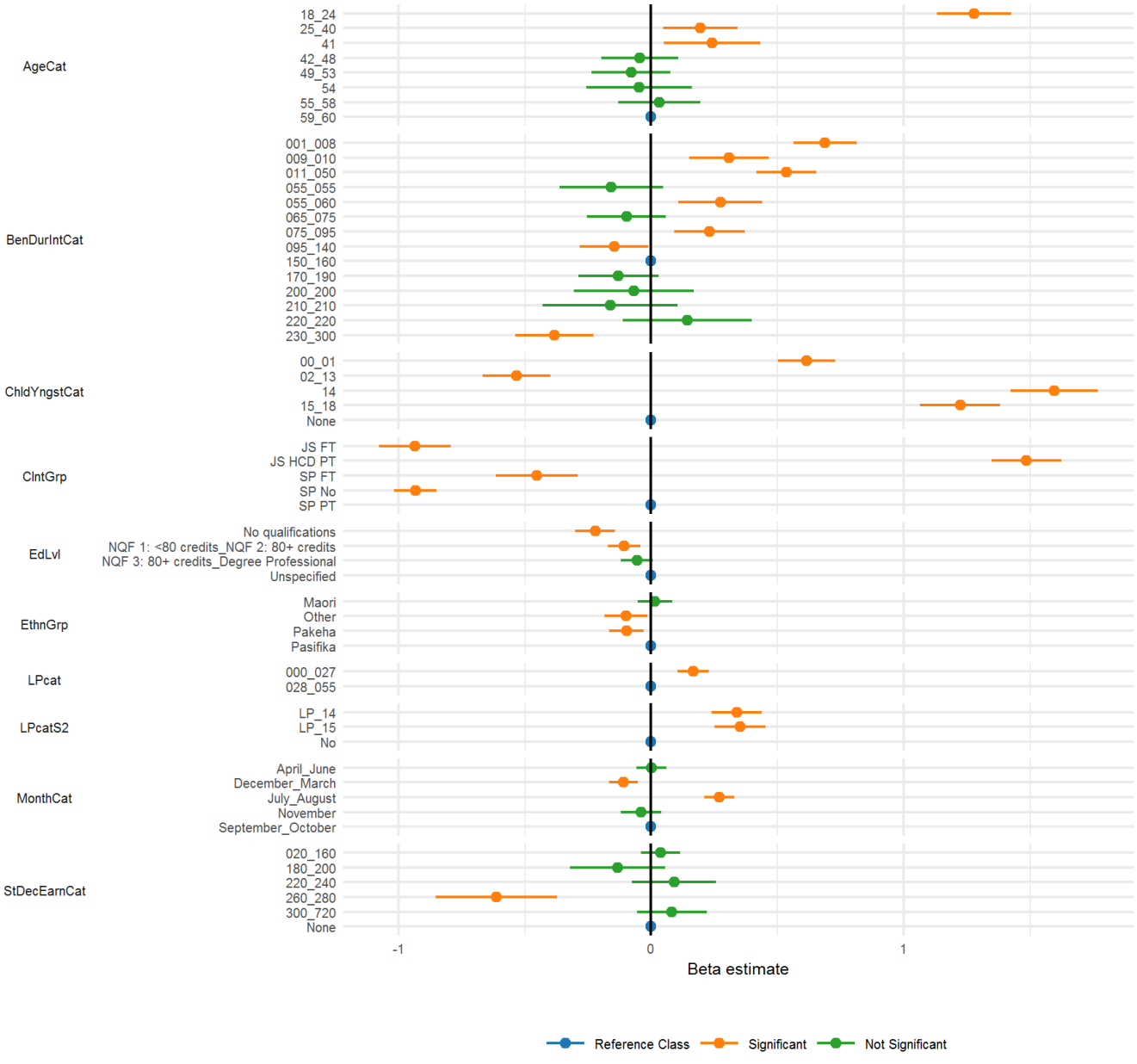
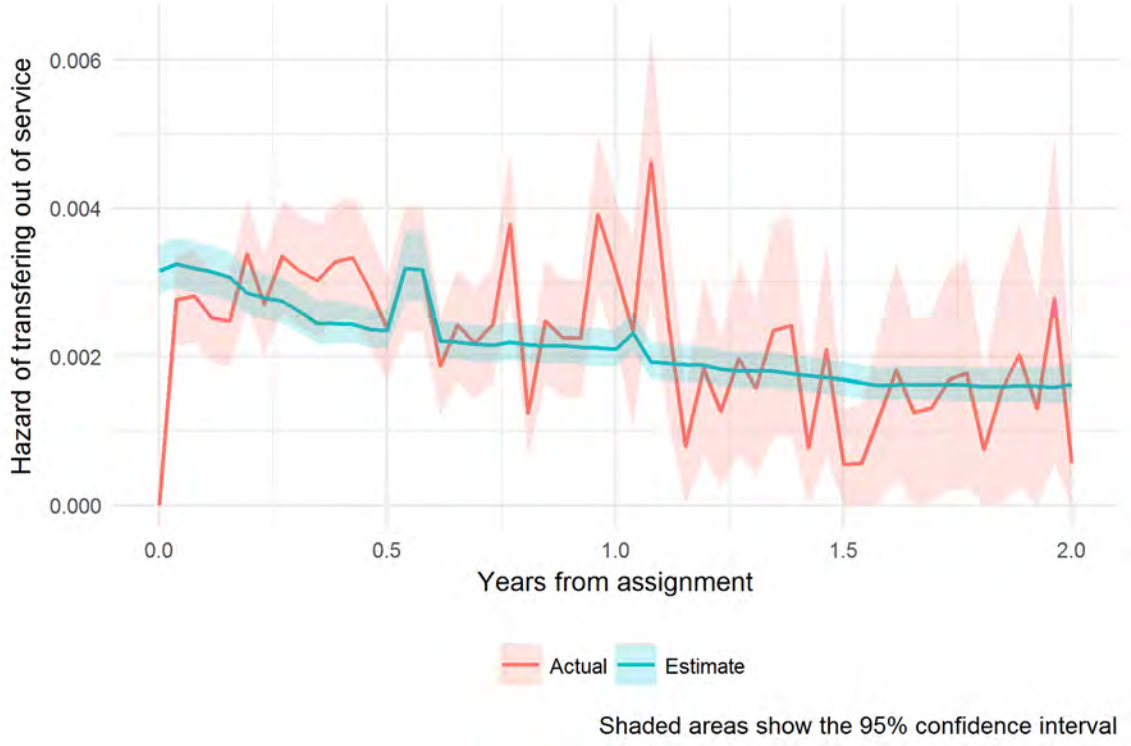


Figure 31: Actual and estimated hazard of transferring out of WFCM General



Income support payment rate

The final two models included in the analysis show the estimated level of income support paid on and off main benefit. These models were estimated using SAS GENMOD using linear regression with a normal distribution for the error term. Figure 32 shows the model estimate for the daily rate of income support payments while on main benefit. Of note is the strong relationship between age-adjusted LET score and daily rate, as well as the reverse relationship for the level of declared income as higher earnings abate income support payments.

Figure 32: Model betas for WFCM General daily income support payment rate while on main benefit

