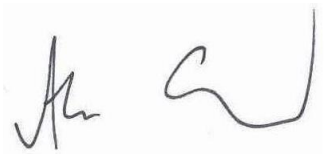
Ministry of Social Development

&

The Treasury

ACTUARIAL ADVICE OF FEASIBILITY: A LONG-TERM INVESTMENT APPROACH TO IMPROVING EMPLOYMENT, SOCIAL AND FINANCIAL OUTCOMES FROM WELFARE BENEFITS AND SERVICES

27 October 2011



Alan Greenfield
Fellow of the Institute of Actuaries of Australia

Sydney
Level 11, 55 Clarence Street
Sydney NSW 2000
P 02 9249 2900
F 02 9249 2999

Melbourne
Level 6, 52 Collins Street
Melbourne VIC 3000
P 03 9658 2333
F 03 9658 2344

ACN 087 047 809
ABN 29 087 047 809
W www.taylorfry.com.au

TABLE OF CONTENTS

APPENDICES

1 EXECUTIVE SUMMARY

1.1 Introduction

The Welfare Working Group (WWG) was established by Cabinet in April 2010. The arms-length group was tasked to conduct a fundamental review of the welfare system and develop options to reduce long-term dependency with a focus on:

- improving work outcomes for sole parents and for people with disabilities and ill health
- how welfare should be funded and any lessons from the insurance industry and ACC in managing forward liability; and
- whether the structure of the benefit system contributes to long-term dependency.

In August 2010, the WWG released an issues paper finding that the benefit system has failed to keep pace with changing expectations about paid work. The WWG also found that there are weak signals about the value of investing early to prevent long-term benefit use, and that the economic and social costs of the current system are high and unsustainable.

The WWG presented its comprehensive set of 43 recommendations to the Government on 22 February 2011. At a high level, the WWG recommends a work-focused welfare system, with a cross-government emphasis on preventing the need for welfare use, with targets and accountability mechanisms to reduce future liability.

Taylor Fry Consulting Actuaries ("Taylor Fry") has been asked by the New Zealand Government Ministry of Social Development ("MSD") and Treasury to provide advice on:

I. *"the feasibility of adopting a long-term investment approach to achieving better employment, social and financial outcomes through the welfare system*

II. *how aggregate future liability in the welfare context could be calculated."*

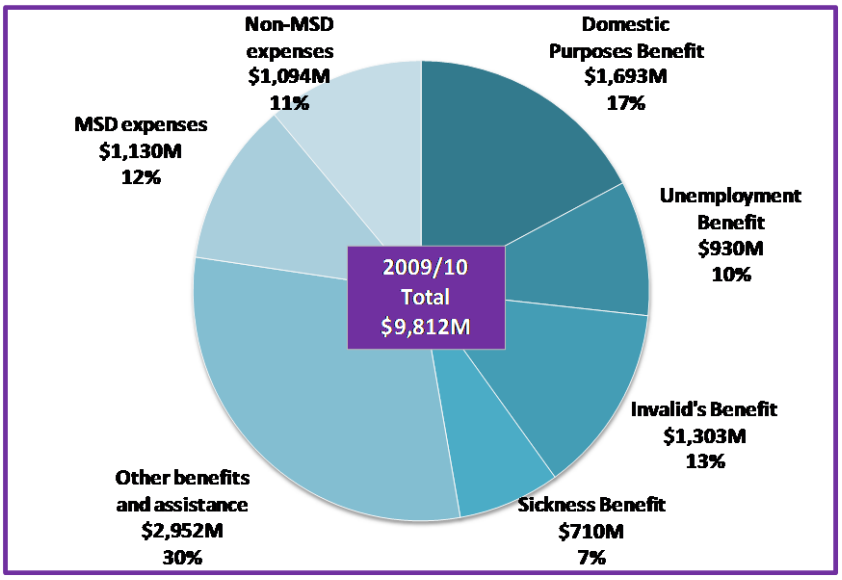
1.2 The benefit system

The benefits within scope are:

- Main benefits
  - Domestic purposes benefit
  - Unemployment benefit
  - Invalid's benefit
  - Sickness benefit
- Supplementary benefits
- Hardship payments

Figure 1.1, reproduced from Section 2, provides a summary of the relative magnitude of the costs under consideration based on payments made for the 2009/10 financial year.

Figure 1.1 **Benefits and costs within under consideration**



1.3 Is the approach feasible?

The question is asked:

Is a long term investment approach to improving employment, social and financial outcomes from welfare benefits feasible?

And answered:

Yes.

This feasibility study requires advice with particular emphasis on considerations required to:

"maintain a focus on employment outcomes, given that welfare is a system of last resort?"

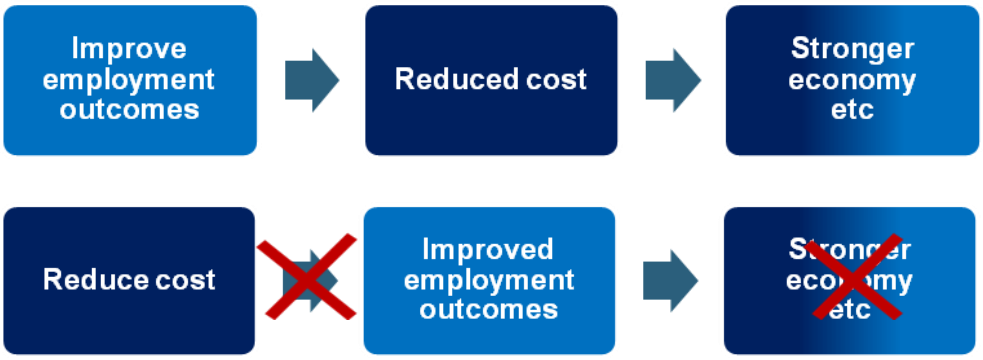
It is evident that without maintaining this focus there is a distinct likelihood that any attempt to improve financial outcomes for the social welfare benefit system may in fact result in a deterioration in overall costs to society. For example, a reduction in the eligibility of benefits, without any reference to initiatives to place people in work would likely result in higher levels of homelessness, increases in the cost of the health and justice systems and lower productivity.

Section 3 discusses the feasibility of the approach and provides a recommended framework to reduce the cost of social welfare by improving employment outcomes with further consequential savings to the health and justice sectors and increased productivity and tax receipts.

We have considered the requirements of such an approach and developed a framework that aims to achieve these outcomes. The recommended framework is drawn from various practices observed in accident compensation schemes, social welfare, and customer focused organisations.

However, to achieve the stated aims listed above it is considered a requirement that the focus should be on improving employment outcomes as the primary driver of reducing cost. A focus on reducing cost without the primary goal of improving employment outcomes might lead to adverse consequences. This position is represented in Figure 1.2 reproduced from Section 3.

Figure 1.2 **Primary focus: improving employment outcomes**



1.4 Recommended framework

In summary the framework required to achieve the stated aims of reduced cost without adverse consequences needs to:

- Have a primary focus on improving employment outcomes;
- Provide a means of overall financial control and accountability for the social welfare system; and
- Link the savings in improved employment outcomes for individuals and groups of individuals to the overall financial control.

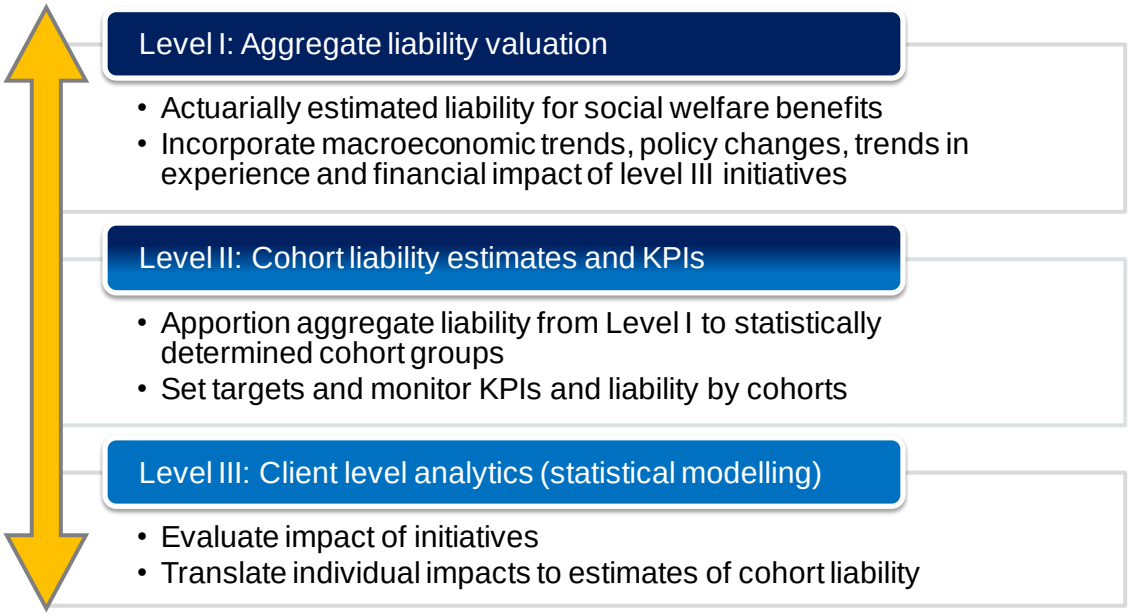
Section 3 outlines a framework to achieve these three critical features of an approach to managing the social welfare system.

The management of the social welfare system can be considered to occur on three levels:

- Level I - Policy decisions;
- Level II - Strategic decisions; and
- Level III - Tactical decisions.

Figure 1.3 shows the recommended framework of analysis and reporting to strengthen the management control at each of these levels.

Figure 1.3 **Proposed framework: 3 levels of analysis**

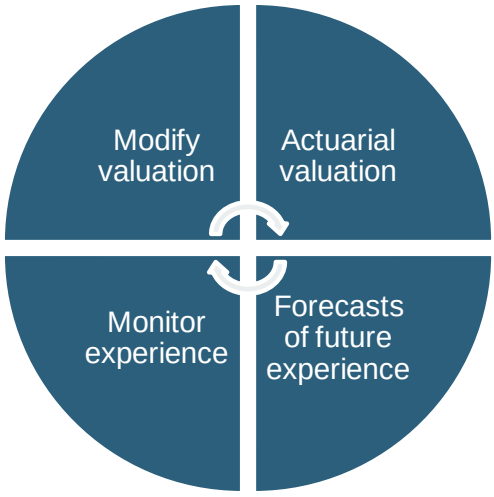


Note: KPIs are "Key Performance Indicators".

1.4.1 Level I – Policy decisions and financial control

The recommended approach for Level I requires the periodic valuation of the liability for social welfare benefits using an actuarial approach. This is the same approach as is adopted in the insurance sector. Figure 1.4 summarises the Level I framework which is detailed in Section 4.

Figure 1.4 **Proposed framework: Level I**



The figure above shows the actuarial control cycle recommended for implementation for the overall financial control of social welfare benefits. It requires:

- A periodic actuarial valuation of the liability for social welfare benefits;
- Forecasts of future experience based on the models and results of the actuarial valuation;
- Monitoring of the experience compared to the forecasts throughout the period between valuations; and
- At the end of the period, an updated actuarial valuation incorporating modifications due to the comparison of actual and forecast experience and other expected changes.

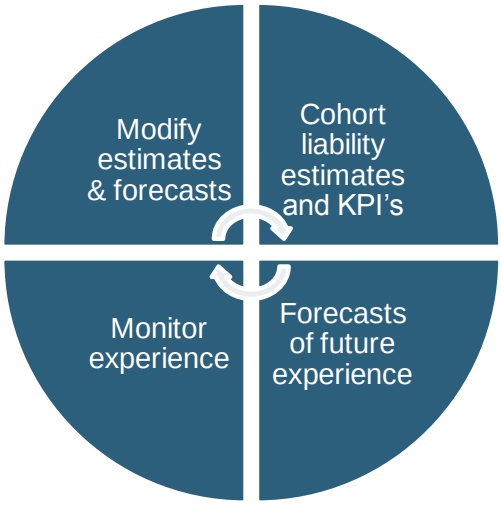
1.4.2 Level II – Strategic decisions and performance measurement

Level II, the cohort-level analysis provides both financial and other performance indicators for groups of similar beneficiaries to enable:

- Strategic decision-making in relation to groups of beneficiaries, particularly in relation to allocation of resources and targeting of services and interventions;
- Target setting and performance measurement of those groups; and
- Evaluation of the financial impact of Level III initiatives to inform the aggregate liability valuation and to explore cost benefit tradeoffs for cohorts of beneficiaries.

Figure 1.5 summarises the Level II framework while Section 5 describes the approach in detail.

Figure 1.5 **Proposed framework: Level II**



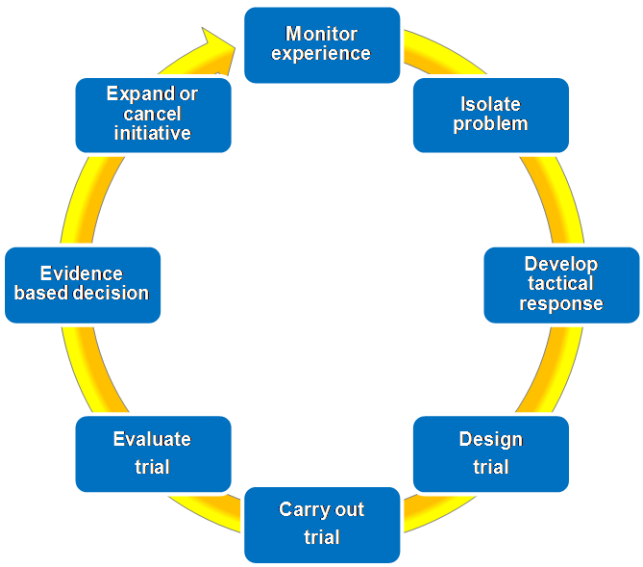
Level II requires a statistical and actuarial approach to:

- Determine appropriate cohorts of benefit recipients that provide management with meaningful groups to analyse and manage. This is typically done using a segmentation analysis;
- Estimate the average future liability by cohort. This will require the allocation of the total liability from Level I to each cohort using a measure of relative cost;
- Derive and estimate a number of key performance indicators for use in the performance measurement of cohorts;
- Forecast the future liability and KPIs by cohort;
- Monitor the experience of cohorts.

1.4.3 Level III –Tactical decisions and client initiatives

The recommended approach for Level III requires the development and evaluation of tactical initiatives to drive behavioural change amongst benefit recipients to reduce the cost of social welfare. This is done by the translation of innovation to concrete initiatives to improve employment outcomes and evaluation of those initiatives. This is in contrast to Levels I and II which enable increase understanding of the overall cost, provide accountability, transparency and financial control, performance measurement and enable strategic decisions relating to allocation of resources. Therefore, it is important to develop a framework that drives innovation at Level III. Figure 1.6 summarises the Level III framework while Section 6 describes the approach in detail.

Figure 1.6 **Proposed framework: Level III**



1.5 Review of MSD capabilities

Generally MSD is well placed to implement the framework recommended in this report. A summary of MSD capabilities and general recommendations in relation to each of the levels of analysis is provided in Table 1.1 which is reproduced from Section 7.

Table 1-1 **Summary of MSD capabilities**

	Personnel capabilities	Recommendation
Level I **Actuarial valuation**	Currently MSD has no actuarial team in place to undertake this part of the framework. MSD do have a forecasting team that provide short term forecasts of numbers of beneficiaries and amount of payments.	At least in the short to medium term appoint external advisers to undertake the required analysis. The actuarial valuation should take account of the MSD forecasts of numbers of beneficiaries and amount of benefit payments.
Level II **Cohort statistical analysis**	Currently MSD has a team of staff with the skills necessary for carrying out this sort of work on an ongoing basis.	Initially it is likely that the team will lack some of the experience necessary to implement the framework suggested. We recommend that MSD partner with an external consultant to build the initial framework and further develop the MSD team's experience and capability.
Level III **Recipient modelling and trial evaluation**	Currently MSD has a small team of staff with the skills necessary for carrying out this sort of work on an ongoing basis.	Initially it is likely that the team might benefit from the experience of an external consultant who has carried out these sorts of projects to: a) Develop the rigorous framework for the approach, including modelling, experiment design and evaluation; and b) Carry out 2 or 3 initial projects to embed the process within MSD.

1.6 Funding

The issue of whether to forward fund a social welfare scheme is a question of high level policy and national economics. Neither of these fields are within our expertise. Nevertheless the following comments drawn from Section 8 are relevant to the discussion:

- Accountability. The existence of a fund (even a partial fund) would very likely give increased emphasis to the amount of the liabilities and to the change in liabilities from period to period. Analysis of the change in funding level from period to period could be used as one of the prime targets for accountability. This is probably the strongest reason for implementing some level of funding.
- Fairness. Intergenerational equity is often given as a reason for funding. However, it is not a simple matter to determine a fair allocation of the cost of the social welfare system and therefore in this context is considered of minor importance.
- Feasibility of full funding. The social welfare system currently provides payments of the order of $10 billion per annum including expenditure in relation to employment services and supports and the costs of administration. The liability for the current recipients will be several times this amount. It is therefore not feasible to consider fully funding this liability even in the medium term.
- Economic reasons.
  - The existence of a fund could be used to dampen or even offset the impact of changes in the economy.
  - Payment into the fund could be made either from existing revenue or by an additional levy of some description. The latter would be a form of additional forced savings which may be considered a prudent economic outcome.

  While these economic reasons seem reasonable in isolation, in fact they form just a part of the overall annual NZ budget. Thus to the extent that funds are diverted from other organisations, or from retiring Government debt, the question becomes one of what is the best use for the funds.

Most importantly, the recommended framework in this report to achieve the investment approach to social welfare can be implemented with or without funding the liability for social welfare benefits.

1.7 Implementation

Section 9 details many of the various issues and risks that are relevant to this proposed approach to the financial control of the welfare system. Some of the more important issues include:

- Reducing cost is achieved by improving employment outcomes as the primary focus (not the other way around);
- The best result would be obtained from implementation of all 3 levels;
- The approach requires significant investment of time & resources;
- The Level I valuation approach needs to be designed carefully;
- Building Level II & III capability takes time. It is important to get it right;
- It is critical that Level III includes a rigorous approach to evaluation of initiatives (controlled trials are best practice in this regard);
- Avoid unnecessary complexity, particularly in the determination of the cohort groups and Level II models;
- Don't underestimate the cultural change & buy in required to achieve the desired outcomes.

2 INTRODUCTION

2.1 Background

Taylor Fry Consulting Actuaries ("Taylor Fry") has been asked by the New Zealand Government Ministry of Social Development ("MSD") and Treasury to provide advice on:

III. *"the feasibility of adopting a long-term investment approach to achieving better employment, social and financial outcomes through the welfare system*

IV. *how aggregate future liability in the welfare context could be calculated."*

"*The advice should consider:*

1. *assumptions - what would be important to consider in order to maintain a focus on employment outcomes, given that welfare is a system of last resort?*

2. *scope of liability - how would the costs of services to reduce benefits fit in if the scope of liability is benefit exposures?*

3. *peer reviewing the work that MSD has already undertaken to measure liability:*

    a. *advise on whether it is robust and meets appropriate external standards*
    b. *what would be needed if it does not?*
    c. *the applicability of existing methods from elsewhere for measuring liabilities to welfare, for example, from social insurance schemes*
    d. *what other information could be used in existing tools to better predict where to intervene to reduce future liability?*

4. *funding – advice on the level of funding that would be required to make partial funding feasible through cyclical variations.*

5. *any significant implementation issues or barriers that would need to be addressed before a liability approach could be adopted, as well as any other information that you consider relevant relating to the feasibility of the proposal."*

Also note that the scope has been restricted by MSD and Treasury to exclude:

- Benefits other than MSD Main benefits, Supplementary benefits and Hardship assistance (i.e. New Zealand Superannuation and Veteran's pension are excluded);
- Financial and other impacts outside of the welfare benefit system (e.g. financial and other impacts on the health, justice and education system and on the ACC);
- Any other Welfare Working Group recommendations, including institutional structure.

2.2 New Zealand social welfare benefits

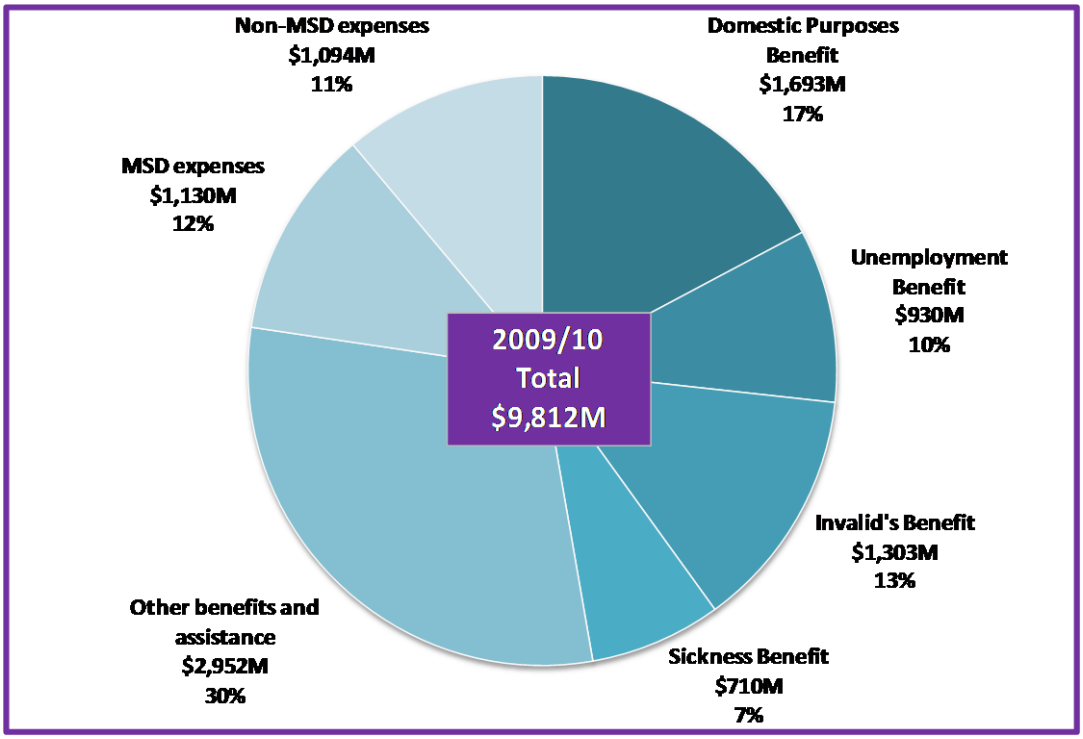
As noted in the previous section, this feasibility study is concerned only with the certain MSD social welfare benefits and the costs of services and service delivery to those benefit recipients.

The benefits within scope are:

- All main benefits, e.g.
  - Domestic purposes benefit
  - Unemployment benefit
  - Invalid's benefit
  - Sickness benefit
- Supplementary benefits
- Hardship payments

Figure 2.1 provides a summary of the relative magnitude of the costs under consideration based on payments made for the 2009/10 financial year.

Figure 2.1 **Benefits and costs within under consideration**



2.3 Interrelationship of social benefits system and other social structures

As noted in Section 2.1 impacts outside the social welfare system are excluded from the scope of this report. That is, impacts on, for example, the justice, health and education systems and the interaction with ACC are excluded. Nevertheless it is worth at least commenting that these interrelationships exist and understanding that changes in one area can impact another. In particular, benefit payment reductions without return to work

are likely to lead to adverse social outcomes such as an increase in homelessness and crime rates and hence higher costs to society. The existence and importance of these relationships is one of the distinguishing features between a social welfare context and an insurance context.

2.4 Scope of report

The two previous sections described the social welfare context in NZ and its relationships to other parts of society. The scope of this feasibility, as required by the contract, has been limited to MSD benefits, excluding the superannuation pension. This is sensible as the external relationships are extremely complex. In any event, provided that a sound approach to the proposed framework in relation to MSD benefits is followed, the impact of changes in MSD benefits will be a proxy for wider economic impacts. This is an important distinction and is discussed in greater detail in Section 3.2

Note also that this report has been written generally at the feasibility level. That is, questions in relation to whether an approach is reasonable and the basic premise and methodology are provided. However, the report does not attempt to cover in significant detail the appropriate methodologies and implementation. We consider this to be outside of scope and properly belongs in a subsequent study related to implementation should the New Zealand Government wish to proceed with this overall investment approach.

2.5 Remainder of report

The remainder of this report is set out as follows:

Section 3	Proposed framework
Section 4	Level I: Policy decisions & financial control
Section 5	Level II: Strategic decisions & performance measurement
Section 6	Level III: Tactical decisions & client initiatives
Section 7	Review of MSD capabilities
Section 8	Funding
Section 9	Implementation

3 PROPOSED FRAMEWORK

3.1 Feasibility of approach

Is a long term approach to improving employment, social and financial outcomes from welfare benefits feasible?

Yes.

Taylor Fry has considered the requirements of such an approach and developed a framework that aims to improve employment outcomes, resulting in a reduction of the cost of social welfare with probable further consequential savings to the health and justice sectors and increased total productivity and tax receipts. The recommended framework is drawn from various practices observed in accident compensation schemes, social welfare, and customer focused organisations. The following sections provide an outline of the approach and the reasoning behind its development.

3.2 Reasoning behind proposed framework

This feasibility study requires advice with particular emphasis on considerations required to:

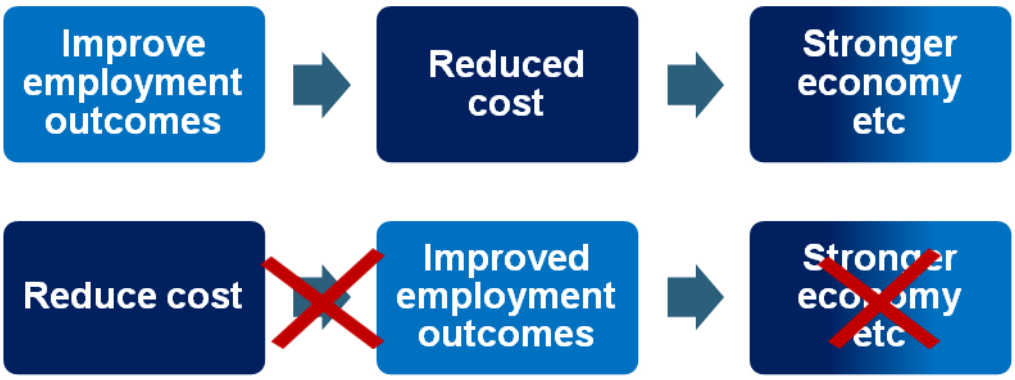
"maintain a focus on employment outcomes, given that welfare is a system of last resort?"

It is evident that without maintaining this focus there is a distinct likelihood that any attempt to improve financial outcomes for the social welfare benefit system may in fact result in a deterioration in overall costs to society. For example, a reduction in the eligibility of benefits, without any reference to initiatives to place people in work would likely result in higher levels of homelessness, increases in the cost of the health and justice systems and lower productivity.

Figure 3.1 contrasts two high level approaches to financial control. The top row highlights that a primary focus on improving employment outcomes (principally workforce participation) will, by definition, lead to a reduced cost and improvements in the economy and society as a whole. The second row highlights that a focus on cost reduction does not necessarily result in improved employment outcomes or any improvements in the economy or society.

Furthermore, the use of better employment outcomes as the primary driver of reductions in social welfare benefits means that estimated savings in the cost of welfare can be used as a proxy for economy-wide savings, e.g. if there are savings in welfare benefits via more people in employment there will also be savings in the cost of the health and justice systems and increases in productivity and taxes.

Figure 3.1 **Primary focus of the financial control of the welfare benefit system**



Therefore the proposed framework needs to provide a principal focus on improving employment outcomes while also achieving financial control and accountability. To enable financial control the framework needs to be able to translate improved employment outcomes into an overall financial impact.

In summary the framework required to achieve the stated aims of reduced cost without adverse consequences needs to:

- Have a primary focus on improving employment outcomes;
- Provide a means of overall financial control and accountability for the social welfare system; and
- Link the savings in improved employment outcomes for individuals and groups of individuals to the overall financial control.

Section 3.3 outlines a framework to achieve these three critical features of an approach to managing the social welfare system.

3.3 Recommended Framework

3.3.1 Overview

The management of the social welfare system can be considered to occur on three levels:

- Policy decisions – governance of the welfare system at the highest level and for the long term, including overall design of the system, rules for eligibility and levels of benefit, i.e. legislation and regulations. These decisions are generally undertaken by Ministers of Government;
- Strategic decisions – management of the welfare system at mid level. These are the sorts of decisions undertaken that generally impact the system over the medium term and include interpretation of the rules for eligibility and allocation of resources to managing the system. These decisions are generally taken by senior public servants and management.
- Tactical decisions – management of the welfare system at the lowest level, i.e. in relation to managing beneficiaries or small groups of beneficiaries. These are

the sorts of decisions that are being made every day by case workers and periodically (e.g. monthly, quarterly) by managers in relation to tactics (i.e. case management, referrals to training, other employment interventions etc.) used to get beneficiaries into employment and hence reduce benefit payments.

Figure 3.2 describes this view of the management of the social welfare benefit system, noting that the strategic and tactical decisions could be considered as comprising operational control, while policy and strategic decisions, could be considered to comprise financial control.

Figure 3.2 **The social welfare system: 3 levels of management control**

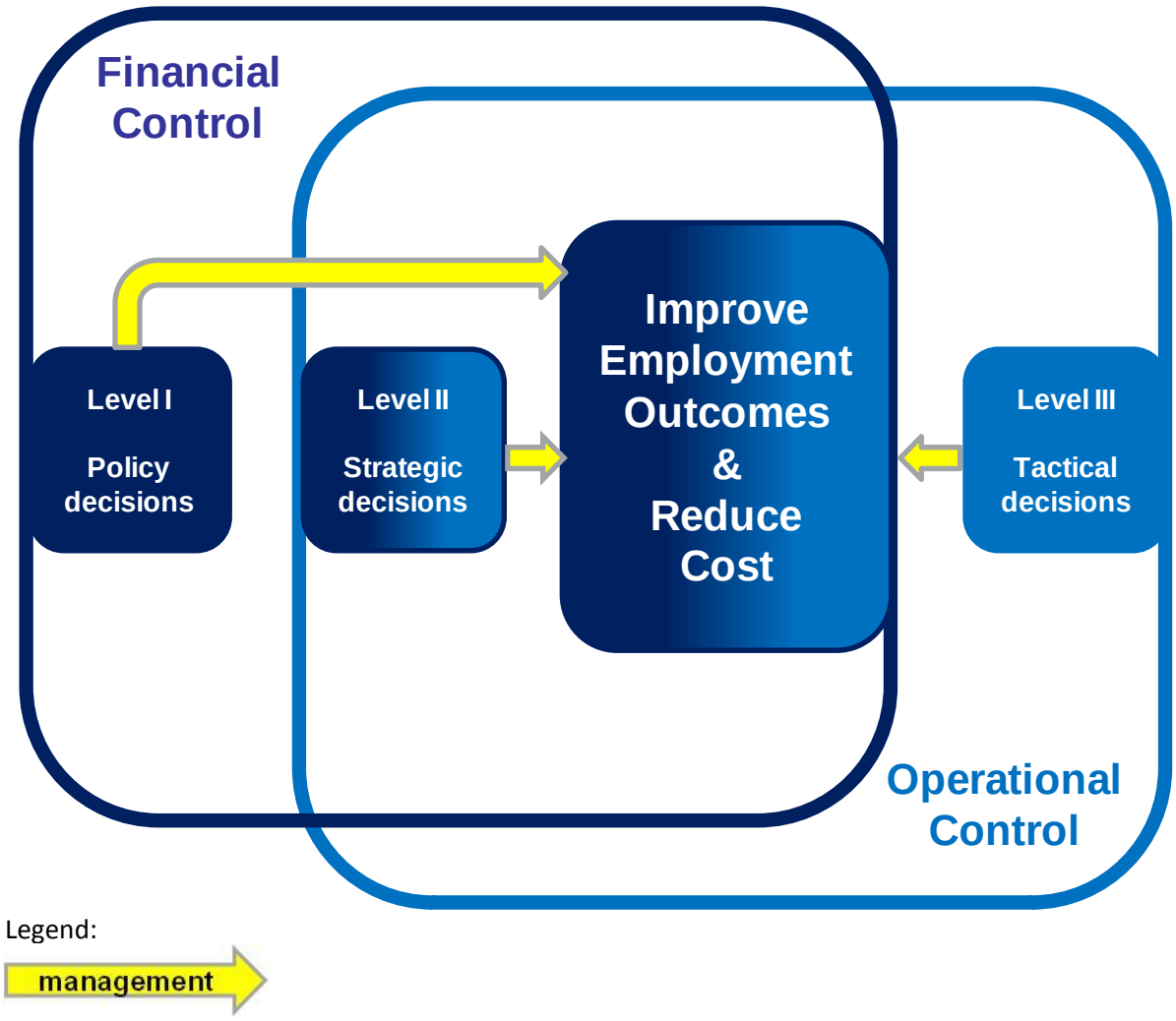
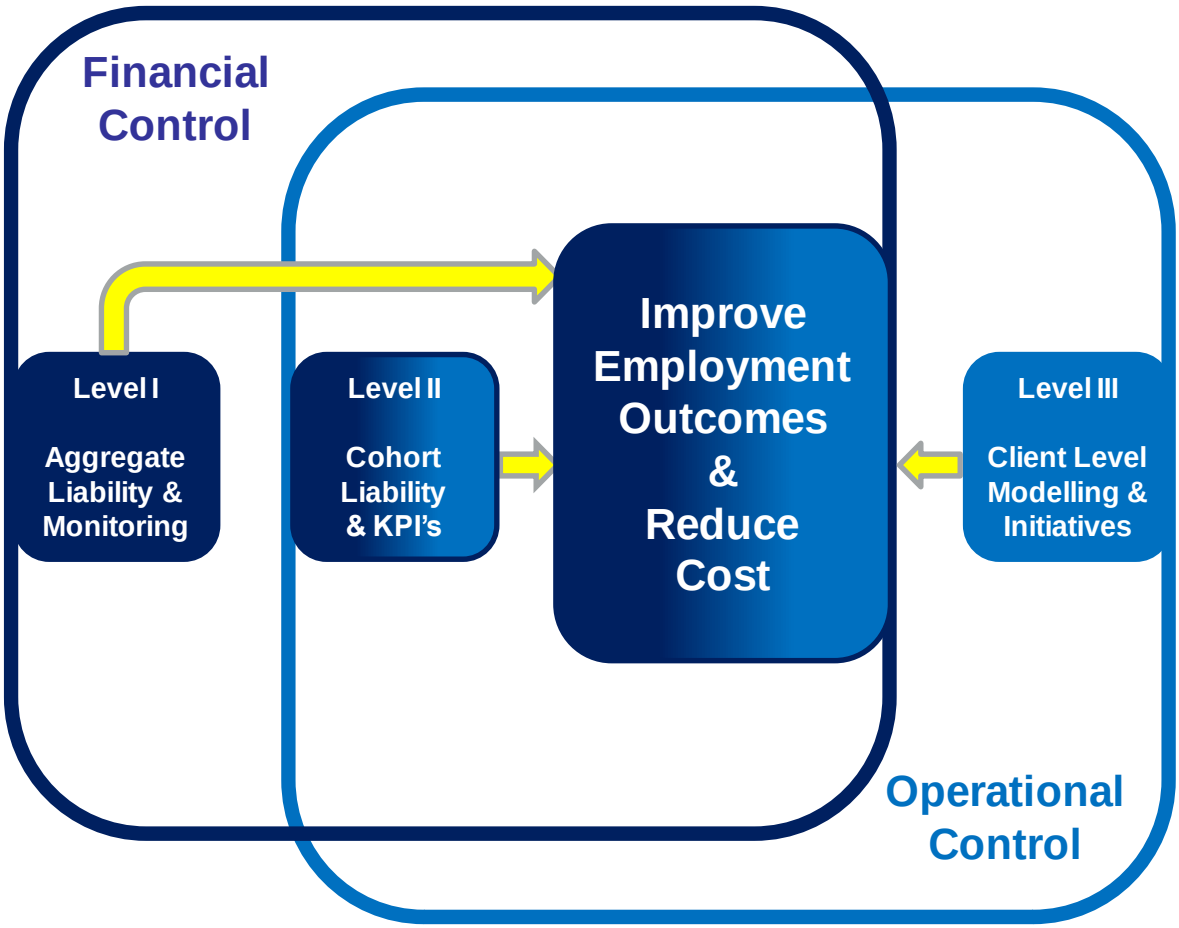