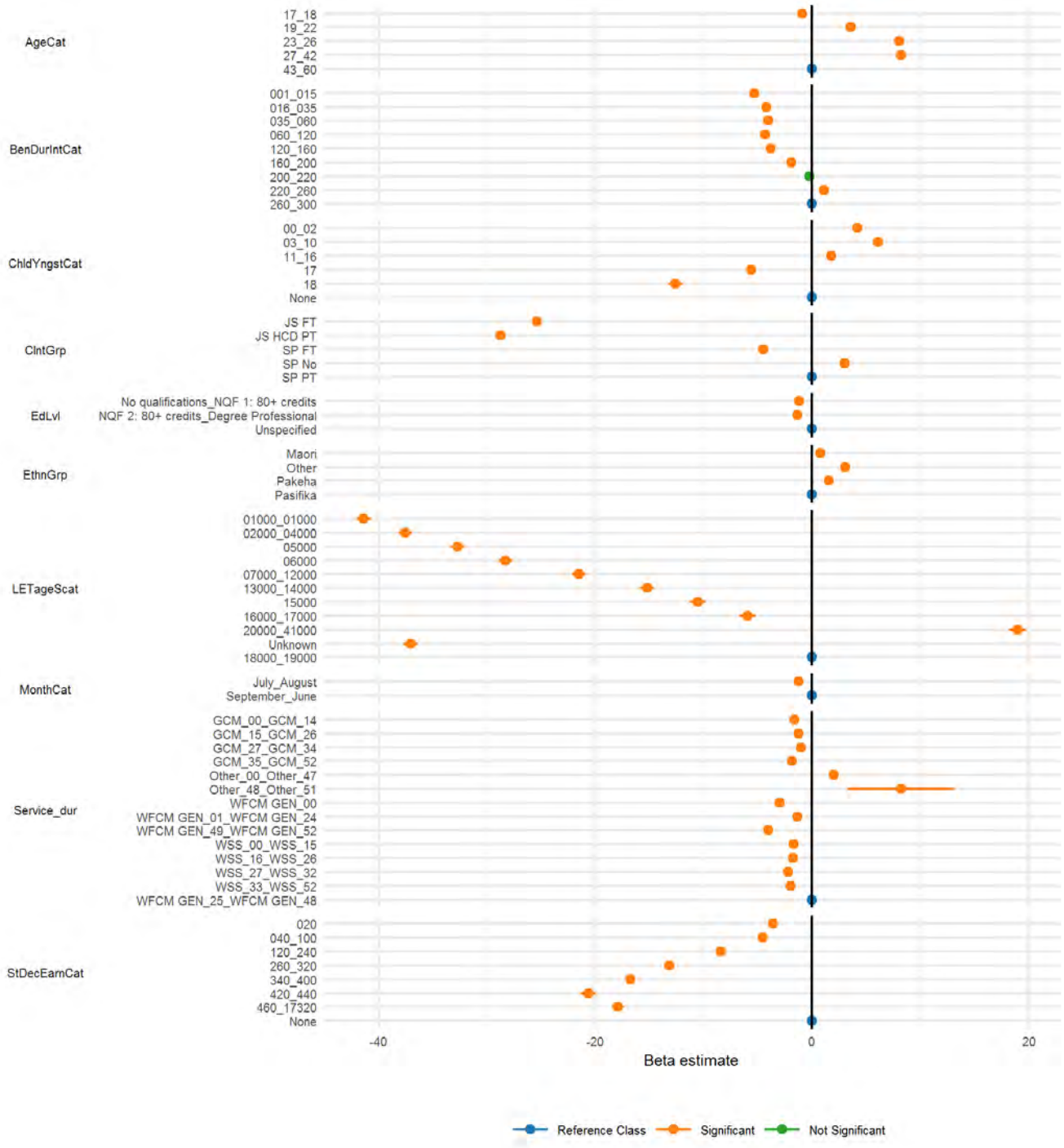


Figure 33 tracks the estimated and actual income support rate for WFCM General participants in the two years after assignment. Overall, the predicted and actual follow each other well. The increase in the rate at one year is likely to result from movement off benefit by jobseekers subject to the reapplication process. We are less sure of the dip just before two years duration.

Figure 33: Actual and estimated daily on-benefit income support rates for WFCM General

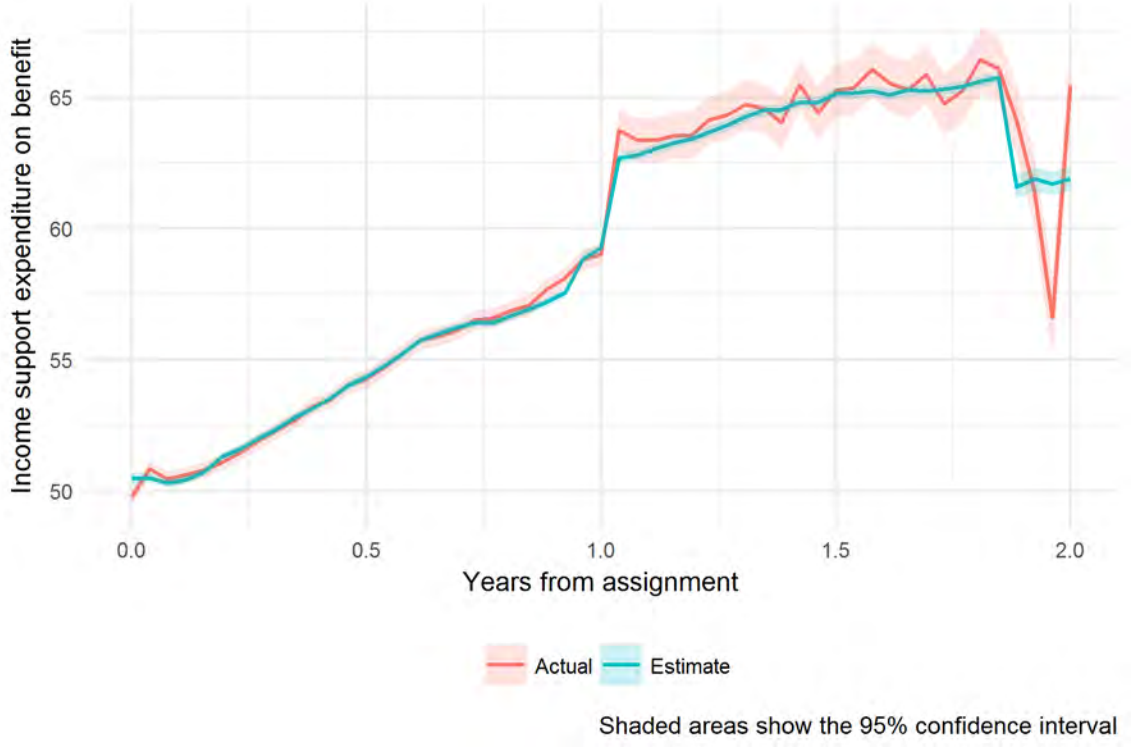


Figure 34 summarises the beta estimates of the off-benefit income support payment rate model. For all off-benefit income support models, there were relatively few variables that influenced the income support rate when off main benefit. The predicated and estimate values are shown in Figure 35 below.

Figure 34: Model betas for WFCM General daily income support payment rate while off main benefit

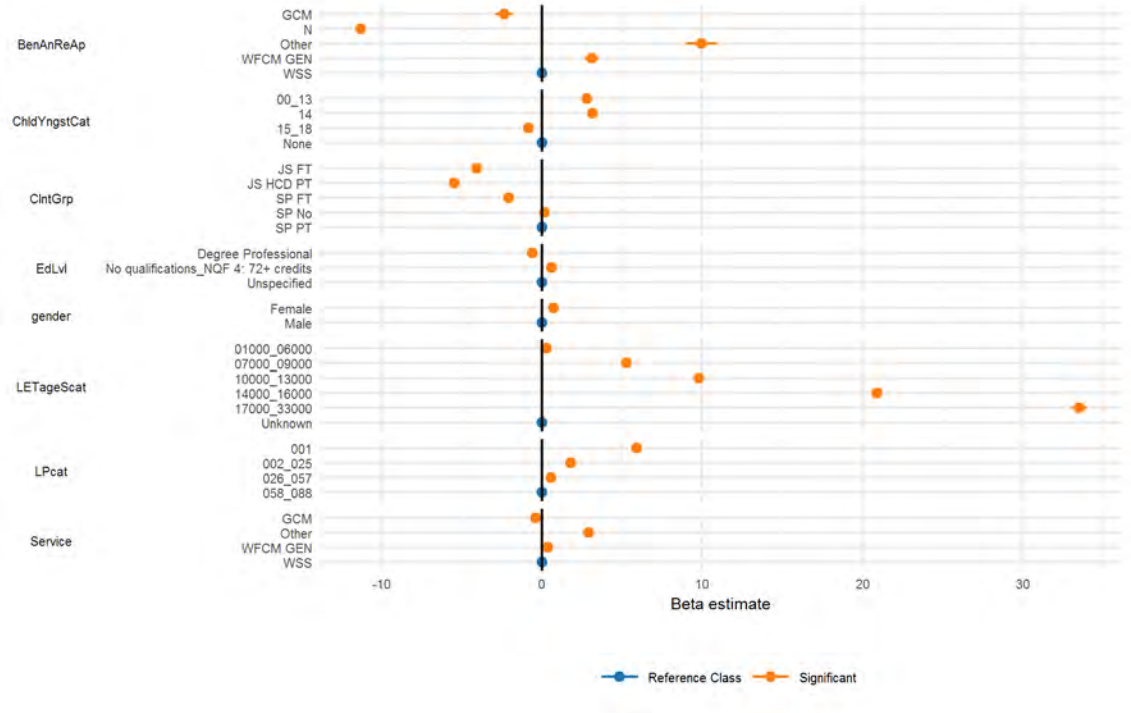
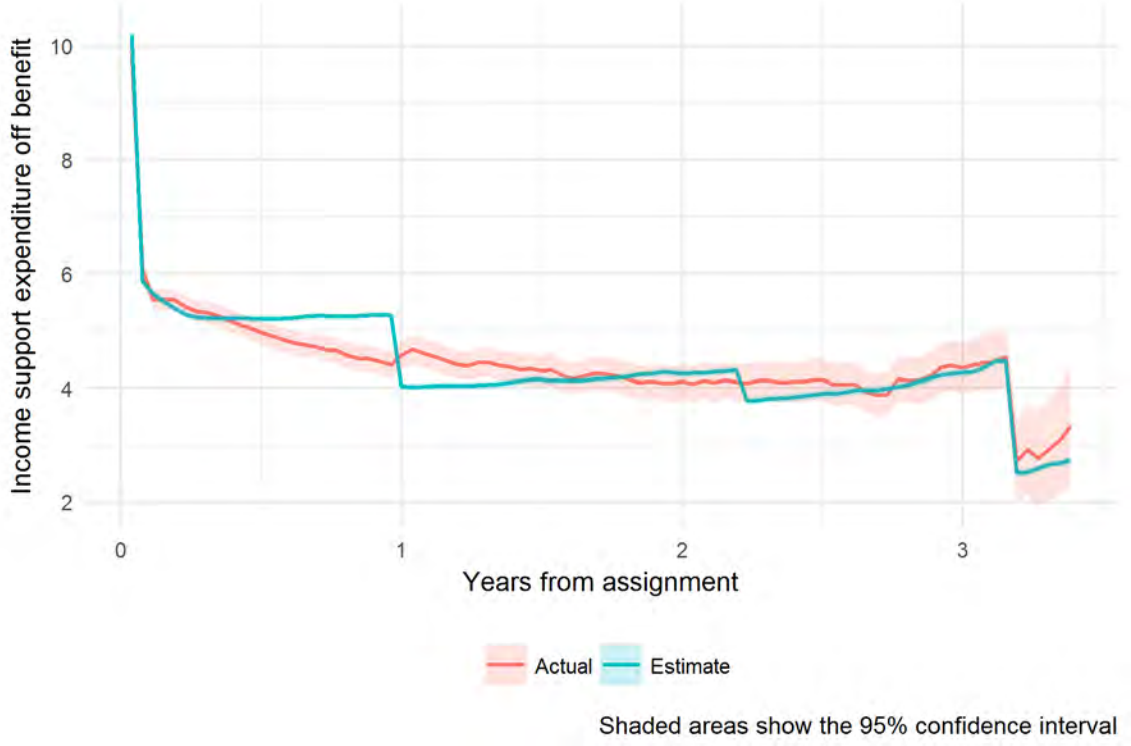