Figure 3.2 assigns each of the decision types to a level of decision making: Levels I to III. The proposed framework requires effective management tools and analysis at each of these levels to aid in the decision making process. Figure 3.3 outlines the recommended approach to each level.

Figure 3.3 **Proposed framework: 3 levels of analysis**



As shown in Figure 3.3 the following analysis and management tools are proposed for each level of decision making:

- Level I, policy decisions: an actuarial valuation of liabilities in respect of welfare benefits and associated forecasts, and monitoring of experience;
- Level II, strategic decisions: actuarial valuation of liabilities and estimated key performance indicators ("KPIs") in respect of statistically relevant and meaningful cohorts of benefit recipients and associated forecasts, and monitoring of experience; and
- Level III, tactical decisions: analytical (or statistical) modelling of beneficiary behaviour and evaluation of initiatives to change behaviour.

An outline of the proposed analysis at each level is provided in Sections 3.3.2 to 3.3.4. Further detail is provided in Sections 4, 5 and 6, respectively.

3.3.2 Level I: Policy decisions

An actuarial valuation of the liability for welfare benefits, the associated forecasts of benefits and monitoring of experience are an integral part of the framework required to understand and control the financial health of the welfare system. They would provide:

- A financial benchmark of the total cost of welfare;
- An understanding of the change in the cost of welfare. In particular, they would provide a means for determining the overall change in the cost of welfare and an approximate estimate of the impact of various factors on the cost of welfare, particularly due to changes in economic assumptions, changes in benefit design, MSD strategies and initiatives and all other external factors (such as the health of the economy, socio-economic and demographic trends) combined. Note that attribution of the change in liability to these various factors is a difficult task and requires additional information including that provided from the recommended Level I and Level II analysis;
- Accountability and performance measurement for those that manage the scheme via the understanding of the changes in overall cost. This is important as it enables the setting of high level targets for management to assist in the overall financial control;
- Transparency in the cost and management of the system;
- A means of analysing the financial impact of policy changes.

3.3.3 Level II: Strategic decisions

The cohort-level analysis provides both financial and other performance indicators for groups of similar beneficiaries to enable:

- Strategic decision-making in relation to groups of beneficiaries, particularly in relation to allocation of resources;
- Target setting and performance measurement of those groups; and
- Evaluation of the financial impact of Level III initiatives.

Of particular importance to better strategic decision making is the use of “lifetime costs” (or estimates of future liability) for benefit recipients. The use of such a measure enables a more complete analysis of cost / benefit comparisons in contrast to a comparison against annual benefit costs. This is the core concept underlying the investment approach.

3.3.4 Level III: Tactical decisions

Level III of the recommended framework is the component which drives improved employment outcomes to reduce the cost of social welfare via innovation of initiatives to improve employment outcomes and evaluation of those initiatives.

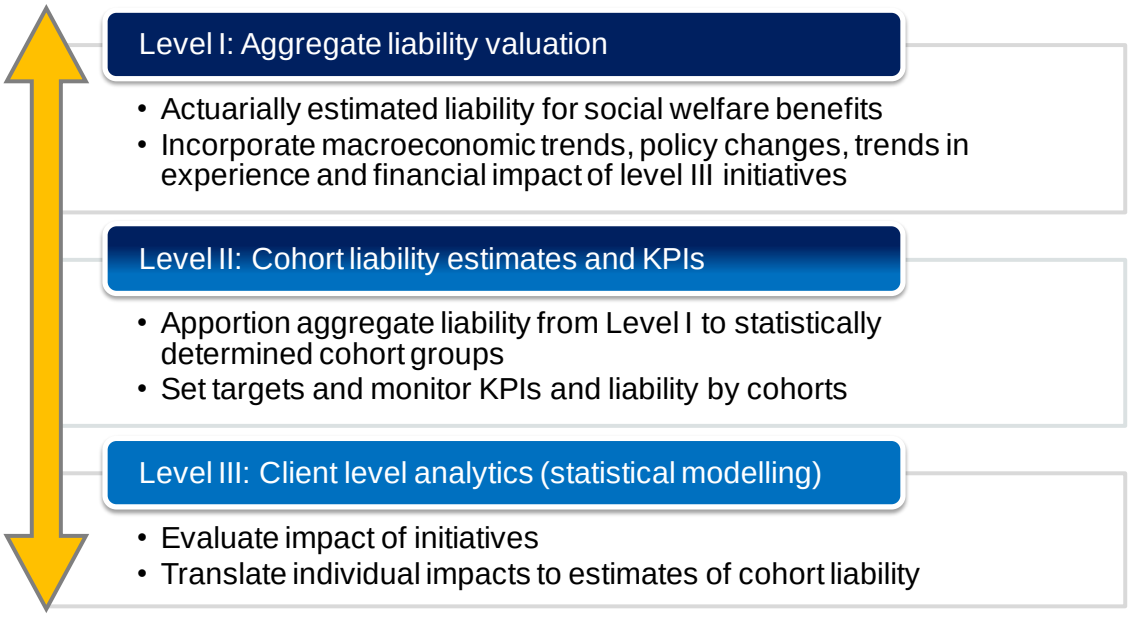
3.4 Recommended Framework – information flow between levels

The proposed framework outlined above requires certain flows of information between the various levels of management and analysis. These linkages are also described in greater detail in Sections 4 to 6. However, in brief they relate to:

- Incorporation of financial impact of Level III decisions on Level I and Level II estimates of liability;
- Incorporation in Level I actuarial estimate of liability of financial impact of:
  - Level II decisions;
  - Trends in experience observed at the cohort level; and
- Apportionment of Level I estimate of total liability to Level II liability estimates by cohort.

These flows are shown diagrammatically in figure 3.4

Figure 3.4 **Proposed framework: information flow**



3.5 Comparison of the proposed framework to current industry practice

The framework outlined above has been drawn from practices in place in various industries and organisations. The sorts of analysis required have been in use for many years in the case of Levels II and III and decades in the case of Level I. Nevertheless, it is our understanding that few organisations, if any, maintain a framework as rigorous and complete as that described. It would certainly be a world first for a social welfare system.

3.5.1 Level I – Aggregate liability valuation

The type of analysis and financial control described as Level I occurs in the vast majority of insurance companies and accident compensation schemes around the world. In fact it is

generally a regulatory requirement in the Western world for insurance companies to undertake actuarial valuations of their liabilities. Principally such valuations are used for the purpose of solvency management and premium setting. In the proposed framework the principal role is one of financial control (akin to solvency management in the insurance context) and cost-signalling (akin to premium setting). Further the financial management of insurance companies generally revolves around setting budgets and forecasts typically derived from the actuarial valuation, and through monitoring and assessment of financial performance based on the comparison of experience with those forecasts. Typically, such monitoring occurs on a monthly basis.

Actuarial practice in this area has been developed over the last hundred or so years with more rapid development over the last 25 years or so due to the increase in computing power. Many of the actuarial models used for valuations that are in place could be considered relatively mature. There is a very high level of understanding within the actuarial profession to enable the application of standard actuarial techniques to the valuation of social welfare liabilities.

3.5.2 Level II – Cohort liability estimates and KPIs

Most insurance companies and accident compensation schemes would undertake cohort level analysis of some description. However, the level of sophistication varies enormously across the industry. Obvious examples of its deployment are claim triage models in workers' compensation which separate administrative only claims from short and long term weekly benefit claims. These sorts of models have been used successfully to improve injury management in this sector.

Insurance companies also monitor their experience by cohort and use the results to manage the business. Action would be taken for cohorts with deteriorating cost trends either via price adjustment, tighter underwriting or claims management practices. E.g. a motor insurer might analyse separately windscreen claims, collision claims and theft claims by region. A deterioration in expected cost trends for a particular cohort would be known relatively quickly via monthly reporting and action undertaken to either alter the trend via tighter underwriting and / or claims management or increase revenue via price changes.

While the analysis occurs in most insurance companies it is important to recognise that the proposed framework requires a deeper implementation and connection to other parts of the framework than is usually applied. In this respect the framework is more holistic than observed elsewhere.

3.5.3 Level III – Client level analytics

The Level III framework is all about innovation. The Level III team would be responsible for formulating initiatives to improve employment outcomes. Initial concepts could come from many sources including academic papers, overseas institutions, New Zealand welfare experience etc. The team needs to be both proactive in looking for areas of improvement as well as reacting to identified trends and problems. Innovation is complemented by evaluation of initiatives through trials.

The analysis component of this level involves statistical modelling of client behaviour using client level data. This sort of modelling is sometimes known as "data analytics" or "data mining" and has proliferated in the last 10 to 20 years due to the increasing collection of

large data sets, the increases in computer power and the development of algorithms to search for patterns in the data. Most large customer focussed organisations now have a team of statisticians and analysts working in this field.

The skill and ability of such teams to make a difference in their various organisations varies enormously. In some organisations analysts are relatively skilled and their management understand the business sufficiently well to produce useful models which add value to the organisation. While this is always the intention, good implementation is the key to success. A successful implementation needs to avoid some of the failures evident in other organisations such as:

- Construction of models without a clear business objective; and
- A lack of planning to enable performance measurement and monitoring of campaigns or initiatives.

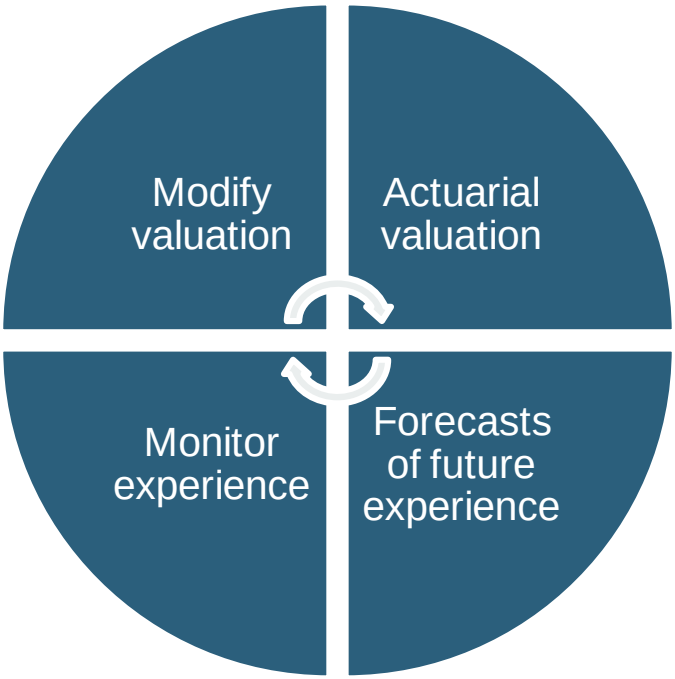
Despite the large differences in the success of data mining teams across various industries and organisations the techniques and methods are relatively well known and can be implemented within MSD as part of the recommended framework.

4 LEVEL I: POLICY DECISIONS & FINANCIAL CONTROL

4.1 Overview

Section 3 outlined the recommended framework to improve employment outcomes via three levels of analysis and control. The recommended approach for Level I requires the periodic valuation of the liability for social welfare benefits using an actuarial approach. This is the same approach as is adopted in the insurance sector. Figure 4.1 summarises the Level I framework while Sections 4.2 to 4.5 describe the approach in some detail.

Figure 4.1 **Proposed framework: Level I**



The figure above shows the actuarial control cycle recommended for implementation for the overall financial control of social welfare benefits. It requires:

- A periodic actuarial valuation of the liability for social welfare benefits;
- Forecasts of future experience based on the models and results of the actuarial valuation;
- Monitoring of the experience compared to the forecasts throughout the period between valuations; and
- At the end of the period, an updated actuarial valuation incorporating modifications due to the comparison of actual and forecast experience and other expected changes.

4.2 The periodic actuarial valuation

4.2.1 Defining the liability

To undertake an actuarial valuation of the "liability" for social welfare benefits it is first necessary to define the scope of the liability.

The insurance context

In the insurance context the liability included on the balance sheet at a particular date is usually defined to be the amount necessary to pay for all incidents that will give rise to claims where those incidents occurred on or prior to the specified date, whether the incidents are known or not. The amount of the liability is usually estimated such that it includes:

- The cost of the payments on the claims;
- The claims administration expenses necessary to administer those claims;
- Any expected future inflation of the claim payments and claims administration expenses; and
- A deduction (or "discount") to adjust for the time value of money.

The social welfare context

In the social welfare context, the definition to be used is not obvious. This is because there is no obvious "incident" at which point the liability can be considered to arise. Various definitions have been suggested for the definition of the incident including:

- Birth of a citizen of New Zealand or arrival in New Zealand of new citizens (immigrants). At this date it could be considered that a "contract" is taken out between the citizen and the Government to provide welfare benefits should the need arise.
- Attainment of age 18 (or for immigrants over 18, the date of arrival) or date of first payment for those under 18. Given that benefits are not generally payable until age 18 the definition above could be restricted generally to those aged at least 18.
- Receipt of a benefit payment at any time prior to the valuation date whether still in receipt of benefits or not. This option of the definition includes both current and past recipients of benefits. The reasoning behind this choice would be along the lines that past recipients not currently on benefits would likely have a much higher chance of receiving benefits in the future and therefore have a much higher average liability than those that have never received benefits. Intuitively this seems like a reasonable compromise between the previous definitions and the following definition.
- Receipt of benefit at the valuation date. For practical purposes this is likely to mean in the week or month prior to the valuation date. This definition could be considered the minimum definition as it includes the liability only for those people currently in receipt of benefits.

For ease of reference each of these possible definitions will be referred to subsequently as:

- Birth definition;
- Age 18 definition;
- Past & current recipients definition; and
- Current recipients only definition.