Figure 35: Actual and estimated daily off-benefit income support rate for WFCM General



Simulating outcomes between service and counterfactual scenario

Having established that our models provide an adequate representation of the observed outcomes of people in each service and comparison group, the next step was to use the model results to examine the difference that each service made to these outcomes.

To achieve this, we ran a simulation that compared the estimated experience of service participants in two states:

- Initial service: the simulated hazard is where people remain in their initially assigned service and is designed to show what happens if people remain in each service for the whole of the follow-up period.
- Counterfactual: people are assigned to GCM instead of their initial service. The GCM counterfactual represents the baseline or counterfactual state. In other words, in the absence of more intensive case management, such as WFCM General, people will at least receive GCM.

Both states are estimated using a person's profile at the start of each service spell and the parameter estimates for the models described above. For each simulated scenario, we estimate the expected outcomes if the group had been assigned to the service (eg WFCM General). We then re-run the same simulation but change the group's service status to the counterfactual (eg for WFCM General, this would have been GCM). All other characteristics and model parameters remained unchanged.

For each simulation run we undertook the following steps:

1. Estimate the duration on main benefit while on the service

The first stage was to estimate the following outcomes for each interval after being assigned to a service:

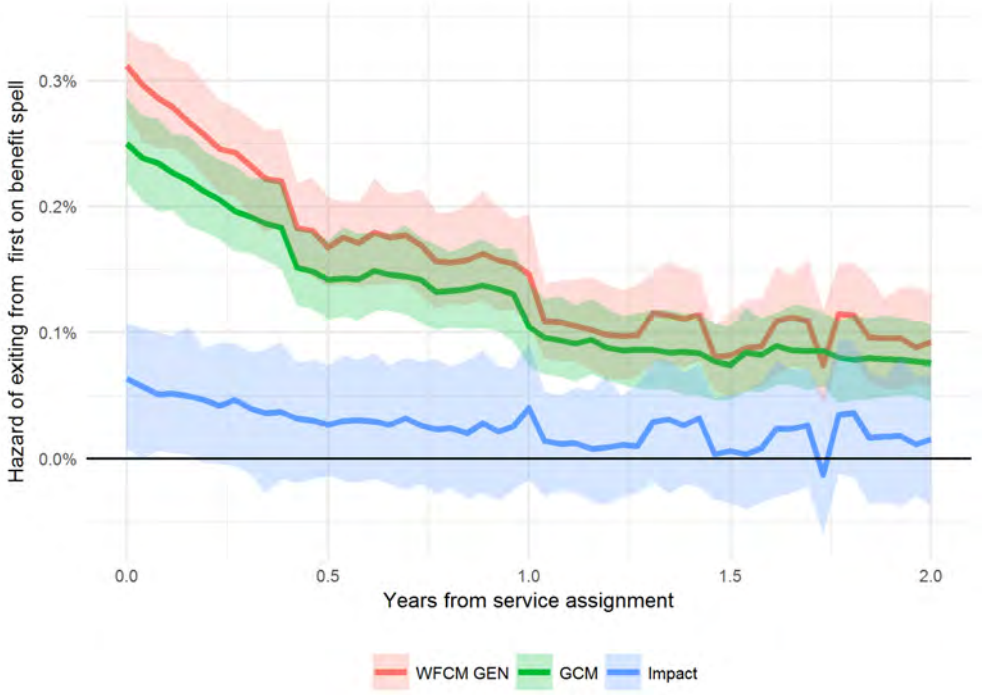
- The hazard of exiting main benefit.
- The hazard of being transferred to another service.
- Expected income support payment rate.
- The average amount of staff time, staff costs, employment subsidy and contract payments based on participant profile and service.

For each lapse period, we take two independent random draws from a uniform distribution between 0 and 1. If the draw for exit was less than the hazard of exiting benefit in the lapse period, then the person was assigned a benefit exit at that lapse period. The second random draw was for the hazard of being transferred out of the service. Again, if the random draw was lower than the transfer hazard rate, then the person transferred out of the service in the interval. If both exit and transfer occurred for the same lapse interval, we randomly selected one or the other as the event for that lapse period. In other words, a person could not exit and transfer within the model. Once a person transferred out of the service or reached the end of the follow-up period, their record was right censored.

The hazard of exiting main benefit while on the service

Figure 36 shows the results of the simulation on the hazard of exiting from main benefit after being assigned to either WFCM General or GCM. How the confidence interval was calculated will be explained below. From the hazard rates we can see that for the same group assigned to WFCM General, the models estimate a higher hazard rate than if the same group had been on GCM, at least initially. As duration on the service increases, the difference in the hazard rate begins to narrow, indicating the impact of WFCM General over GCM is diminishing.

Figure 36: Modelled hazard of exiting benefit under WFCM General and GCM scenarios



The hazard of transferring out of service

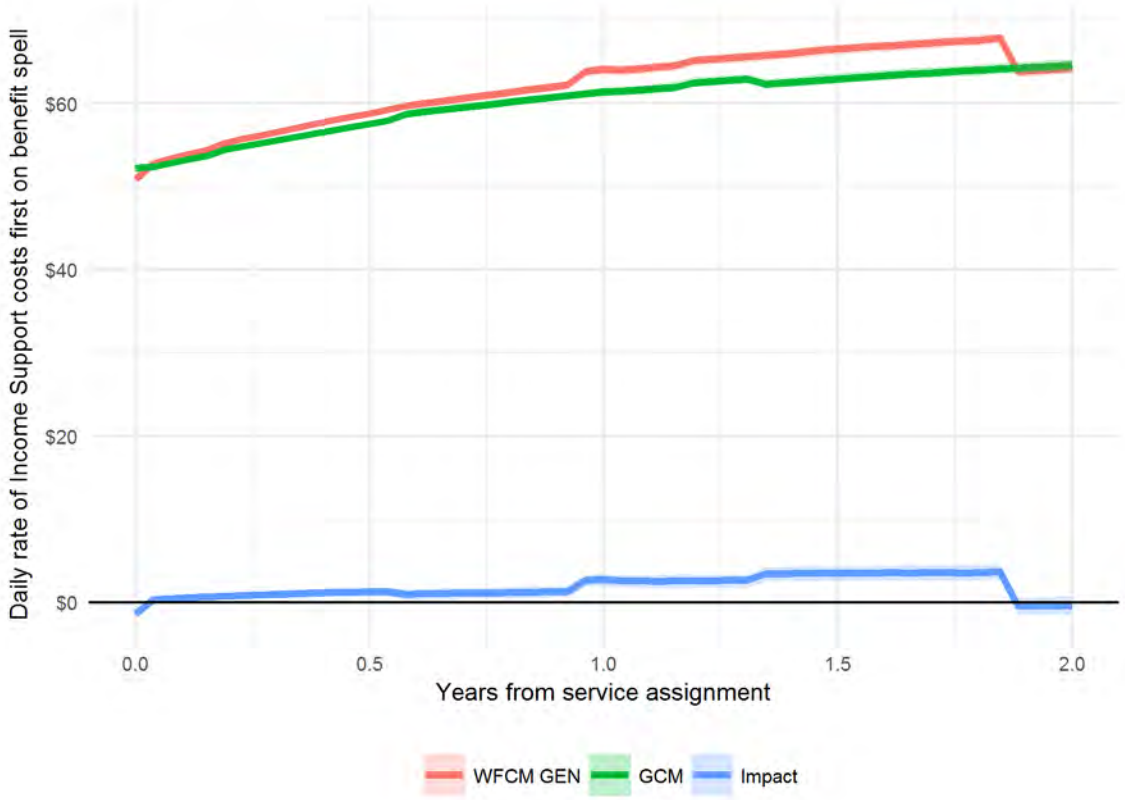
Figure 37 shows the hazard of transferring to another service. In the simulations, we used the same expected transfer hazard for both services as we wanted to standardise the probability of transferring out the service between the two scenarios. In the case of WFCM General, the transfer hazard rate is very low, an order of magnitude less than the probability of exiting from main benefit while on the service.

Figure 37: Modelled hazard of transferring to another service WFCM General and GCM scenarios

Income support payment rate

While on the service we calculated the average income support rate between the two scenarios. In most cases, there was no material difference in the daily rate of income support payment between the service and the counterfactual. But for WFCM General, it appears that being assigned to the service results in a slight increase in a participant's daily rate over being in GCM (Figure 38).

Figure 38: Modelled daily on main benefit income support payment rate for WFCM General and GCM scenarios



Case management and employment costs

In addition to estimating the above outcomes, we also included the cost of case management and employment assistance. These values were based on the average values for each lapse period controlling for some basic client attributes of main benefit and age (over or under 24). In subsequent versions of the analysis, we plan to include these outcomes through model estimates.

Figure 39 tracks the weekly rate of all service delivery costs while on the service (in real dollars and excluding indirect costs). Overall, participating in WFCM General results in a higher average rate of assistance than for GCM, peaking at just over $10 a week. The difference between the two scenarios decreases over time.

Figure 39: Estimated daily case management costs for WFCM General and GCM scenarios

2. Estimating outcomes while off main benefit

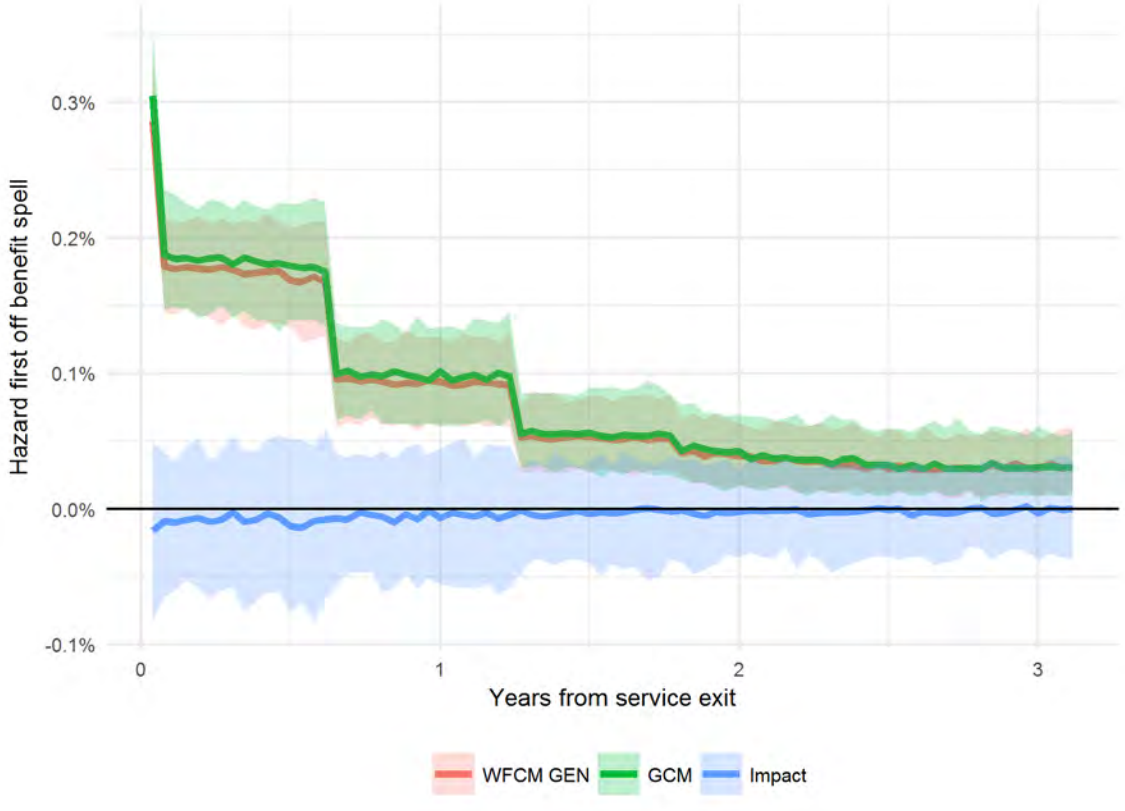
The second step in the process was to estimate the outcomes for those individuals who had exited from main benefit while on the service in the initial simulation run. What this means is that the population who exit will change with each simulation run. We used the same profile information the person had when assigned to a service, except for duration on benefit at exit. Duration on benefit at exit was adjusted according to the lapse period if the individual exited benefit in the initial model. The only time-varying characteristics included in the off-benefit models relate to calendar time (ie month).

Across the services, we found fewer differences in the off-benefit outcomes of participants between the service or counterfactual service. Only for some services were participants found to be likely to return to main benefit based on which service they were on.

The hazard of returning to main benefit

Figure 40 summarises the estimated hazard of returning to main benefit after exit. Common to all services, the hazard of returning is highest immediately after exit before falling at a steady rate with duration off benefit. For WFCM General we found a small estimated reduction in the hazard of returning to main benefit in the first year after exit.

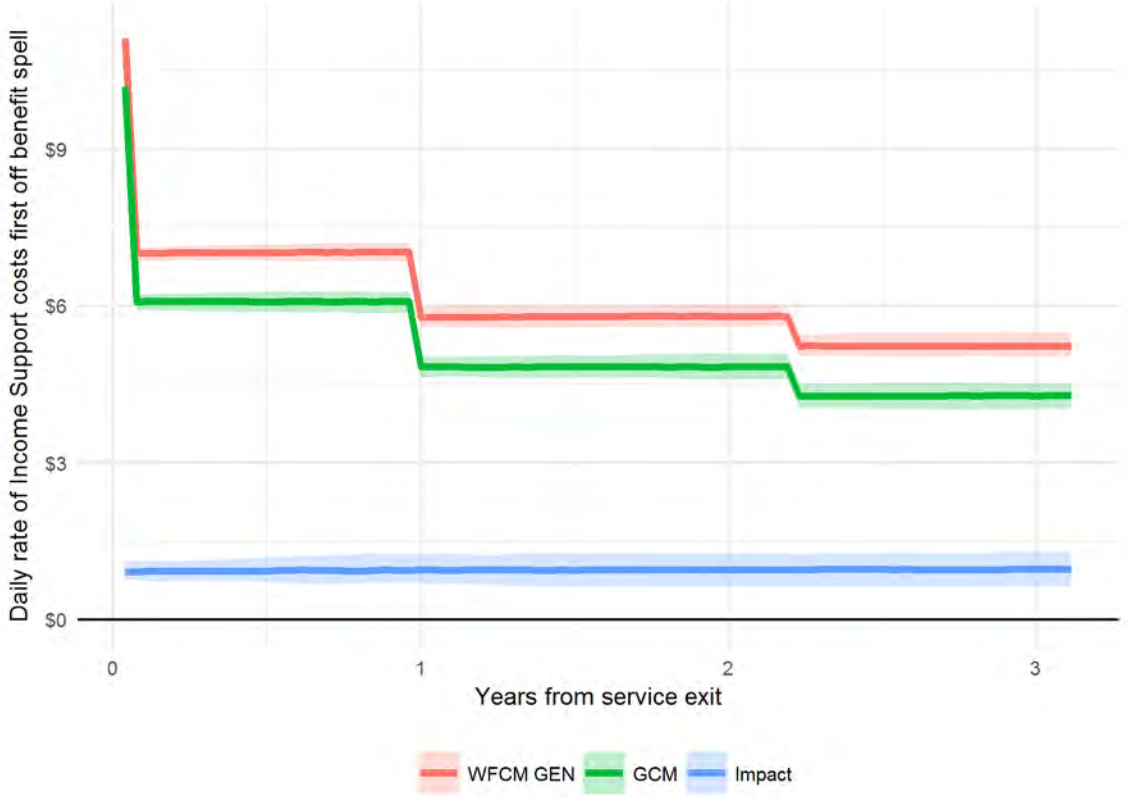
Figure 40: Estimated hazard of returning to main benefit for WFCM General and GCM scenarios