Note that on the attainment of age 65, with a few exceptions, there is no liability as the benefits within scope of this report cease at that time.

Applicable standards

As far as we are aware there are currently no actuarial or accounting standards in place dealing with the actuarial estimation of the future liability for social welfare benefits. However, based on our discussions with NZ Treasury there has been some attempt by the International Accounting Standards Board and there are ongoing discussions in this regard. In the absence of an accepted standard it would be necessary for an agreement to be reached in relation to the definition of the liability prior to a first valuation.

In addition to the question of definition of liability, standards also describe other attributes of the estimated liability. In the absence of an agreed standard we believe it would be appropriate to employ the most important requirements of the accounting and actuarial standards which are relevant to insurance liabilities. In particular we recommend imposing the following key requirements:

- The liabilities calculated should be a central estimate. Within the context of insurance contracts, this means that the value should be a "best-estimate". That is, the true ultimate cost of the claims in current dollar terms should be equally likely to be higher or lower than the estimate.
- The liability should incorporate the time-value of money. In other words, the estimate should be the value of all future cash-flows, discounted to current dollars. The rate of discount should take into account both the risk-free rate and the nature of the liabilities.
- The liability should take into account the administrative costs involved in handling the claims. Within the social welfare context, this would incorporate both the cost of case management as well as the cost of intervention programs.
- The total liability should be analysed and reported separately for each principal benefit type.
- Relevant future economic factors should be taken into account. This may include factors such as wage inflation and the unemployment rate.
- The effect of any change of basis since the previous valuation should be quantified.

Another factor that is often considered is the volatility of the estimated liability. This is often referred to as the uncertainty in the estimate of the liability. In the context of insurance, companies are required, at least in Australia, to report provisions that exceed the central estimate of the liability such that there is a 75% chance of proving adequate. In the context of the New Zealand social welfare system our view is that estimation of the uncertainty in the liability is an unnecessary complication and should be deferred, perhaps indefinitely. The requirement for insurance companies to estimate liabilities at a higher likelihood of sufficiency than 50% is due to the prudential requirements of the regulator to

protect policyholders. In the case of the New Zealand social welfare system the "policyholders" (i.e. those in receipt of benefits at any point in the future) are protected by Government and full funding is unlikely any time in the short or medium term.

Financial signals

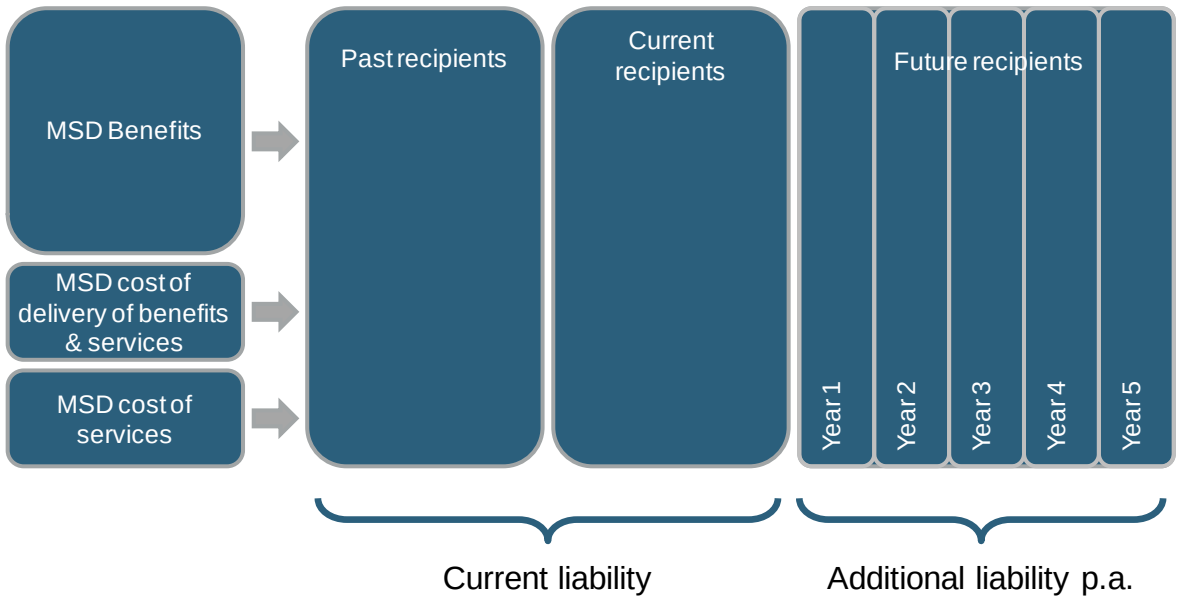
Given that there are no applicable standards to rely on it is worthwhile considering other characteristics relevant to the choice of definition. In particular:

- It is important that the estimated liability for welfare benefits can be used to send the right signals to inform strategic decision-making. In this regard the definition needs to capture changes in the cost of welfare for both current and future benefit recipients to assist in investment and other financial decisions.
- The definition chosen should, subject to meeting other criteria, be the simplest to establish.

Recommended definition

The birth and age 18 definitions provided above involve additional complexity in the actuarial valuation. Further given that a majority of the population never receive a benefit it seems unnecessary to use either of those definitions. The most obvious choice seems to be either the past & current recipients or current recipients only definition. Figure 4.2 can be used to understand the difference between these choices.

Figure 4.2 **Definition of the liability**



Under the "past & current" option noted above the liability to be estimated would include both the components for past and current recipients. Under the last option the liability would be estimated for current recipients only. In either case, the actuarial valuation would need to consider and estimate the new (or additional) liability arising from new recipients for each year (perhaps up to 4 or 5 years) into the future. Under the "past & current" option these "new" recipients would never have received a benefit previously

whereas under the "current only" option some of these new recipients would include past recipients.

Section 5 and 6 provide some detail on the analysis and modelling required for Levels II and III of the recommended framework. It is likely that from time to time a useful input to those analyses will be the liability in relation to past recipients. Furthermore, the data would not change retrospectively from valuation to valuation under this definition whereas it would change using the "current only" definition. However, one advantage of the "current only" definition is that it is likely to be easier to model.

At this stage we do not have a firm opinion on whether the definition of liability should be "past & current" or "current only". There are advantages to both and the decision is best left until some initial exploratory analysis is undertaken in relation to a valuation of the social welfare benefits.

Note that the choice of definition is further complicated by the need to determine the liability by benefit type. There are several ways in which future benefit payments can be allocated to one benefit payment type or another. For example, the liability by each benefit type could be determined as the sum of all forecast future payments for each benefit type or by the allocation of all forecast future payments for a beneficiary to the current benefit type. Other allocations are also possible.

4.2.2 Estimation of the liability

The estimation of the liability itself should be carried out using one of a number of standard actuarial techniques. To some extent this will be the choice of the valuation actuary. However, given that there are some similarities between social welfare benefits (both in duration and nature) to ACC benefits it is worth reviewing the approach that is taken for those liabilities.

ACC Scheme: estimation of liabilities

The Accident Compensation Corporation ("ACC") is a Crown entity, established by the New Zealand Government, to provide comprehensive, 24-hour, no-fault personal injury cover for all New Zealand residents and visitors to New Zealand. ACC's role is to prevent injury, treat it where it occurs, and rehabilitate people to productive life as soon as is practicable.

ACC cover is currently managed under five accounts:

- Work account;
- Earners' account;
- Non-earners' account;
- Treatment Injury account; and
- Motor Vehicle account.

The aggregate outstanding claims liability and unexpired risk liability[1] for the ACC scheme is valued annually by external actuaries and the estimate appears on the government's

[1] Note that the unexpired risk liability for ACC incorporates the "Unearned premium liability" and "Unearned premium liability deficiency" items in the accounts.

balance sheet , where it forms the bulk of the "insurance liabilities" item. At the June 2010 balance date, the ACC liabilities were valued at $27b.

Each ACC account is valued separately due to the differing dynamics and nature of claims present in each, and also due to the need to set levy rates for each account. The future payments under each account are further broken down into 12 payment types. The major payment types are:

- Weekly compensation
- Social rehabilitation – serious injury
- Social rehabilitation – non serious injury
- Medical costs
- Elective surgery
- Claims handling expenses

Each payment type is subject to different patterns of claim incidence and recurrence, as well as being impacted differently by economic factors such as inflation. As an example, we discuss one approach to valuing weekly compensation payments.

Weekly compensation refers to income replacement benefits paid to injured persons. The amount of the payment is subject to wage inflation. An individual may go on and off of weekly benefit payments repeatedly as, for example, they have subsequent surgeries in relation to the original injury. Moreover, the total benefit paid to an individual may vary as they return to part-time work and the level of their wages fluctuates below the weekly benefit amount.

A typical actuarial approach to valuing weekly compensation payments might break the analysis down into the following components:

- The number of people receiving weekly compensation benefits for the first time and the timing of that benefit payment in relation to the original injury;
- The proportion of recipients currently on benefits that will receive a benefit payment each period into the future. This can be expected to vary with the time since the original accident; and
- The average benefit amount that a person receives. Again, this can be expected to vary with the time since the original accident and whether or not the payment is a first payment.

A model of the structure outlined above is known as a Payments per Active Claim ("PPAC") approach. Within the structure there is considerable flexibility in the level of sophistication of the underlying models. These may range from judgementally selected factors to a detailed statistical model.

Valuations for each account/payment type combination are combined to give a central estimate of the outstanding claims liability for the scheme as a whole. A margin is added to the estimate to account for the possibility that the central estimate is insufficient to meet the future liabilities. This margin is calculated so that there is a 75% chance that, in combination with the central estimate, the amount reserved is sufficient to meet all future payments.

There are obvious parallels between ACC's weekly compensation benefits and the main benefits within the social welfare system. From an actuarial valuation perspective, other

benefit types under the ACC Scheme would have similar characteristics to some of the Supplementary and Hardship payments and services provided by MSD. Hence it is feasible for a valuation of welfare benefits to proceed in a similar manner to the valuation of ACC liabilities or indeed other those of other workers' compensation schemes. In the following section, we outline possible approaches to valuing the aggregate liability of the social welfare system.

4.2.3 An ACC-like methodology for valuation of the liability

The approach adopted in the ACC valuation is standard for insurance portfolios where claimants are paid benefits as long as they satisfy some eligibility criteria. This aligns with the structure of the social welfare system. We believe that an ACC-like approach is likely to be the most suitable for the aggregate-level actuarial valuation. As noted the specific approach adopted by the ACC's actuary is known as a "Payments Per Active Claim" method. Within the welfare context, one might rename this as a "Payments Per Active Client" method. We recommend using this methodology for the valuation of the main benefits of the social welfare system. Alternative actuarial methods may also be relevant for some of the smaller types of benefits (particularly supplementary benefits and hardship payments).

4.2.4 Data

A valuation of the aggregate liability would require the following items of data:

- A unit record file containing details of all past benefit recipients, such as date of birth and gender. Any data that identifies a particular individual should be stripped from this data prior to being provided to the actuary for valuation. However, each record should be assigned a unique alphanumeric identifier.
- A transaction data file containing all past benefit payments. This data would include the unique alphanumeric recipient identifier, the amount and date of the payment, the type of benefit and the period to which it relates.

We believe that MSD already collects and maintains data suitable to produce the above extracts. However, we have not reviewed sample datasets to confirm this.

In addition to the data needed to conduct the valuation, it would also be preferable to have summary reports of data from an alternative source in order to reconcile and verify summary information derived from the extracts. These sources may include:

- Summary reports produced by MSD on numbers of recipients by benefit type;
- A summary of payments from accounting type sources in order to reconcile the payment data from the extracts.

4.2.5 External versus internal valuation

In order to carry out the aggregate level valuation as described above, it is necessary to employ a team of sufficiently experienced actuaries and actuarial analysts. There is a decision to be made as to whether this team is internal to MSD or external.

While there might be some advantages to an internal team we recommend that, at least initially, an external actuarial team be engaged to conduct the valuation. Reasons for this include:

- Quantification of the liability and particularly any changes in the estimate of the liability from valuation to valuation is likely to be politically sensitive. An external team will have a greater level of independence from political or management pressure;
- A valuation of this size is likely to require a team of perhaps 5 to 10 individuals depending on the scope of the analysis. Recruiting a sufficiently skilled and experienced team of this size could take some years.
- An external team would be able to undertake the analysis immediately as they will already have in place the infrastructure and experience necessary to conduct the work;
- The nature of the work is highly seasonal. i.e. there is a large amount of work at the time of valuation and much less work during other periods. This is easier for an external consultant to manage than for MSD. For MSD it would mean hiring sufficient staff to complete the work during the peak period and then finding sufficient analysis for those staff to undertake during other periods.

In the longer term MSD might consider employing its own actuarial team. Critical to success of the team would be the skill and experience of the team leader. Note though that the Australian experience of accident compensation schemes has not seen significant internal hiring of actuarial staff. There are approximately 15 accident compensation schemes in Australia. While three of these have small actuarial teams, all of them use external actuaries to undertake their liability valuations. The perceived independence from Government is seen as a critical factor in the choice of valuation actuary.

4.2.6 Frequency of valuation

Actuarial valuations of Australian workers' compensation scheme liabilities are typically carried out every six months. Due to the scale of the work required to carry out the valuation more frequent valuations is not realistic. We recommend that in this instance valuations of the social welfare benefits be carried out either on an annual or six monthly basis. Final choice will depend on the scope and difficulty of the analysis and whether there is sufficient value to be gained by carrying out the analysis more frequently than once per annum. Depending on the underlying trends and numbers of initiatives it might be worthwhile undertaking six monthly valuations for main benefits and only annual valuations for the remainder.

4.3 Forecasts of future experience

The actuarial valuation described in Section 4.2 produces a series of forecasts by benefit type of each of the main components of the underlying analysis. Given the size of the Scheme we recommend that the analysis provides forecasts which would be monitored on a monthly or quarterly basis. The exact form of the forecasts to be provided would ultimately depend on the underlying modelling. However, it is likely to be segmented by all of the following factors:

- Benefit type
- Model component (i.e. number of active beneficiaries, average benefit payment)
- Month or quarter of year
- Duration on benefits.

4.4 Monitoring

The actuarial valuation described in Section 4.2 would be carried out annually or every six months. However, between valuations the experience of the scheme would be monitored on a regular basis. Depending on the forecasts produced, the monitoring could be carried out either monthly or quarterly. Automatic production of periodic reports would compare actual experience to the forecasts. These reports would be considerably more useful if they include some indicator of the statistical significance of differences observed between actual and forecast. A hypothetical sample report is shown in Table 4.1.

Table 4.1 **Sample report for Month of July 2012: Unemployment benefits**

Duration on benefits	Number of benefit recipients		Ratio of actual to expected	Significant difference?[1]
	Actual	Forecast		
< 1 month	12543	11946	105%	
1 – 2 months	8278	8447	98%	
2 – 3 months	5464	4967	110%	X X X
3 – 6 months	7140	7516	95%	
6 – 12 months	3335	3625	92%	
12 – 24 months	5336	5929	90%	X
24+ months	3736	4151	90%	X
Total	**45832**	**46581**	**98%**	

Note: 1. xxx indicates a result that lies outside the 95% level of confidence. x indicates a result that lies outside 90% level of confidence.

Such reports would be analysed by management and the actuary to gain an understanding of trends in the welfare system, unexpected changes in costs and as an early warning of deterioration. As evidence of changing trends emerges the periodic reporting becomes a valuable source of information to be incorporated in the next actuarial valuation. This constant monitoring allows for unexpected trends to be identified early and appropriate attention given and strategies put into place where necessary.

Actuarial monitoring of costs at an aggregate level provides an extremely useful management tool, helping to ensure that the long-term costs of the scheme are sustainable. A rigorous monitoring framework is particularly useful in the period following the introduction of a scheme or significant changes in legislation and policy, where the financial impacts of changes are not clear from the outset.

4.5 Incorporating experience into the subsequent liability valuation

At each valuation the modelling and forecasts need to be modified to take account of the emerging experience since the previous valuation. The updated valuation would allow for the changes observed or expected due to:

- External factors such as the health of the economy and the level of unemployment;
- Demographic changes;
- Policy changes;
- Trends in the behaviour of beneficiaries;
- Client level initiatives carried out to improve employment outcomes;
- Any other observed differences between forecasts and experience.

In this way, the aggregate liability figure would change to incorporate the latest experience. To the extent possible the total change since the previous valuation would be attributed to each of the various factors.