Income support payment rate

While people are off main benefit they can still receive income support such as Accommodation Supplement. Figure 41 tracks the daily income support rate for people who exited main benefit from WFCM General or GCM. It appears that, as for main benefit income support rates, people who exit from WFCM General receive a higher daily rate than GCM. However, the difference is modest at just $1 a day.

Figure 41: Modelled daily off main benefit income support payment rate for WFCM General and GCM scenarios



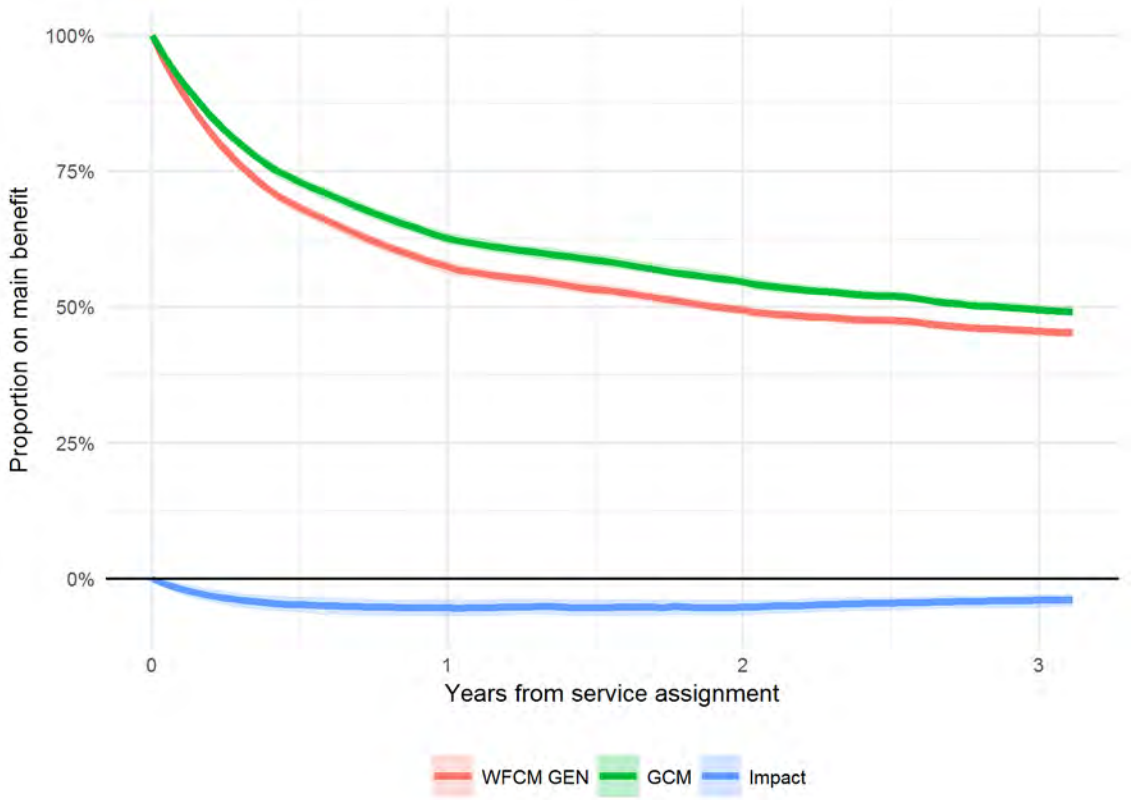
3. Combined on and off main benefit spells

The last step in the simulation was to combine the on and off main benefit spells into a single lapse period history. For those lapse periods where either a person had remained on main benefit for the whole of the service period, had transferred to another service or had returned to main benefit we assume they experience the same outcomes irrespective of the scenario. Therefore, any difference in their combined on and off benefit outcomes was because of the impact of service on the first on and off benefit spells only.

Time on main benefit

Figure 42 summarises the estimated proportion of time on main benefit in each lapse period after assignment if participants were assigned to WFCM General or GCM. The difference between WFCM General and GCM is driven by the difference in hazard of exiting main benefit (Figure 36) and the hazard of returning to main benefit (Figure 40). The impact on the probability of being off main benefit is largest in the first year after starting the WFCM General. In the follow-up period, we see the gradual reduction in the difference in outcomes between the two scenarios. Over the long term, the impact will stabilise at a constant rate.

Figure 42: Modelled proportion of time on main benefit for WFCM General and GCM scenarios



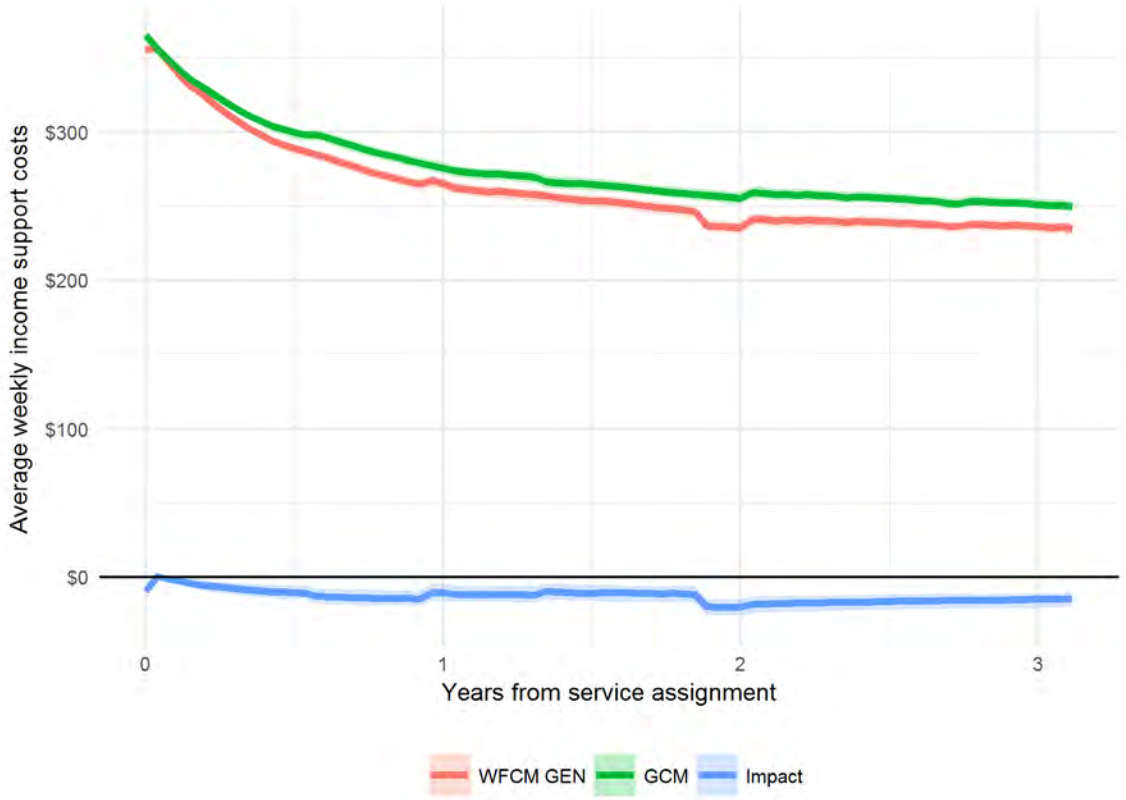
Income support costs

Using the time on main benefit each month, we can multiply the time on benefit by the income support daily rate for both the on and off main benefit period. Figure 43 tracks the total weekly cost for each service scenario. As expected, the income support cost chart closely follows that of the proportion on main benefit.

We then also calculated the total case management assistance costs based on the average weekly expenditure and the number of people who remain in each service.