5 LEVEL II: STRATEGIC DECISIONS & PERFORMANCE MEASUREMENT

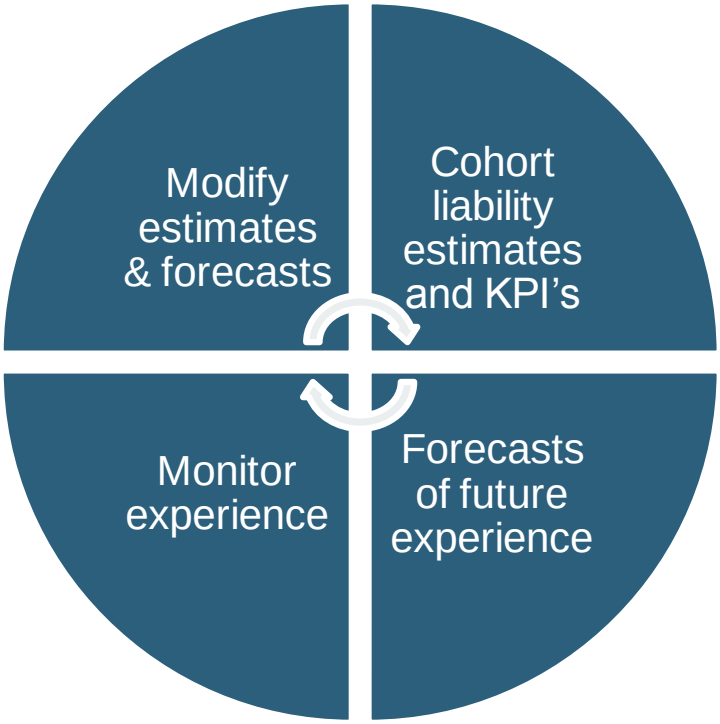
5.1 Overview

Section 3 outlined the recommended framework to improve employment outcomes via three levels of analysis and control. Level II, the cohort-level analysis provides both financial and other performance indicators for groups of similar beneficiaries to enable:

- Strategic decision-making in relation to groups of beneficiaries, particularly in relation to allocation of resources and targeting of services and interventions;
- Target setting and performance measurement of those groups; and
- Evaluation of the financial impact of Level III initiatives to inform the aggregate liability valuation and to explore cost benefit tradeoffs at the portfolio level.

Figure 5.1 summarises the Level II framework while Sections 5.2 to 5.4 describe the approach in some detail.

Figure 5.1 **Proposed framework: Level II**



Level II requires a statistical and actuarial approach to:

- Determine appropriate cohorts of benefit recipients that provide management with meaningful groups to analyse and manage. This is typically done using a segmentation analysis;
- Estimate the average future liability by cohort. This will require the allocation of the total liability from Level I to each cohort using a measure of relative cost;
- Derive and estimate a number of key performance indicators for use in the performance measurement of cohorts;
- Forecast the future liability and KPIs by cohort;
- Monitor the experience of cohorts.

5.2 Cohort Design

The design of cohorts is an important first step in the development of the Level II framework. Good cohort design will enable better strategic decision making and provide a sound basis for performance measurement. The cohorts should be designed such that the groupings:

- Provide good discrimination between the average liability for each cohort;
- Are Simple. i.e. have as few characteristics as possible to achieve good discrimination;
- Are Meaningful; and
- Are Evidence-based.

The above criteria are based on our experience of several accident compensation schemes. The following discussion is a summary of our reasoning.

Statistically meaningful differentiation between cohorts

Strategic decision making will benefit from a good understanding of which are the low and high cost groups. While these groupings may seem obvious or intuitive (e.g. unemployed youth, older disabled people, teen mothers etc.) it will be necessary to carry out a statistical analysis to search for the best discriminators or predictors of cost. These are sometimes not the obvious characteristics previously suspected.

Simple & Meaningful

The best discrimination of cost will come from an extremely complex model. While this might have some use in certain circumstances, in this case such complexity will defeat its purpose. The aim will be to determine cohort groupings that allow for meaningful analysis and management. For example, it would be difficult to consider an appropriate strategy for all:

- Single fathers with 3 children living in Auckland with no tertiary education aged 30 to 35 with past episodes of depression. This is meaningful but too complicated. While there might be an appropriate tactic for such individuals at Level III there is no place for such groupings at the strategic level.

- Unemployment benefit recipients with a risk score of 20 to 50. This is simple but has no intuitive meaning.

However, a cohort grouping of unemployed males, aged 18 to 22 who have been on benefits for 6 weeks or more would be both simple and meaningful and would ultimately prove more useful. Note that this is a hypothetical example only.

Evidence-based

A cohort-level future liability tool should be determined through an evidence-based approach, with statistical analysis accompanying the inclusion of all drivers.

Organisations that have tried to create 'expert' models for riskiness without appropriate statistical analysis have generally been unsuccessful. E.g. we have seen organisations attempt to develop a model of riskier cohorts by consulting an array of experts in order to determine the major drivers of risk and their impacts. While the models might seem correct intuitively they suffer from the inability of intuition to deal with the effect of more than two or three drivers simultaneously.

We recommend carrying out a statistical study known as segmentation analysis for each benefit type to determine appropriate cohorts. At Taylor Fry we use software by Salford Systems such as CART® or TreeNet® to carry out this sort of analysis.

5.3 Cohort level analysis

Once the various cohorts have been determined it will be necessary to undertake some actuarial and statistical modelling to estimate a range of indicators to provide insight into their performance. There are two relevant types of indicator:

- Financial, i.e. related to the average cost of benefits
- Non-financial, e.g. level of education etc

The first of these can be summarised by a single indicator, i.e. the estimated future liability. Estimation is discussed in Sections 5.3.1. The non-financial indicators are likely to be more numerous and varied and are considered in Section 5.3.2.

5.3.1 Estimation of cohort level future liability

Estimates of the average future liability by cohort will need to be modelled based on the relative costs of each cohort. This is likely to involve separate modelling of:

- Duration on benefits by cohort; and
- Average benefit paid by cohort.

Our recommendation is to build both of these models using specialised statistical techniques (e.g. survival analysis and generalised linear modelling). MSD already have some experience using these sorts of techniques and building this type of model. MSD capabilities are reviewed in Section 7.

There are four key requirements of the cohort level future liability estimate:

- To enable the total aggregate future liability to be appropriately apportioned between cohorts. This will be used for to provide accountability and performance measurement at the cohort level;
- To enable the evaluation of Level III initiatives via a cost / benefit analysis (see Section 6.2.6)
- To measure the impact of an MSD project or initiative on the aggregate level future liability; and
- To be relatively simple to update to allow regular re-calibration to reflect changes in the aggregate future liability.

5.3.2 Estimation of cohort level KPIs

The first stage of this part of the framework would be to determine the appropriate KPIs for estimation. This would involve consultation within MSD to determine which are the non-financial indicators that are most relevant to assist in managing beneficiaries and in understanding the success or otherwise of initiatives. Some further discussion and research is needed to develop a good set of indicators but they are likely to include figures related to workforce participation, and possibly a range of health, education and justice outcomes.

Once the KPI's are agreed it would be a matter of analysing the data to determine current and historic average KPIs to provide an understanding of any recent trends and for forecasting future expectations.

5.4 Forecasts of future experience & Monitoring

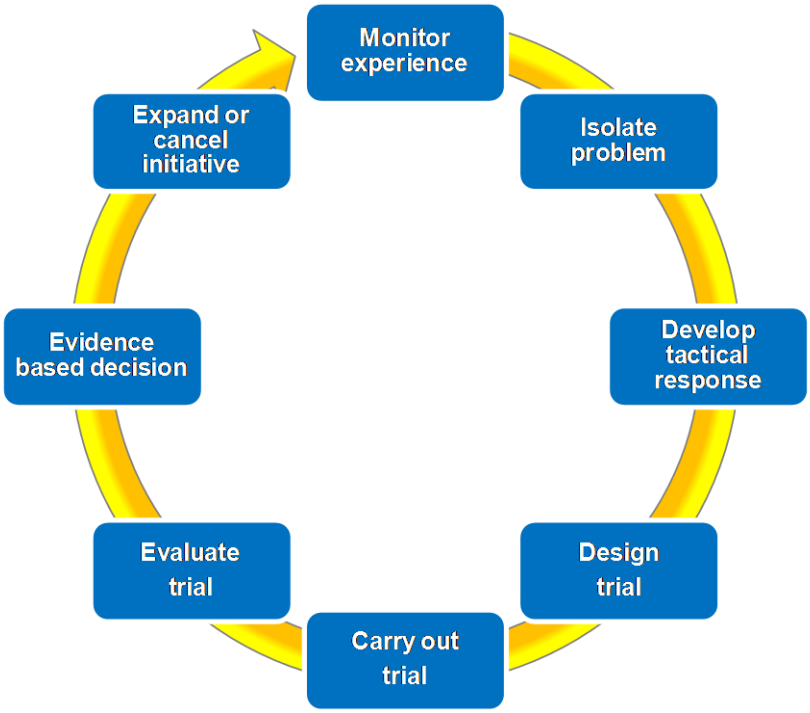
The cohort level analysis described in Section 4.2 will produce a series of forecasts by cohort of each of the main components of the estimated future liability and the KPIs. The exact form of the forecasts to be provided would depend on further discussion and exploratory research. For example, a useful KPI might be the percentage of beneficiaries remaining in work 52 weeks after going off benefits. Targets could be set based on historical averages and trends by cohort. Monthly reporting would compare experience to targets.

6 LEVEL III: TACTICAL DECISIONS & CLIENT INITIATIVES

6.1 Overview

Section 3 outlined the recommended framework to improve employment outcomes via three levels of analysis and control. The recommended approach for Level III requires the development and evaluation of tactical initiatives to drive behavioural change amongst benefit recipients. Figure 6.1 summarises the Level III framework while Section 6.2 describes the approach in some detail.

Figure 6.1 **Proposed framework: Level III**



6.2 Level III Process

Level III of our recommended framework is the component which drives improved employment outcomes to reduce the cost of social welfare. This is done by the translation of innovation to concrete initiatives to improve employment outcomes and evaluation of those initiatives. This is in contrast to Levels I and II which enable increase understanding of the overall cost, provide accountability, transparency and financial control, performance measurement and enable strategic decisions relating to allocation of resources. Therefore, it is important to develop a framework and that drives the innovation at Level III. The following sub-sections outline our proposed structure for this approach.

6.2.1 Monitor experience

The aim of monitoring experience as part of the Level III framework is to be able to develop initiatives to improve employment outcomes in the social welfare system whether widespread or particular to a small group of beneficiaries. In this context we take "monitoring experience" to have an extremely wide definition. It encompasses not only experience of the social welfare system in NZ, but also all related international experience, relevant academic research (international or local), and indeed anything innovative that is thought to be capable of improving employment outcomes.

Monitoring scheme experience

The monitoring reports set up as part of Levels I and II would provide a starting point for some of the analysis at Level III. As an example, a review of experience at the cohort level may find that the experience for a large group has deteriorated. However, the cohort analysis would not be able to inform the specific nature of the deterioration. Hence it is important to put in place a process for analysing in greater depth any issues that arise out of Level I and II monitoring. This may be via further review of data or by surveys of recipients, case workers or service providers.

Other sources of innovation

There will also be the need to undertake additional analysis and surveys, continual reviews of academic literature and international experience and, for example, engagement with key stakeholders. The aim will be to discover any issue or idea that can be used to improve employment outcomes.

In summary, at Level III the challenge will be to understand what is happening at the client level and what innovative approaches can be taken to improve their employment outcomes.

6.2.2 Isolate the problem

The result of the continual monitoring of experience described above will be the discovery of issues within the social welfare system that present the possibility for improving employment outcomes.

6.2.3 Develop a tactical response

Once issues are discovered it will be necessary to develop a tactical response which may improve the employment outcome. This may involve a literature search, a discussion with other social welfare and disability schemes, brainstorming within MSD, advice from external providers etc.

6.2.4 Design trial

Having decided on one or more tactical response(s) it will be necessary to determine:

- Whether the approach is indeed successful; and if so

- Which outcomes are improved and by how much;
- The effect that the approach might have on the portfolio liability; and
- Whether the approach is justified in cost / benefit terms.

The best way to have an accurate understanding of the value of an initiative will be to undertake a controlled trial.

There are 3 possible approaches that could be used by MSD to evaluate initiatives:

- **Random or pseudo-random controlled experiments** where subjects are randomly or near-randomly allocated to an intervention group or a control group and then observed.
- **Matched population experiments** which are a slight variation on the random controlled experiments. In this case stratified groups are defined and a population of subjects is allocated randomly to an intervention group or a control group and then observed.
- **Retrospective cohort studies** where groups of subjects are defined at a point in time – based on the outcome – and followed from that point in time to observe the factor of interest. This requires longitudinal data.

Random or pseudo-random controlled experiments are the best means to evaluate initiatives. This method isolates the intervention itself as the only distinct factor between the intervention group and the control group. Thus, it can be determined conclusively the impact that intervention has on a population. Designed properly, it will also reveal the segments in each group that respond best to the intervention. We recommend this level of evaluation whenever possible.

However, we understand that controlled experiments may be considered undesirable in certain situations. Whilst we still urge use of the random controlled experiment whenever possible, MSD may alternatively design:

- Concurrent experiments, where each population is subject to a different intervention rather than one with and one without. The experiment than judges the superior of the two. The flaw in this experimental design is that you do not measure the difference from not intervening – the cost/benefit. Also, one would likely need some prior evidence that both interventions have promise.
- Matched population experiments where two regions with similar risk characters and demographics are the intervention group and the control group. This is indistinguishable from a phased roll-out of the intervention, so is less likely to be considered undesirable. In all other respects, it is the same as the random controlled experiment. The difficulty lies in ensuring sufficient similarity in the two groups. This can be dealt with by stratification prior to random selection.

We recommend that the matched population experiment is the minimum standard experimental design for the initiatives to be evaluated and to enable the financial impact to be assessed for incorporation within the Level I and II estimates of liability. Evidence from a retrospective cohort study is usually too weak to be able to determine adjustments to aggregate and cohort level liabilities.

6.2.5 Implementation of the trial

Once the design of the trial is determined it will need to be carried out and evaluated. Due to the need to observe clients over time, this phase of each initiative is usually the most time consuming. This increases the importance of well-considered objectives and design prior to implementation.

During this phase performance tracking and monitoring of the initiative will be important as it will indicate when:

- The design phase has not adequately controlled for other risk factors or the environment;
- The service providers and case workers are not carrying out the initiative correctly; and
- The implementation of the initiative has caused the behaviour of clients or service providers to change in a way that affects its effectiveness.

It will also provide an early indication as to the level of effectiveness of the initiative and indicate when sufficient data has been collected to properly evaluate the initiative.

6.2.6 Evaluation of the trial

Once sufficient time has elapsed and data collected the impact of the trial needs to be determined in a statistically rigorous manner. Ultimately the questions that need to be answered include:

- Whether the approach has been successful; and if so
- Which outcomes have been improved and by how much;
- The effect that the approach might have on the portfolio liability; and
- Whether the approach is justified in cost / benefit terms.

The answer to the first question will require an analysis of both the financial outcomes and other social outcomes. The financial impact can be determined using a cost / benefit analysis where the cost is the average cost of implementing the initiative for each beneficiary and the benefit can be evaluated by determining the savings in future liability using the estimated liabilities from the cohort level analysis. Social outcomes can be tracked via changes in the cohort level KPIs.

Once success or failure has been determined the initiative can be implemented more widely or cancelled. It will be important to keep track of successful initiatives to inform the subsequent Level I actuarial valuation. This is very important as without this process of informing the high level liability the feedback would be lost and the ability to achieve overall financial control would suffer.

7 REVIEW OF MSD CAPABILITIES

7.1 Summary

Generally MSD is well placed to implement the framework recommended in this report. A summary of MSD capabilities and general recommendations in relation to each of the levels of analysis is provided in Table 7.1.

Table 7-1 **Summary of MSD capabilities**

	Personnel capabilities	Recommendation
Level I **Actuarial valuation**	Currently MSD has no actuarial team in place to undertake this part of the framework. MSD do have a forecasting team that provide short term forecasts of numbers of beneficiaries and amount of payments.	At least in the short to medium term appoint external advisers to undertake the required analysis. The actuarial valuation should take account of the MSD forecasts of numbers of beneficiaries and amount of benefit payments.
Level II **Cohort statistical analysis**	Currently MSD has a team of staff with the skills necessary for carrying out this sort of work on an ongoing basis.	Initially it is likely that the team will lack some of the experience necessary to implement the framework suggested. We recommend that MSD partner with an external consultant to build the initial framework and further develop the MSD team's experience and capability.
Level III **Recipient modelling and trial evaluation**	Currently MSD has a small team of staff with the skills necessary for carrying out this sort of work on an ongoing basis.	Initially it is likely that the team might benefit from the experience of an external consultant who has carried out these sorts of projects to: a) Develop the rigorous framework for the approach, including modelling, experiment design and evaluation; and b) Carry out 2 or 3 initial projects to embed the process within MSD.

In addition to the personnel capabilities there are also data and systems requirements. In relation to data it seems that MSD retain the appropriate data. However, it is likely that some development will be required to provide data in a form suitable for both the Level I and Level II framework. In relation to systems we are not aware of MSD's current infrastructure nor do we have sufficient expertise to judge its adequacy. However, it is likely that some infrastructure development will be required to provide the IT resources necessary for the both the Levels II and Level III framework. In addition IT development will be required to disseminate Level II information to frontline caseworkers. This is likely to be one of the larger projects necessary to implement the proposed framework.