We can convert these results into cumulative estimates (ie the total time on benefit or income support costs since assignment to WFCM General or GCM).

Figure 43: Modelled income support costs for WFCM General and GCM scenarios



Calculation of net-return and return on investment

Based on the simulation results outlined above, we now turn to the calculation of the net-return and the overall return on investment (ROI). The calculation of the ROI is based on the following equation:

$$wROI = \frac{((IS_c - IS_s) + CM_c)}{CM_s}$$

Where:

wROI: welfare Return on Investment

IS_s: estimated income support payments in the service scenario

IS_c: estimated income support payments in the counterfactual service scenario

CM_s: estimated case management and employment assistance costs in the service scenario (ie the gross investment cost)

CM_c: estimated case management and employment assistance costs in the counterfactual service scenario.

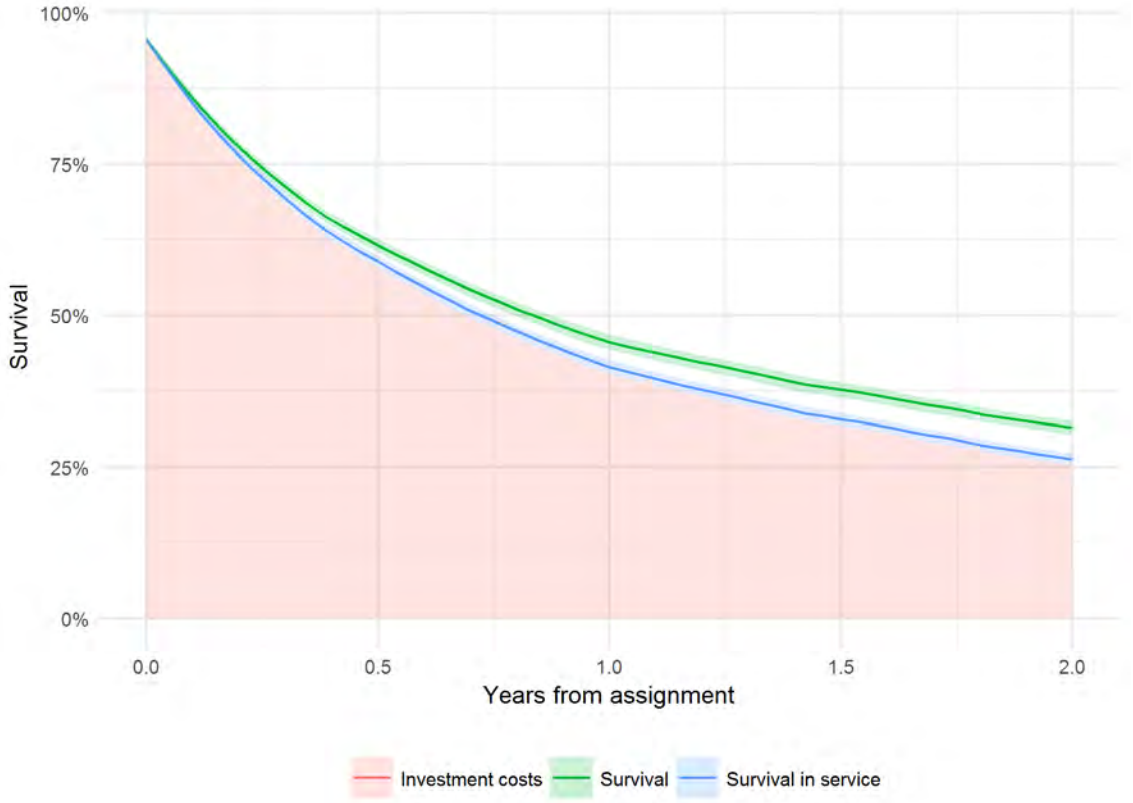
A wROI value over 1 indicates the net-return ($(IS_c - IS_s) + CM_c$) exceeds the gross investment cost of the service.

Gross and net investment costs

The first step to estimating the wROI was to estimate the gross investment of assigning a person to a service and the net-investment by subtracting the investment if the person had been assigned to GCM instead. While the net-investment is not used in the wROI calculation itself, it is a useful metric to understand the level change in investment between the service and the counterfactual. If this change is relatively small, then this would raise questions about how successful the service assignment process was in shifting investment towards the target group.

Investment costs are included only during the first benefit spell and while the individuals remain on in the service initially assigned to. Duration on the service is therefore estimated through the hazard of both exiting benefit (Figure 36) and transferring to another service (Figure 37). Taken together we can estimate the expected duration that individuals remain on the service and the baseline service (eg GCM) as shown from WFCM General in Figure 49 below.

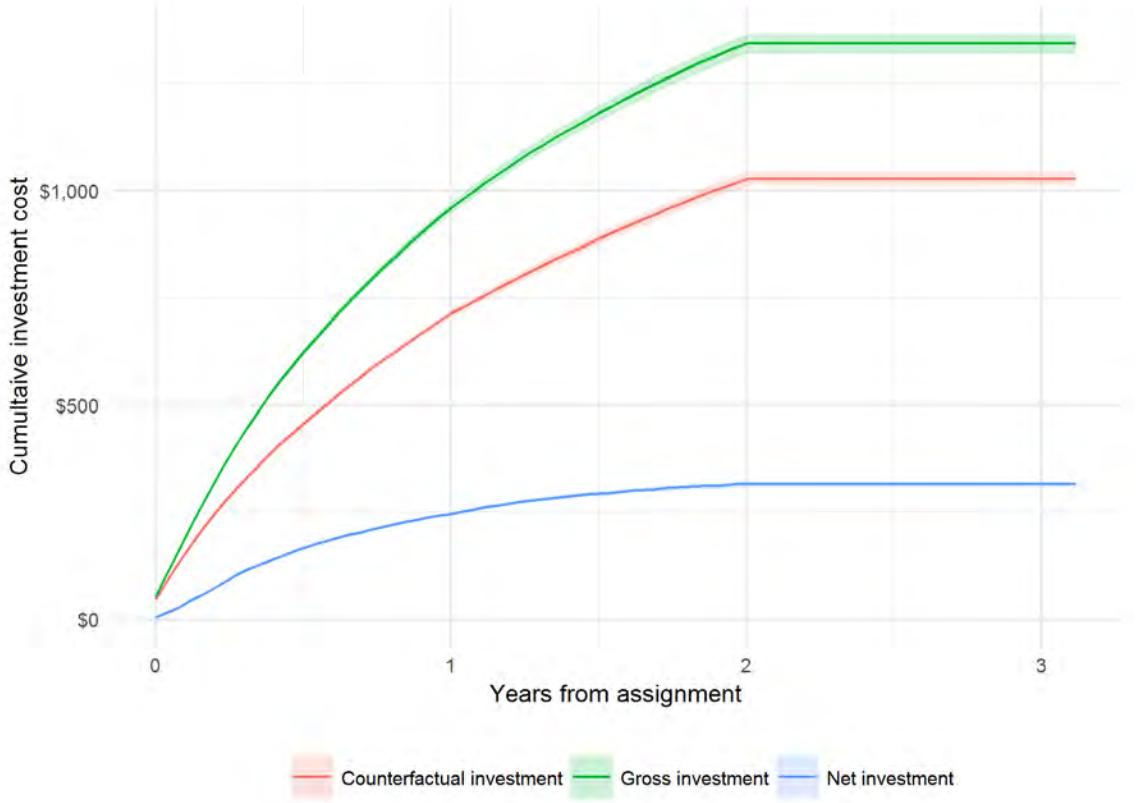
Figure 44: Survival on first benefit spell and survival in service while on first benefit spell for people assigned to WFCM General



The curve represents the survival on first benefit spell from assignment to the service, while the blue line represents the survival on WFCM General conditional on surviving on the first benefit spell. The difference between the two is instances where people have been assigned to a higher priority service within their first benefit spell. The red shaded area is where the investment cost of each service is based; namely, while people are in the service and have not exited from main benefit or been assigned to another service. If a person is assigned to another service, then this record is right censored from that point onward. In other words, we assume that any effects of a service occur while on the service and do not have any influence on people's outcomes once they have been assigned to another service.

Investment costs are based on the average weekly costs of people in each service (see Figure 39) multiplied by the number of people (i) who have not exited main benefit and (ii) on the service in each interval (Figure 49). The investment period ends after two years for WFCM General. At the end of two years the cumulative investment if people had been assigned to WFCM General is estimated to be $1,361. The estimated investment if these people had been in GCM over the same period was estimated to be $1,041, a net investment of $322.