A detailed review for each of the recommended levels of analysis follows.

7.2 Level I: Actuarial liability valuation

MSD have been developing a methodology to estimate future liability through the use of a client based benefit projection tool. We discuss the specifics of this technique in Section 7.3. We believe this tool was initially designed with two objectives in mind:

- To provide estimates of future liability for individuals; and
- To provide an estimate of the aggregate-level future liability.

While not immediately obvious, these objectives are considered to be more or less incompatible. In relation to estimates of future liability for individuals MSD's current methodology is discussed further in Section 7.3. In relation to its use for an aggregate liability valuation we make the following comments:

- Estimates from cohort- or client-level models are driven primarily by individual risk characteristics. Hence, these types of models are good at informing the drivers of risk at an individual level, with models that are reactive to the changing effect of these drivers. However, this reactivity can cause unwarranted volatility at the aggregate level. Hence it is likely that at the aggregate-level the estimated future liability would be relatively unstable.
- At a cohort- or client-level, it is difficult to assess portfolio-level influences such as economic cycles, demographic changes and social trends above the individual risk characteristics. Hence important trends and changes are likely to be missed.
- The current methodology does not appear to allow for new clients or recurrent clients i.e. projections cannot include the entire client portfolio when forecasting forward. Thus funding and monitoring attempts would need to be made on incomplete information about the liability.
- The current methodology includes only a limited allowance for economic drivers and no allowance for demographics. This limits the ability to use the tool to forecast the impact of adverse economic or demographic scenarios. i.e. the ability to undertake scenario testing for policy changes is limited.
- The cohort- or client-level tool is an atypical way to determine an aggregate-level liability. There are existing methodologies that are broadly accepted and would provide a more robust estimate of the aggregate-level future liability. Thus it is unlikely that the cohort-level tool would meet recognised actuarial or accounting standards for the determination of an aggregate liability.

In summary, using MSD's current methodology to value to aggregate liability is not recommended. Nevertheless the approach could be adopted, with some modification, for the Level II analysis which is discussed further below.

7.3 Level II: Cohort future liability

In this section we discuss MSD's capabilities in regard to the cohort-level future liability by identifying the tools that are currently in use or in development. Our advice is in relation to the feasibility of adapting these tools to the structure recommended – in terms of both personnel and resources.

Technical aspects are not discussed here but are considered in Appendix B. These aspects are not relevant to the feasibility of adopting MSD's approach but may be relevant to successful implementation.

7.3.1 Future Liability

A client-level future liability tool has been developed by MSD. We understand that initially it was envisaged that this future liability tool would serve multiple objectives:

- To provide estimates of future liability for individuals; and
- To provide an estimate of the aggregate-level future liability.

The conflict in meeting both of these objectives with a single model was discussed in Section 7.2.

We appreciate that some facets of MSD's future liability model would not continue to be developed if the recommended approach to a separate aggregate-level future liability is undertaken. Thus, we limit our analysis here to the model components relevant to the cohort- and client-level structure and aggregate-level future liability disaggregation. We provide further dissection of MSD's current model in Appendix B.

The approach used to estimate client-level future liability is represented in Figure 7-1. The Likelihood of Long-Term Benefit Receipt (LLTBR) model is used to determine the cohorts. Each of these cohorts has a duration model calibrated to determine its expected future time on benefits and a benefit model overlaid to determine the cost of this time on benefits.

Figure 7-1 **MSD approach: estimation of future liability**

LLTBR model for cohorts by risk score

The Likelihood of Long-term Benefit Receipt (LLTBR) score aims to:

- Estimate the probability that that a client remains on benefits long-term;
- Be always up to date;
- Be low-cost and be able to be generated automatically; and
- Be flexible enough to incorporate new information and meet operational requirements.

In its current implementation the model provides the rank of the riskiness of clients i.e. it is a client risk score approach to cohort determination.

The model is discussed in greater detail in Section 7.3.2. However, it is worth commenting here that this is a reasonable approach to use but that as suggested in Section 5.2, we believe a cohort design that is based on meaningful characteristics (rather than a statistical score) would provide wider application. Hence we do not recommend this model for determination of cohorts.

Benefit duration model by cohort

The survival analysis approach taken for modelling benefit duration model is generally reasonable. However, the current execution produces a relatively high level of uncertainty. This is principally due to the extrapolation of 48 months of client observations to almost 50 years. In its current form it is difficult to measure its accuracy and it is likely that it will deteriorate relatively quickly over time. However, the same approach with a slightly different treatment and assumptions, would allow a calibration over the entire MSD data history – currently 18 years – and increasing to 50 years as experience emerges over time.

Alternatively other data sampling methodologies for constructing long-duration models that would likely provide greater accuracy could be used.

In summary adapting the current methodology is feasible. Certain facets of the modelling could be improved resulting in a more stable prediction. An expanded review of the modelling is provided in Appendix B.

In addition to the comments on the modelling itself we note that we are uncertain as to the degree of testing of the benefit duration model. It is important that extensive diagnostic tests are performed on the model calibration as well as an evaluation of its performance on a holdout sample dataset. Furthermore, a process for monitoring the performance of the model over time should be put in place prior to implementation.

Model of time in receipt of supplementary assistance

The model of time in receipt of supplementary assistance is derived as a proportion of the clients on main benefits.

We have a number of concerns in relation to this model which are outlined in Appendix B.5.3. In particular:

- The structure of this model seems counterintuitive. It would seem more natural to model the proportion of clients on supplementary benefits as a proportion of

clients that are off main benefits (rather than as a proportion of clients on main benefits). While this might give more intuitive results it is unlikely to have any impact on the model.

- The extrapolation of 4 years data to, in some cases, 40 years or so is likely to introduce instability in the model and reduce its accuracy. Given the uncertainty involved it would be preferable to develop a technique that required less extrapolation.

Model of income support entitlement

The model of income support entitlement is based on observed average payments summarised according to:

- Benefit type;
- Number of children;
- Age group; and
- Partner status.

Projections of benefit payment amounts are derived using the starting averages described above with transitions to different benefit amounts based on observed transitions to other benefit types over the initial four years as well as age, and cessation of dependent children based on age of youngest child.

While it is difficult to comment on the approach given the information provided there are some simplifying assumptions that have been taken that may have a material impact on the results:

- No transitions are assumed after the first 4 years;
- Benefit rates based on the number of children assume clients have no more children; and
- Change in benefit rates appears to occur when the youngest child reaches a 18 rather than changes in benefit as each child reaches 18.

Future liability reference table

The output of the benefit duration model in development is a future liability reference table from which a user can look up a client's future liability by cross-referencing LLTBR cohort, age, gender, benefit type, partner status, number of children and age of youngest child. There are a few minor concerns in relation to the final results in the liability reference table and these are described in Appendix B.5.5.

Summary

As noted in Appendix B.4.3 the DFL tool described above provides average estimates of liability at the cohort level where each cohort is defined by LLTBR, age, gender, benefit type, partner status, number of children and age of youngest child.

The recommended framework in this report requires liability estimates calculated at cohort level (i.e. Level II of the framework, see Section 5.3). Based on this proposed framework a modified version of the DFL tool may be appropriate for use within Level II. Critical to this decision will be the extent to which the cohort segmentation determined to

be most appropriate for Level II coincides with the cohort segmentation used in the DFL tool. It is likely that in addition to the various model modifications suggested in this Section and Appendix B that there will be the need to redevelop the model using a revised segmentation including benefit type as a predictor of duration.

7.3.2 Likelihood of long-term benefit receipt 2010 model

The LLTBR was initially developed as a client-level tool. For this purpose it remains appropriate. In developing the future liability, MSD currently use the LLTBR to determine cohorts. That is, the LLTBR is the risk score. The LLTBR model ranks risks based on risk-profiling models that have been in production at MSD for some time. The model outputs the ranked likelihood of being on benefits two years from present. To determine this, it uses clients' risk characteristics.

The LLTBR appears to be competent at discriminating between risks. Moreover, the processes followed by MSD in testing this model are good practice. We have some suggestions regarding model specification, but these do not affect the feasibility of using the model and are discussed in Appendix A.

LLTBR by benefit type

Currently, the LLTBR does not distinguish very well between risks within a benefit type. For example, Unemployment Benefit clients have LLTBRs under 15 per cent, whereas Invalid Benefit clients have LLTBRs over 90 per cent. In our opinion it would be preferable to retain the separate risk scores by benefit type

7.3.3 Personnel and resources

The time and resources required to construct, test and implement the cohort-level future liability is dependent on the degree of departure from MSD's current models.

MSD appear to have the capabilities to construct and monitor the cohort-level future liability models. The current projects demonstrate an adequate appreciation of the techniques required to build successful models. Some of the recommendations made in this report may be new to MSD staff. However, after consultation with MSD staff, we believe that any need for external providers in regards to the cohort-level analysis would only be for transferring knowledge to internal resources. That is, we do not see a need for extensive external involvement in maintenance of a cohort-level future liability framework on an ongoing basis.

During the initial development phase of cohort level models MSD would likely find that some external advice might be helpful to speed up the development of the framework and build on the capability of the internal team.

7.4 Level III: Client-level analysis

Currently, MSD have many means of intervention in a client's case. We have seen numerous tactical tools used, such as the LLTBR, and others under development. It is not in

the scope of this feasibility study to make recommendations or review each MSD targeting tool. Instead we have restricted ourselves to a high level review and offer the following comments:

- The **Modelling Process** was examined in the case of the LLTBR. The process adopted was robust and good practice. In particular we note that MSD tested the performance of the model on a holdout data set. We consider this to be of critical importance to a strong modelling process. We have reviewed this model in detail in Appendix A where we offer some suggestions for its improvement.
- We believe the **Implementation Process** could be improved. In particular we recommend a robust framework of controlled trials and evaluation be implemented.
- Despite only limited interaction with MSD analysts the **Technical Capabilities** of MSD appear relatively strong. Our review of the LLTBR and future liability models, has given us confidence that MSD have strong technical capabilities. We believe the client-level tool structure that we recommend could capitalise on these capabilities.

8 FUNDING

8.1 Background

As noted in Section 2 we have been asked to provide "*advice on the level of funding that would be required to make partial funding feasible through cyclical variations".*

In the context of the social welfare system funding here refers to the payment in advance of monies to build up a fund to pay future welfare payments. Under the current system there is no forward funding and welfare benefits are paid when they are due from general revenue. This is known as "pay-as-you-go".

The requirement to provide advice on the level of funding necessary to achieve a feasible level of partial funding is difficult without undertaking some sort of financial projection including forecasts of future benefit payments which is outside the scope of this report.

We have therefore restricted our advice to some general comments on funding and some hypothetical examples. Furthermore, it should be noted that our expertise is relevant to the detail of how much to fund and over what period to achieve a certain level of funding i.e. questions related to the detail of funding projections. The question of whether to fund or not and at what level is generally outside our area of expertise although we have made some general comments in relation to these questions below.

8.2 The question of funding

The issue of whether to fund a social welfare scheme is a question of high level policy and national economics. As noted, neither of these fields are within our expertise. Nevertheless the following comments are relevant to the discussion.

8.2.1 Accountability

The existence of a fund (even a partial fund) would very likely give increased emphasis to the amount of the liabilities and to the change in liabilities from period to period. This would be an outcome of setting targets of a particular level of funding which would be affected by the amount paid into the fund **and** the change in the total liability. Analysis of the change in funding level from period to period could be used as one of the prime targets for accountability.

This is probably the strongest reason for implementing some level of funding.

8.2.2 Fairness

Intergenerational equity is often given as a reason for funding. For example, it seems "fair" that each generation pay for its own social welfare benefits. However, it is not a simple matter to determine a fair allocation of the cost of the social welfare system.

Matters such as population ageing, increasing lifetime earnings of each generation (due to productivity growth and increasing workforce participation rates for females), progressive tax policy (i.e. the principal that higher earners pay more than lower earners) and in fact all the other transfers between generations (e.g. technology, changes in the environment, infrastructure) mean that any conclusion about the fairness of the costs of funding the social welfare system cannot realistically be determined.

8.2.3 Feasibility of full funding

The social welfare system currently provides payments of the order of $10 billion per annum including expenditure in relation to employment services and supports and the costs of administration. The liability for the current recipients will be several times this amount. It is therefore not feasible to consider fully funding this liability even in the medium term. The only real option for fully funding the system might be to fully fund all new recipient payments and continue with PAYG for current recipients.

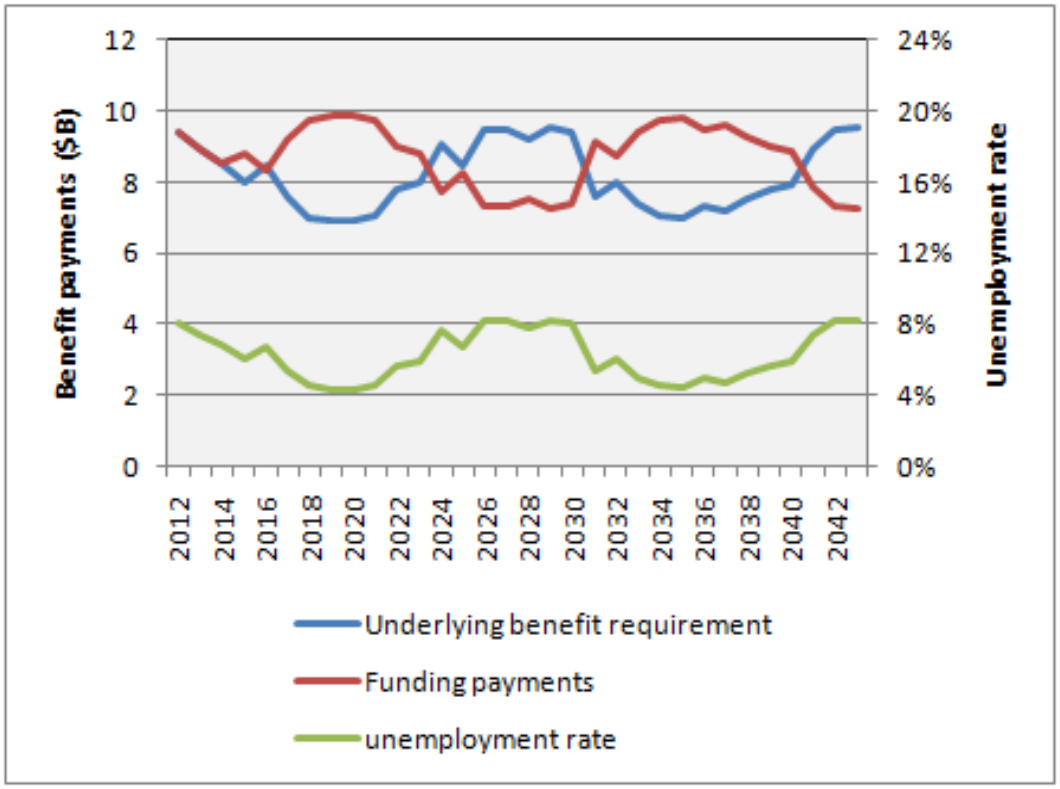
8.2.4 Economic reasons

The existence of a fund could be used to dampen or even offset the impact of changes in the economy. E.g. in poor economic periods the payment of benefits could be subsidised by drawing down on the fund while the reverse could occur in prosperous periods. Figure 8.1 highlights the possibilities.

Also payment into the fund could be made either from existing revenue or by an additional levy of some description, should government wish to consider changes of this nature at some point. The latter would be a form of additional forced savings which may be considered a prudent economic outcome.

While these ideas seem reasonable in isolation, in fact they form just a part of the overall annual NZ budget. Thus to the extent that funds are diverted from other organisations, or from retiring Government debt, the question becomes one of what is the best use for the funds. For example, it may be that funding can take place but that this will require additional government borrowings. Such alternatives would require consideration by the various Government Ministries, particularly Treasury.

Figure 8-1 **Hypothetical example of counter-cyclical funding for social welfare**

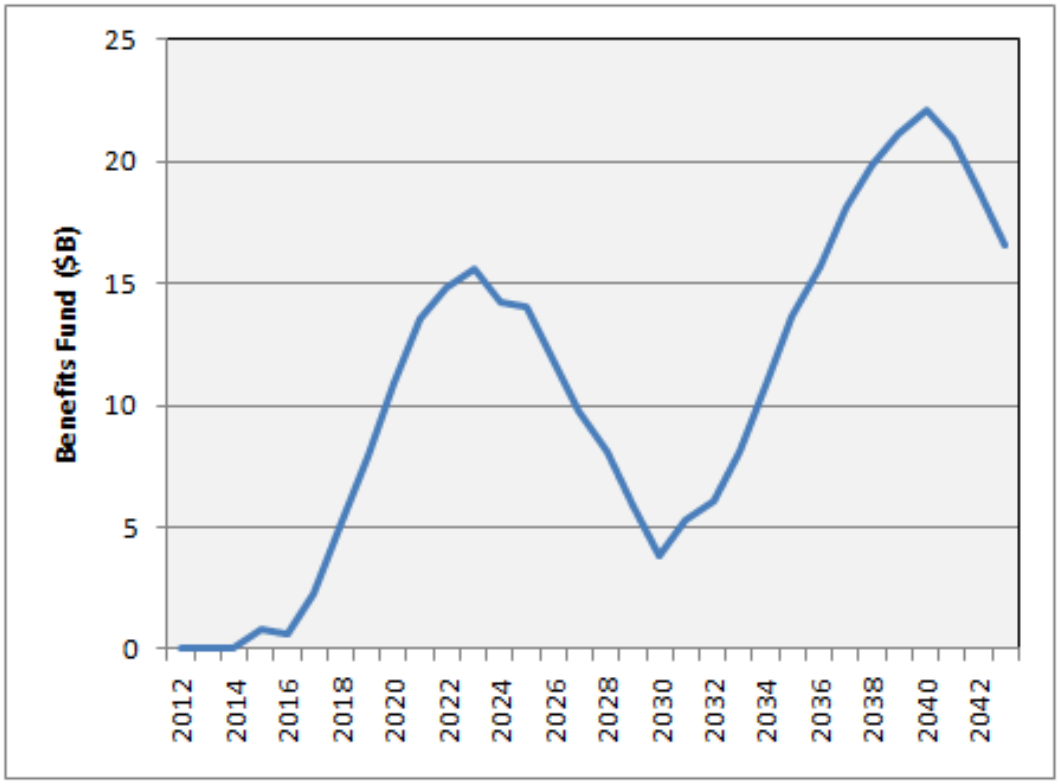


The figure above provides a simplified example of a hypothetical counter-cyclical funding approach for the social welfare benefits. In the example the green curve shows a hypothetical forecast of the unemployment rate as a proxy for the health of the economy. It can be observed that in poor economic times (high unemployment) the underlying requirement to pay benefits is higher than in prosperous times (see the blue curve). This need to pay higher amounts during poor economic times is a double hit to the economy as there is less government revenue but higher expenses.

A countercyclical approach might attempt to reverse this double hit by paying more when times are prosperous and less when the economy is poor, thereby enabling the government to spend more on priming the economy when needed. While the same could be done by financial management of government funds (borrowing more in poor times and paying down debt in good times) this method may provide an alternative. The red curve in the figure above represents a plausible funding arrangement using this approach to counter the economic cycle.

The resulting build up and draw down of the fund under this approach for this hypothetical example is shown in Figure 8.2.

Figure 8-2 **Hypothetical example of funding level**



In any event, we believe the question of funding is not really relevant to the recommended framework. That is, **the approach recommended in this report can be implemented with or without funding.**

The heightened awareness of the amount and change in liabilities could still be brought about, for example, with appropriate targets for senior management in relation to a notional level of funding or the total amount of the liability.

8.3 Funding details

Should the Government decide to fund the social welfare benefits system, either partially or fully, some options in relation to the speed and level of funding would need to be analysed by undertaking forecasts of the various cash flows.

8.4 Investing the funds

Should a funding option be followed the Government will need to consider investment options for the funds. Given that the ultimate size of even a partially funded Scheme is likely to be in the billions of dollars serious consideration will need to be given to the investment strategy.

Note also that the management of the funding pool itself, including how it is invested, would affect the level and volatility of funding.

9 IMPLEMENTATION

9.1 Background

As noted in Section 2 we have been asked to provide advice in relation to "*any significant implementation issues or barriers that would need to be addressed before a liability approach could be adopted".*

While we have significant experience in the design and implementation of business modelling projects our expertise does not extend to the sort of advice given by management consultants in relation to management structure, cultural change etc. Nevertheless we attempt to cover a wide range of potential issues in the discussion that follows.

9.2 General risks and issues

The following is a list of risks and issues relevant to the entire approach and framework. Later sections deal with specific components of the advice.

- The framework requires commitment and buy-in at all levels, from politicians, senior management, analysts and case workers. Without this buy-in the framework will do nothing more than provide an estimate of the social welfare liability.
- While the ultimate aim is to reduce the financial cost of the social welfare system the primary focus needs to be on improving employment outcomes. Provided this focus remains, the overall financial cost will reduce and social outcomes improve. A focus on reducing social welfare costs without improving employment outcomes may have adverse consequences in areas such as health (particularly mental health) and justice.
- To achieve significant behavioural change amongst beneficiaries not currently participating in the workforce (i.e. to increase the rate of return to work) would require a change in their obligations. This could potentially be achieved through the single core benefit (i.e. "Job Seeker Support") approach recommended by the Welfare Working Group, or by changing the existing legislation to require similar sorts of obligations on recipients as is required by the unemployment benefit.
- The financial framework itself is not a silver bullet. The framework enables management to control the system by providing the right information to make better decisions.
- While each level of analysis and control could be implemented separately the best result will come from implementing the entire framework.
- This new approach will require significant investment of time and resources. While this is minor in comparison to the welfare benefit system, additional budget and dedicated teams will be needed to carry it out.

- Building the right framework (models, analysis and reporting) for each level II takes time. However, it is important to do it well and the initial development phase should not be rushed.
- Significant cultural change is required at several levels:
  - The biggest change will be around work expectations for most benefit recipients. Case workers will be at the forefront of dealing with this change. Community attitudes will also be important.
  - The level of financial control and the rigorous nature of the modelling and monitoring and the requirement for controlled experiments will require a cultural shift within MSD managers and analysts.
- A successful implementation of the proposed framework will require that business processes are aligned with the desired outcomes. In most respects this is not within our core expertise to comment. However, in the case of business modelling where we do have significant expertise it is fair to say that we often see instances of business models developed which do not align with the purpose for which those models had been developed. Considerable effort should be given to planning and design of the various models, analysis and reporting to be developed to ensure that it meets the required objectives.

9.3 Level I framework

Issues specific to the proposed Level I framework (i.e. the aggregate actuarial valuation of welfare liabilities and subsequent monitoring) include the following.

- There is the potential for the actuarial estimate of the welfare liability to be subject to pressure from senior management and politicians. Particularly in the quantification of the change in the liability estimate from year to year and in determining the causes of change. This pressure is likely to be more difficult for an internal actuarial team.
- MSD currently have an econometric modelling team which provides forecasts of numbers of benefit recipients and amount of benefit payments. The actuarial valuation will provide similar forecasts. However, the use of the two sets of forecasts is very different and each will have its strengths and weaknesses. The actuarial valuation will focus on long term payment forecasts whereas the econometric team will be focussed on short term forecasts. It is likely that the actuarial team will benefit from access to short term forecasts and could be used as an input to the valuation. Management will need to understand the differences and use the various forecasts where appropriate.
- Care needs to be taken in the design of the liability valuation approach. While there are many methods and techniques that can be used some will integrate better with the cohort approach and provide better information for policy decisions.
- The actuarial valuation itself needs to be rigorous and based on high quality modelling. The information to be gained from the financial control of the welfare system at the aggregate level will depend on an accurate recognition of the various underlying trends. This will enable a better understanding of the drivers of the cost of welfare and hence improve decision making.
- As far as we're aware there are currently no actuarial or accounting standards relevant to the estimation of social welfare benefits. It will be necessary to

agree to some appropriate definitions and scope prior to undertaking the actuarial valuations. We would recommend that from an actuarial perspective either the relevant Institute of Actuaries of Australia or New Zealand Society of Actuaries professional standards be applied to the extent possible. This would cover matters such as the need to determine the liability on a central estimate basis, to include the cost of administering the benefit system and to allow properly for inflation and investment earnings. There are also standards in relation to the detail that needs to be included in reporting. Perhaps the most difficult issue will be to decide the scope of the liability in regards to past, current and future benefit recipients. Some discussion of this has been covered in Section 4.2.1. However, while this is a difficult decision, ultimately any of the recommended definitions will be reasonable and should not impact the usefulness of the framework.

9.4 Level II framework

Issues specific to the proposed Level II framework (i.e. the cohort actuarial liability, KPIs and monitoring) include the following.

- It is important that the Level II selection of cohort segments is carried out using formal statistical techniques. Human judgment applied to similar situations is typically sub-optimal.
- The selection of key performance indicators need to be meaningful for frontline staff.
- Avoid unnecessary complexity in the design of the cohort indicators to allow ease of use for frontline staff and MSD management.

9.5 Level III framework

Issues specific to the proposed Level III framework (i.e. the statistical modelling of client behaviour, innovation and initiatives and evaluation of trials) include the following.

- It is extremely important that the Level III framework includes a rigorous approach to evaluation of initiatives (controlled trials). Without controlled trials a significant benefit of the framework would be lost.
- Trial design itself must be rigorous to allow confident decision making and evaluation of financial impacts.
- The process of finding problems and developing solutions is a creative one. The team of staff leading this part of the approach needs to work continuously looking for innovative solutions to get beneficiaries back to work.
- Failure of some trials is to be expected and is part of the process. This may require some cultural shift among senior managers. In fact, if trials are not failing it probably means not enough are being carried out.
- While statistical analysis of individual client behaviour is a very useful tool, care needs to be taken in order to manage personal information in a sensitive and secure fashion.

9.6 MSD capabilities

Issues specific to MSD's capabilities include the following.

- At Level I MSD either needs a new team or an external consultant. An internal actuarial capability is difficult to develop in the short term. In either case the actuarial team needs appropriate qualifications, skills and experience.
- MSD currently has some Level II & Level III tools and capability under development. During the development phase it's important to set up appropriate methodologies and framework. This will improve the likelihood of success via rigorous analysis and decision making and enables knowledge transfer between staff.

9.7 Funding

Issues in relation to funding have generally been covered in Section 8. In summary they relate to:

- What level of funding is to be targeted?
- What is the time horizon to achieve that target?
- Issues in relation to the investment of the funds including choice of investment vehicle, investment strategy (e.g. asset allocation, hedging strategies, duration matching etc) and investment managers; and
- Whether funding level responds to the health of the economy.

While the single most important issue in relation to funding is whether to fund or not, the question of funding is not really relevant to the recommended framework. That is, **the approach recommended in this report can be implemented with or without funding.**

APPENDIX A LIKELIHOOD OF LONG-TERM BENEFIT RECEIPT TECHNICAL REVIEW

A.1 Scope

This review is based on the following documents provided by MSD:

- A report titled "Likelihood of Long-Term Benefit Receipt 2010 model: Technical report" dated February 2011 and prepared by the Centre for Social Research and Evaluation;
- An MSD memorandum entitled "LLTBR 2011 v1 update project" dated 23 February 2011 from Jared Forbes and Marc de Boer; and
- An MSD Project Brief for the project "LLTBR 2011 model update" dated 28 March 2011 and prepared by Jared Forbes and Marc de Boer.

Our review of these documents has been augmented by discussions between Ash Evans (of Taylor Fry) and various MSD staff, including Marc de Boer on 22 June 2011.

However, our review is necessarily more restricted than a full technical review. In particular, we have not reviewed:

- The code and processes used to create the modelling datasets and the checks conducted on the data, including reconciliations of these to independent sources;
- The final model specification, including parameter estimates;
- Full model diagnostics on a "hold out" sample of the data including
  - Actual versus expected analyses for various subsets of the data;
  - Actual versus expected analyses across different variables within these sub-populations; and
- Results for alternative modelling techniques applied to the data.

A.2 Model purpose

A.2.1 General

Our understanding of the primary purpose of the model is that is intended to help case managers decide the right level of support for each client. To do this, it gives each client a score between 0%-100% which, broadly speaking, reflects the probability that this client will still be on benefits in two years.

Before we review the technical aspects of the model, it is worth considering how useful this score will be in achieving its stated purpose and whether there are alternative tools which may achieve that purpose. It may well be that such issues have been already considered in detail by MSD. Nevertheless, we include a brief discussion below the definition of risk used for the model and the merits of complex versus simple models.

A.2.2 Risk definition

The LLTBR gives a single risk score across all benefits. One of the consequences of this is that case managers will see, for instance much higher average risk scores for the Invalid's benefit than the Unemployment related benefit. Other things being equal, and providing that case managers have sufficient authority, the model will drive more support for recipients of the Invalid's benefit and less for recipients of the Unemployment related benefit.

Such an outcome might well be one of the planned consequences of the model. However, it might equally be that there has been a higher level decision (e.g. a policy or strategic decision) regarding the resource allocation between recipients of these two benefits. If so then the current structure of the LLTBR will potentially work against this decision.

Another choice for the LLTBR structure would be that it assigns a relative score within each benefit type. As an example, it might identify as high risk the 30% highest risk recipients within each benefit type. This would allow case managers to make support decisions based on, for instance, how risky an Unemployment related benefit recipient was relative to other recipients of the Unemployment benefit, rather than in relation to all recipients. A model structured in this way is more likely to preserve higher level resource allocation decisions between benefit types. It would be a very easy task to convert the current model to one which gave a relative risk score within each benefit type.

A further advantage of a move to a relative risk score for each benefit type is that one could then vary the two year time horizon for each benefit type. For instance, the average duration of payment for IB recipients is very much longer than it is for UB recipients and it seems likely that the choice of a longer time horizon for assessing risk would be appropriate for IB beneficiaries. Although it is difficult to be sure without carrying out some analysis of the data, we believe that the choice of different time horizons for each benefit type could lead to a material improvement in the performance of the LLTBR as a tool for assessing the relative risk of beneficiaries within each payment type.

Finally, we are not certain of the rate of transfer or conversion between the different benefit types and how this has been incorporated. If there is a significant conversion so that, for instance, SB beneficiaries become IB beneficiaries over time (or vice versa) then it will be important that the outcome variable definition is based on benefit payments of any type to each beneficiary, not just the type being paid to that beneficiary at the end of the snapshot month.

A.2.3 Simplicity versus complexity

Note that the discussion that follows is intended merely as a review of the use of the LLTBR in the context of its intended use and suggestions as to how this might be improved. It is not intended as a criticism of statistical risk profiling. In fact we are strongly supportive of the use of statistical risk profiling techniques. Further, the discussion is not intended as offering definitive solutions as to how to target the right assistance to clients, which is a significantly complex problem.

It is worth noting that the current model is, as far as front line staff are concerned, at the complex end of the range of possibilities. We do not mean that it is complex in a technical sense but, as far as the presentation of its results to end users is concerned, it is essentially

a black box. The end user sees a risk score but does not know how or why such a score has been assigned.

If this single score is the primary model output then this creates an opportunity in that one can use a technically complex model with many risk drivers. The current LLTBR is of what we consider to be a moderate complexity so, in a sense, it does not take full advantage of the opportunity created by the use of single model score. Its performance could likely be enhanced and we consider the details of such enhancements in later sections of this appendix.

However, the use of a single model score also has a number of disadvantages. Some of these are acknowledged in the material we have reviewed. The main issues are:

- Risk is not necessarily correlated with an improvement of outcome if support is provided. In fact, the clients with high risk scores face complex barriers and may be resistant to outcome improvement and therefore a poor group to which to offer increased employment support. We think the targeting of support would benefit from a more sophisticated approach than a simple differentiation on the basis of the risk of long duration;
- Case managers are not given any indication of what support is likely to be effective. Two different high risk beneficiaries are likely high risk for different reasons and respond to different support;
- Case managers will not know how and why the model is delivering its results. This can lead to a lack of buy-in.

The other end of the spectrum is a simple, single rule of eligibility. The effect of a very simple rule, which is illustrated by MSD in its documentation, is the use of past duration of benefit as an eligibility rule for support. See for example Figure 1 of 23 February document listed in section A.1 above. This shows a very poor targeting performance for the use of duration when compared to the LLTBR. However, it is worth noting that this is somewhat of an extreme comparison. It is likely that the evaluation of eligibility rules which, in addition to duration, used the type of benefit, age and maybe a very few other characteristics would result in dramatically improved performance. Although these would likely not perform as well as the LLTBR they may not be very much worse. The use of such eligibility rules would improve the clarity around who gets support, why and what support might be appropriate. We suggest that some investigation of the merits of enhanced eligibility rules versus the black box nature of the LLTBR would be worthwhile.

We now focus our review on the LLTBR itself.

A.3 Summary of Model

The LLTBR consist of four logistic regression models fitted using Proc Logistic in SAS. Each separate regression model is for a separate benefit type i.e. one of UB, DPB, SB and IB. The general process in fitting and applying the models is to:

1. Prepare the data;
2. Select the variables to appear in the model;
3. Fit the model;
4. Transform the model to get a risk score;
5. Carry out various diagnostic tests.

We comment on the various stages below.