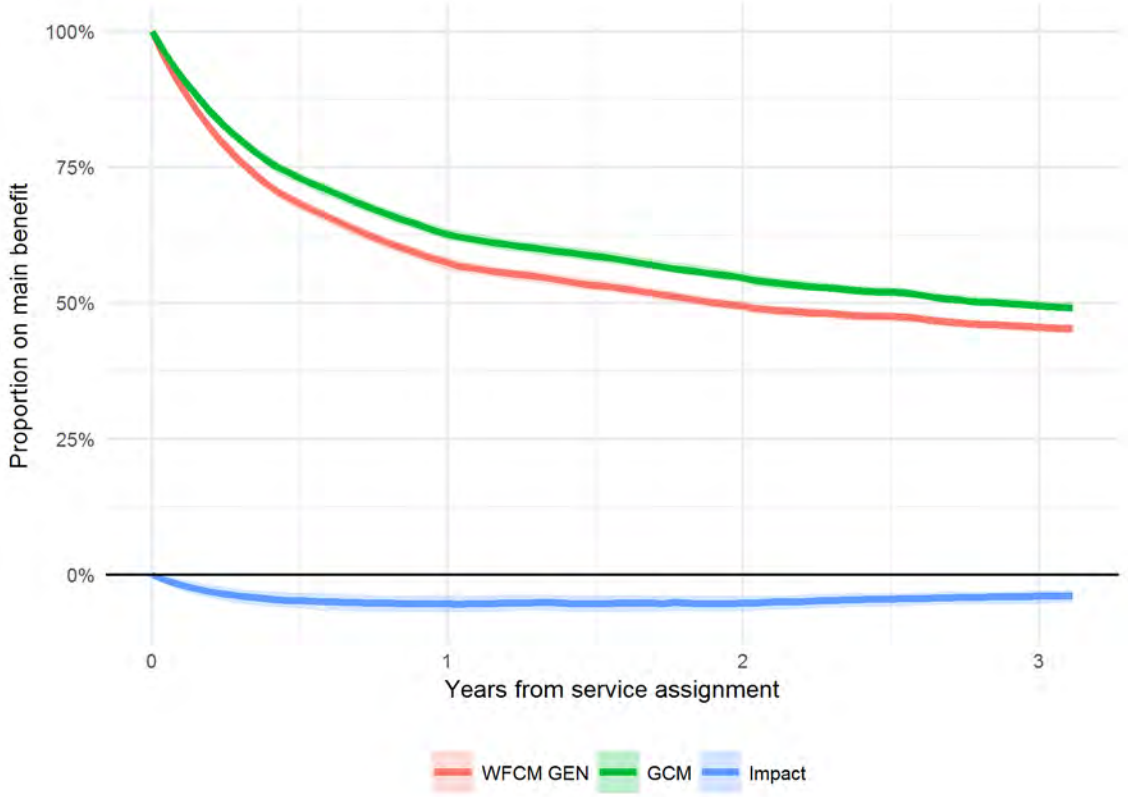
Figure 45: Cumulative gross, counterfactual and net investment for WFCM General



Net-return: impact on income support costs

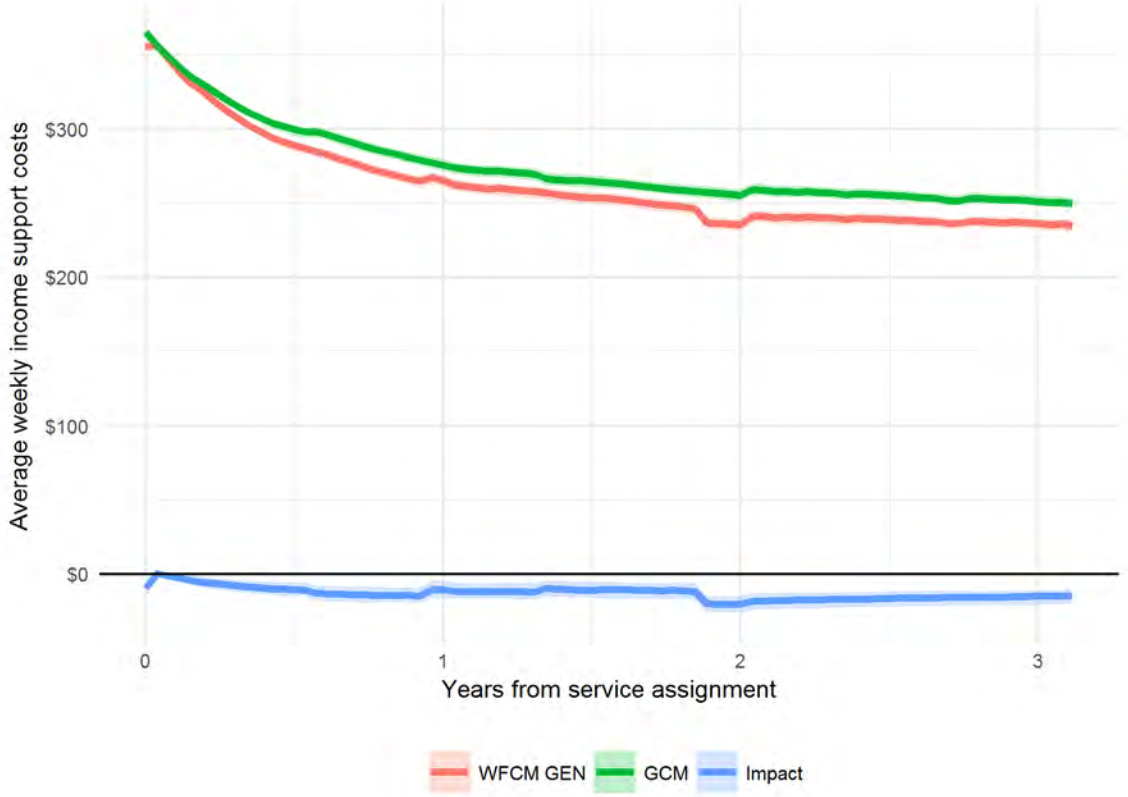
The net-return is primarily driven by the differences in income support costs when assigned to the service, and the baseline comparison. Income support costs in each scenario are based on the total amount of time people spend on their first benefit spell up to the end of the control group holdout period (Figure 36) and if they had exited while on the service the time before they return to main benefit (Figure 40). When we combine these two transition states we can estimate difference in time on benefit between the service and baseline scenario. Figure 46 shows the estimated proportion of people on benefit if assigned to WFCM General or GCM over the three-year follow-up period. The difference in proportion is determined by the reduction in time on first benefit spell as well as an increased time off benefit after exit.

Figure 46: Estimated proportion on main benefit between WFCM General and GCM scenarios



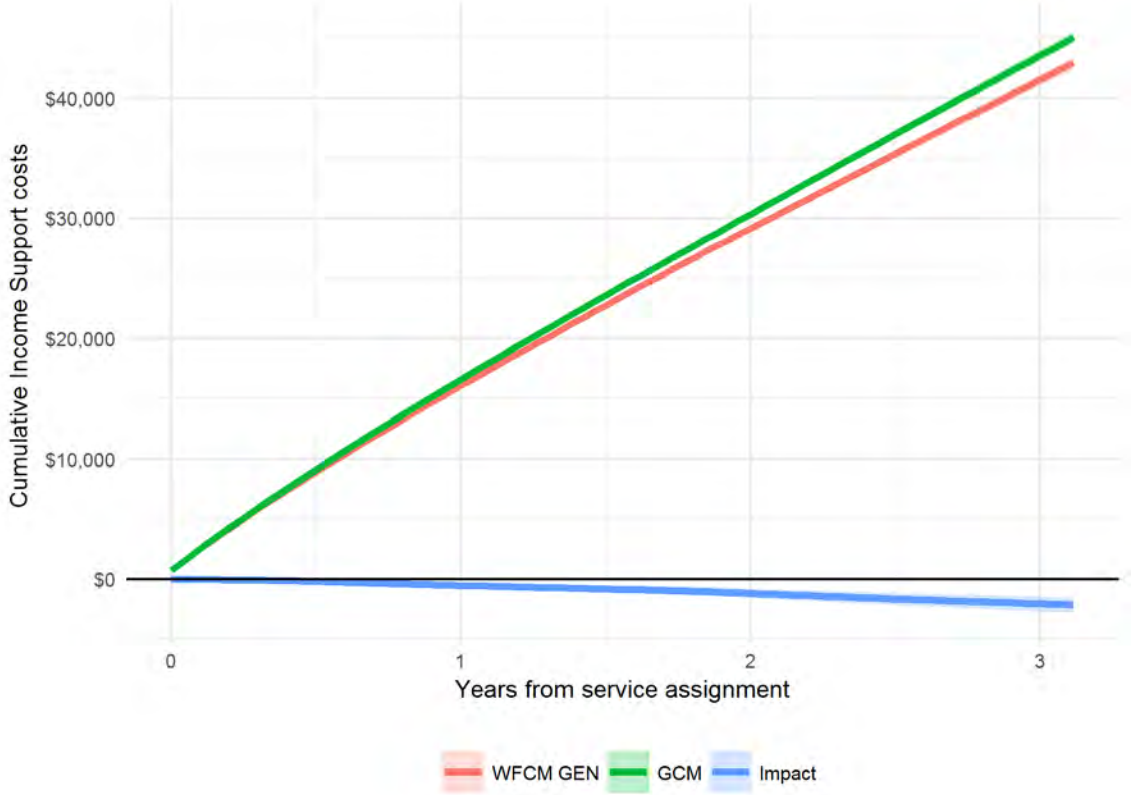
Using the duration on first benefit spell and off benefit spell, the next step is to multiply these values by the estimated income support payments for periods on Figure 38) and off main benefit (Figure 41). When combined with time on main benefit we can estimate the relative income support payments between WFCM General and GCM over the three-year follow-up period (Figure 46). The blue impact line indicates the interval net-return in income support costs.

Figure 47: Weekly income support costs between WFCM General and GCM scenarios



We can use these interval impacts to determine the cumulative impact of assignment to WFCM General or GCM over the duration of the follow-up period as shown in Figure 48. At the end of the follow-up period, we estimate the income support costs for when people are assigned to WFCM General to be $42,952 after 3.1 years compared with $45,057 if they had been assigned to GCM instead. The net-return is the difference between these two figures, namely $2,105.

Figure 48: Cumulative income support costs between WFCM General and GCM scenarios



Discounting

The final step in the calculation of ROI was to discount values to net present value. All monetary values included so far are in real (2017 dollars). We used a constant two percentage point discount rate. The discount rate represents the expected risk-free rate of return, which is itself based on long-term government bond rates. Given the relatively low discount rate and short follow-up periods, discounting has a small impact on the analysis.

Table 14: Effect of discounting on WFCM General welfare ROI values

Estimate	Initial value	Discounted value
Gross investment	$1,361	$1,343
Counterfactual investment	$1,041	$1,027
Income support savings	$2,105	$2,000

We have not included any uncertainty penalties in the discount rate. Uncertainty penalties reflect the view that causal effects over the medium to long term are more difficult to reliably estimate. Adding an uncertainty penalty into the discount rate ensures that any potential positive or negative medium- or long-term impacts make a smaller contribution to the overall cost-benefit results. The challenge with uncertainty penalties is arriving at an independent value of what these would be and whether they would vary according to the robustness of the underlying causal estimates.

Welfare ROI estimate for WFCM General

Based on the above results we can now calculate the wROI for WFCM General.

$$2.26 = \frac{(\$2{,}000 + \$1{,}027)}{\$1{,}343}$$

Calculation of uncertainty

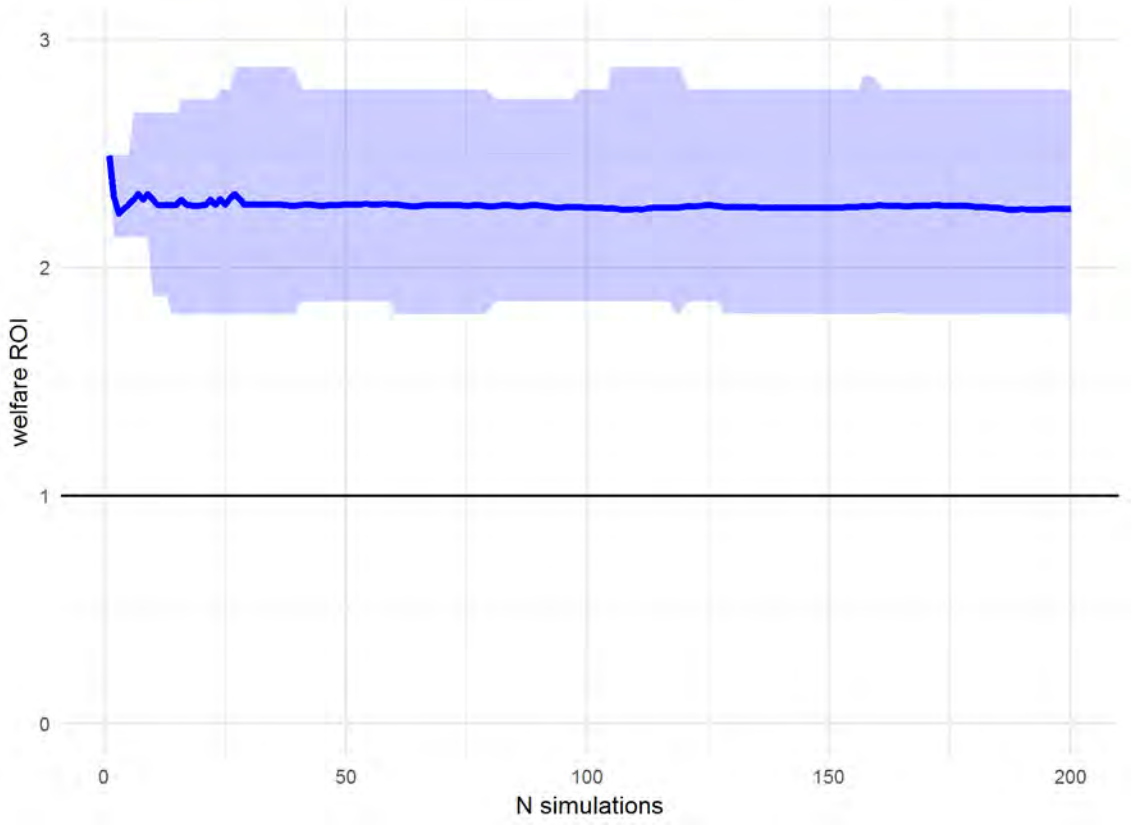
Each of the components that make up the simulation has a level of uncertainty. Therefore, it is important to reflect this uncertainty in the final results of the simulation. To achieve this, we used a Monte Carlo approach whereby we ran multiple versions of the simulation model. Each run took different expected beta estimates to produce a range of likely values from each of the models in the simulation.

The selection of beta values for each run was based on the parameter space for the beta estimates (Gentle, 2003). The parameter space was defined by the model's beta estimate and associated beta co-variance matrix. At the start of each iteration, we took a random draw from the parameter space for each model. We did not include any co-variance between models; instead, beta values were selected independently for each model.

Based on this random draw, we calculated the modelled and counterfactual durations and associated impact estimates. Repeating this process 200 times generated a distribution of expected values based on the parameter space. We took the 97.5 and 2.5 percent intervals of this distribution as our confidence intervals (ie 95 percent) for the results presented in this analysis.

Figure 49 plots the expected wROI for WFCM General as the number of simulations increases. Over the first 25 simulations, we see volatility in the mid-point estimate, after which the estimate stabilises. Looking at the confidence interval, the width of the interval continues to vary until around 125 simulations, at which point the intervals remain broadly stable. Based on these results we considered 200 simulations to provide enough estimates to give us both a stable mid-point estimate as well as a stable confidence interval.

Figure 49: Estimated wROI for WFCM General by number of simulations



Works cited

Allison, P. D. (1984). *Event History Analysis: Regression for Longitudinal Event Data*. *Quantitative Applications in the Social Sciences* (Vol. 46). Retrieved from http://books.google.com/books/about/Event_History_Analysis.html?id=XkACeUEiEsUC

Bandookwala, M., Kemp, C., Anderson, D., & Bly, M. (2014). *Work Focused Case Management for clients with health condition or disability (WFCM: HCD): Process Evaluation Report*. Wellington: Ministry of Social Development.

Crane, B., & Kemp, C. (2013). *Work and Income Service Model Trial: Process Evaluation Report*. Wellington: Ministry of Social Development.

de Boer, M., & Ku, B. (2017). *Service Delivery Cost Allocation Model for Individual Outputs: 2017 version*. Wellington: Ministry of Social Development.

Gentle, J. E. (2003). Random Number Generation and Monte Carlo Methods. *Statistics and Computing*, 229−281. http://doi.org/10.1007/b97336

Hall, P., Herdina, N., & Henshaw, K. (2016). *Intensive Client Support (ICS) trial evaluation: Interim 12 month evaluation*. Wellington: Ministry of Social Development.

MSD. (2016). *Dataset Documentation: Staff time tracking*. Wellington: Ministry of Social Development.

MSD. (2016). *Effectiveness of Contracted Case Management Services Sole Parent Employment Service and Mental Health Employment Service Trials Evaluation: Final Report*. Wellington: Ministry of Social Development.