A.4 Data

A.4.1 Process

The data appears to have been prepared as a series of monthly snapshots with the months running from October 2004 to August 2008. The predictors are based on data known at the end of the monthly snapshot. The outcome variable is a binary response, set to one if the beneficiary was on benefit for the full two years after the end of the monthly snapshot (actually more than 725 days).

Observations are sampled from the full range of monthly snapshots. 60,000 records for each benefit type were sampled.

There are also some other predictors added which relate to participation in various programs over the outcome period. These will be used to calculate a risk score which removes the effect of participating in these programs.

A.4.2 Taylor Fry comments

It is standard practice to reconcile the data against other sources and reports and to fully check the data preparation code. We have not verified that this has been done although the documentation we have reviewed appears to indicate an appreciation of the importance of data quality. It is worth noting that, in our experience, errors in data preparation are responsible for a significant proportion of model failures. This is because model diagnostics, used for checking model performance, are all dependent on correct data.

As a minor comment, it seems to us unnecessary to restrict oneself to 60,000 observations per benefit type. Modern model fitting and variable selection software will cope with many more observations, if they are available. Although we think that 60,000 is certainly enough to fit good models, if a model uses a large number of predictors (as these do) there would likely be a small but material improvement from increasing the number of observations.

Finally, in order to implement some of our later recommendations, we suggest that the data is divided into three parts: a learn sample, a test sample and a hold-out sample. The relative sizes of these samples are normally something like 50%, 25% and 25% respectively. One option is to randomly sample in these proportions across the whole time period. Another, preferable, option is to select the hold out sample from a later time period. We comment on the use of each of these samples later in this appendix.

A.5 Variable selection

A.5.1 Process

A random sample of 20,000 from each sample of 60,000 was used as follows:

1. The categorical variables were transformed to a series of binary variables;
2. Continuous variables had alternative, transformed versions calculated. The transformations used were log and square root;
3. All variables were tested for their association with the outcome variable using Pearson correlations. Variables with p-values of greater than 0.2 were discarded;
4. The remaining variables were modelled using Proc Logistic and variables with a p-value of greater than 0.5 discarded;
5. Step 3 was repeated, discarding variables with a p-value of greater than 0.2;
6. A stepwise selection using a selection threshold of 0.05 was then used. We are unclear if this was a forward or backward process.

The remaining 40,000 from each sample were then used to fit another Proc Logistic model and any non-significant variables removed (we are unsure of the p-value criteria used for this step).

A.5.2 Taylor Fry comments

Whilst the process is a reasonably comprehensive conventional statistical approach to variable selection, it no longer represents best practice for this type of problem. The main issues are that:

- It makes no allowance for interactions between variables e.g. the case where two variables are each poorly correlated with the outcome variable but a combination of them is highly correlated. This can result in too few variables being selected;
- It does not allow for effects from continuous variables which are not linear dependencies or linear dependencies of a transformed version. This can result in too few continuous variables being selected;
- It does not allow for variables which are correlated with each other (as well as being correlated with the target). This can result in too many variables being selected and instability in the resulting model.

Whilst the list of variables which results from the current process will give a reasonable model it will likely be significantly sub-optimal.

Instead, we recommend a process along the following lines, applied separately for each benefit type:

1. Do not transform either the categorical or continuous variables (although categorical variables with more than 10-20 levels may need some subjective grouping);
2. Apply a suitable “black-box” data mining algorithm to the learn sample. This will fit a model and give a list of the variables used in the model, ordered by their importance. The algorithm which we use for this purpose is Treenet which is an implementation of the gradient boosting algorithm MART, sold and distributed by Salford Systems. There is also a freeware version available in the package “R”;

3. Take the list of variables from step 2 above and evaluate the area under the ROC curve on the test sample. Shave (i.e. delete) a few variables from the bottom of the variable importance list. Refit the model. The area under the ROC curve for the test set will increase. Repeat this step until the area starts to decrease. This will give a reduced list of candidate variables;
4. Take the list of continuous variables from step 2 above. Apply a hierarchical clustering procedure (we suggest Proc Varclus in SAS) to segment the list into clusters of correlated variables. Choose one variable from each cluster, based on its ranking in the list from step 2. Some clusters may need two variables;
5. The list from step 4 and the categorical variables from step 3 are the final candidate variables for the regression;
6. Again apply the data mining algorithm and evaluate its performance. This will give a baseline against which to assess the final fitted logistic model.

We are confident that the application of this procedure will result in a marked improvement in the performance of the LLTBR.

A.6 Fitting of the model

A.6.1 Process

There is little description of the adopted process in the documentation. However, from what we can tell, the fitting process seems to be one of simply taking the list of candidate variables from section A.5.1, choosing one of the transformations of the continuous variables and fitting the model with Proc Logistic on a sample of 40,000 observations (a different set than the one used for the variable filtering).

A.6.2 Taylor Fry comments

As with the variable selection, we regard this as a conventional and reasonable treatment but one which now falls short of best practice. In particular:

- It is unnecessary and probably materially inaccurate to use the transformed continuous variables as predictors in the way implemented in the current process;
- Likewise, it is not best practice to deal with categorical variables by simply fitting binary variables for significant levels, although the effects of this on the final model are likely less significant than those of the continuous variable process;
- There is no apparent, comprehensive process for dealing with interactions between predictors;
- We do not use Proc Logistic on a regular basis but our reading of the documentation is that it does not adjust the standard error estimates for variable selection for any under or over-dispersion.

Instead, we suggest a process applied to the learning sample along the following lines:

- Continuous variables should be transformed into categorical variables with each level being a range of values, all with significant exposure;

- All these transformed continuous variables and all categorical variables in the candidate variable list are included and fitted in what is known as a "saturated" model;
- This model should be fitted in such a way that the standard errors are adjusted for any under or over-dispersion. One solution is to use Proc Genmod which calculates and applies the relevant adjustment;
- The saturated model should be gradually simplified by
  - Fitting splines to replace the categorical versions of the continuous variables;
  - Grouping the levels of the categorical variables together, based on pairwise tests of significance between levels;
- Search for interactions between the simplified variables and include significant interactions;
- In the assessment of which splines and interactions to fit, reference should be made both to the p-values of the parameters and to model diagnostics calculated using the test sample.

This suggested process is labour intensive. It requires specialist model fitting software and an experienced analyst. However, it will likely produce a simpler, more robust and more predictive model than the baseline "data-mining model" produced as part of step 6 in section A.5.2, and the current LLTBR.

A.7 Transform the model to generate risk scores

A.7.1 Process

The model score from each of the four models should be the predicted probability of a beneficiary being on benefit for two years. It should be sufficient to simply use the model score as the risk score for each beneficiary. However, for reasons which we have not analysed in detail, it appears that the model score does not agree with the actual probability of a beneficiary being on benefit for two years for various subgroups of the population. To correct this discrepancy, the observations for all the four models have been combined, ranked and grouped on the basis of model score. The average, actual probability of each group being on benefit for two years was calculated. In cases where a group had a lower model score but a higher actual average probability than the next highest ranked group, this group is combined with its nearest neighbour. This process is continued until all such discrepancies disappear.

A.7.2 Taylor Fry comments

We regard this process as reasonable although we have not reviewed it in detail nor seen the detail of the calculations. However, our preliminary view is that the reasons for the discrepancy are connected with what we regard as the less than ideal fitting process described in section A.6.1. If this were to be amended to the process suggested in section A.6.2 then we are confident that the process described in section A.7.1 would no longer be necessary.

A.8 Final model diagnostics and final model fit

A.8.1 Process

We have not seen or reviewed the full set of diagnostics produced for the LLTBR. So in this section we limit ourselves to briefly describing the process we think should have been adopted.

A.8.2 Taylor Fry comments

Extensive diagnostics of model performance should be calculated using the hold out sample. These should include:

- ROC curves;
- Actual versus expected plots against all predictors for various sub-populations;
- Actual versus expected plots against the model score; and
- Various diagnostics showing the incremental loss in model performance from adopting a series of simpler models.

Finally, although not absolutely necessary, we suggest that the entire model is refit using the combined learn, test and hold out samples. Note that this refit is simply a re-estimation of parameters; the model structure is not changed in any way. Neither are the reported diagnostics updated.

APPENDIX B DYNAMIC FISCAL LIABILITY TECHNICAL REVIEW

B.1 Scope

This review is based on the report titled "Dynamic estimation of the fiscal liability of Work and Income clients: Technical report" working paper dated August 2011 (referred to subsequently as the "DFL Report"), prepared by the Centre for Social Research and Evaluation and provided by MSD.

We have restricted our review to the assessment of the suitability of the DFL Report for the estimation of an aggregate liability for welfare benefits. Although we consider it out of scope to undertake a review with respect to its suitability for the determination of an individual benefit recipient's liability we have provided some brief comments in this regard towards the end of this Appendix.

Further, while we have had discussions with MSD staff (particularly Marc de Boer on 22 June 2011) in relation to a previous version of the DFL Report (dated October 2010) we have not had discussions in relation to this current version. We understand the version under consideration here contains some significant differences in assumptions and approach compared to the previous version.

Note that our review is necessarily more restricted than a full technical review. In particular, we have not reviewed:

- The code and processes used to create the modelling datasets and the checks conducted on the data, including reconciliations of these to independent sources;
- The approach used to fit the models;
- The final model specifications, including parameter estimates;
- Model diagnostics for determining goodness of fit, in particular
  - Actual versus expected analyses for various subsets of the data;
  - Actual versus expected analyses across different variables within these sub-populations; and
- Final results and the underlying calculations.

B.2 Purpose of the DFL Report

B.2.1 Individual versus aggregate liability

Our understanding of the primary purpose of the DFL Report is to describe *"the dynamic fiscal liability (DFL) tool."* The tool has been designed by MSD as an initial attempt to build a platform to provide an estimate of the liability for **individual** clients while they are on main benefit. In future iterations, it is anticipated that the model would be updated to align with a formal actuarial valuation of the liability,

The DFL Report also states in regards to the estimates of individual liability calculated by the DFL tool that:

"*While it is simple to add individual liability to arrive at a total figure, this figure does not represent the aggregate liability.*"

In summary, as stated in the DFL Report the aim of the DFL tool is to provide an estimate of the liability for individual benefit recipients but **not** to provide an estimate of the aggregate liability.

This is a very important distinction. We note that while the DFL Report does caution against adding the individual liability figures to arrive at an estimate of the aggregate liability it does in fact do so. Until a formal actuarial valuation is carried out we strongly advise against reporting such results. Any use of such results for policy advice may cause difficulties later.

B.2.2 Definition of liability

The DFL Report estimates the liability for future cash flows for an **individual** client at a point in time. As such there is little need for consideration or discussion related to the appropriate definition of the aggregate liability for the social welfare system. However, the report does consider this briefly on page 15 under a section titled "Liability of clients commencing benefit". This section considers "stock", i.e. clients on benefit at 1 June 2010 and "inflow", i.e. clients commencing benefit during financial year 2010/11. The estimated liability for these clients are added (note that some clients are both stock and inflow and are not double counted) resulting in a total liability for the "population" of clients on benefit through 2010/11.

From an accounting and actuarial perspective the above definition is flawed. While there does remain some significant exploratory analysis and consideration required as to the definition of liability it will certainly be an amount related to a specified group of clients (See Section 4.2.1 of this report) **at a point in time**. It will not include the liability for clients coming into the specified group over a period of time. However, the estimation of the additional liability arising over the course of the next year or period of years will be an important component of the forecasting of future funding. These valuations coupled with the periodic monitoring of experience between valuations will form the basis of the financial control of the social welfare system.

B.3 Summary of analysis

The estimated liability for individuals is calculated using the DFL tool. The tool has been constructed according to the approach represented in Figure B-1.

Figure B-1 **MSD approach: estimation of future liability**

The overall structure outlined above requires models fitted for estimating:

- The duration a client will be on main benefits;
- The proportion of clients in receipt of supplementary assistance; and
- The expected amount of income support for both main benefits and supplementary assistance.

The duration on benefit and amount of income support are combined for each combination of client characteristic. For duration on benefit the relevant client characteristics are:

- LLTBR cohort;
- Gender; and
- Age.

For amount of income support the relevant client characteristics are:

- Benefit type;
- Partner status;
- Age;
- Number of children; and
- Age of youngest child.

The combination of these amounts results in a lookup reference table with future liability estimates for all possible combinations of the client characteristics used in the modelling.

Note that the models listed above include adjustments to take account of various features in the data as follows.

- The model of duration is adjusted to allow for deaths.
- The model of duration is extrapolated to retirement age (65) based on the models fitted to past data.
- The model of income support is extrapolated to retirement age (65) with adjustments for some transition between benefit groups according to age of client and transition between child and non-child rates based on the age of the youngest child;
- Future cash flows are discounted using a net (or real) investment return to determine liability values.

Each of the models and adjustments undertaken are considered briefly in Appendix B.5 below.

B.4 General comments

Before we review the technical aspects of the analysis, it is worth considering certain general aspects related to the approach. In particular we consider whether the type of approach adopted is appropriate, further consideration related to the definition of liability and commentary in relation to the use of an individual or cohort based liability tool.

B.4.1 Approach

The approach used could be considered a "bottom-up-approach". It uses varying estimates of the components of cost (e.g. benefit duration) at an individual level. While this might give useful information for case workers and management at the individual level the estimates are likely to be less accurate in total compared to a standard actuarial valuation approach. A standard actuarial approach would analyse the components of cost at a higher summarised level. Such an approach would more easily cope with underlying trends in the experience of the social welfare benefit system.

This distinction is extremely important. A model of individual client behaviour is unlikely to be able to capture or adequately model trends due to changes in:

- Client behaviour;
- the economy; and
- legislation or policy.

It is important that these sorts of trends are modelled well as the financial control framework suggested in this feasibility study relies on an understanding of the impacts of these trends on the long term cost of social welfare. For this reason the approach adopted in the DFL Report is considered unsatisfactory for the purpose of calculating aggregate liability, i.e. for meeting the requirements suggested in the Level I framework of this report.

However, the DFL tool or some modified version of it might be suitable for estimation of cohort level liabilities under the Level II framework of this report. As noted above it is not within scope of this review to undertake a full technical review nor advise on the modifications that we believe would improve the results of the DFL tool. Nevertheless we have made some brief comments to possible areas of enhancement in Appendix B.5.

B.4.2 Definition of liability

As noted above the DFL tool provides estimates of future benefit cost for individual clients on benefit at a particular point in time. The issue of definition of liability as used in the DFL Report is discussed briefly in Appendix B.2.2. A more general discussion is provided in Section 4.2.1. That section concluded that one reasonable choice for the definition of liability would be based on all clients in receipt of benefits as at the date of valuation ("current recipients" in the notation used in Section 4.2.1 or "stock" as used in the DFL Report). Further, it is recommended that the estimated liability for "future recipients" (equivalent to "inflow" in the DFL Report) would be an important part of the financial control framework.

The DFL tool as currently designed was not intended to, and will not be able to calculate the estimated liability for future recipients. Some modification of either the tool itself or the application of its use would be required to adapt the tool for this purpose. In the DFL Report the liability for inflow clients is estimated after the end of the year to which they relate. In the context of the financial control of the social welfare system this estimate would be required in advance of the year to which it relates, perhaps several years in advance.

A separate forecast of cohorts of future recipients would be required to be able to adapt the DFL tool for the purpose of estimating the liability for cohorts of future recipients.

B.4.3 Individual versus cohort based liability estimate tool

The DFL tool has been described above as a tool for estimating the future liability for an **individual client**. While it has been constructed to meet that objective the underlying models do not distinguish between all clients at the individual level. In fact it would be more appropriate to consider the DFL tool as providing estimates of liability at the cohort level where each cohort is defined by LLTBR, age, gender, benefit type, partner status, number of children and age of youngest child.

The recommended framework in this report requires liability estimates calculated at cohort level (i.e. Level II of the framework, see Section 5.3). Based on this proposed framework a modified version of the DFL tool may be appropriate for use within Level II. Critical to this decision will be the extent to which the cohort segmentation determined to be most appropriate for Level II coincides with the cohort segmentation used in the DFL tool. It is likely that in addition to the various model modifications suggested in this appendix that there will be the need to redevelop the model using a revised segmentation including benefit type as a predictor of duration.

B.5 Technical review

B.5.1 Background

Our technical review is restricted in that we have not reviewed:

- The code and processes used to create the modelling datasets and the checks conducted on the data, including reconciliations of these to independent sources;
- The final model specification, including parameter estimates;
- Full model diagnostics on a “hold out” sample of the data including
  - Actual versus expected analyses for various subsets of the data;
  - Actual versus expected analyses across different variables within these sub-populations; and
- Results for alternative modelling techniques applied to the data.

Also, as noted above the approach used by the DFL tool is not considered appropriate for the estimation of an aggregate liability. Therefore, as we consider it out of scope to undertake a full technical review with respect to the DFL tool’s suitability for the

determination of an individual benefit recipient's liability, the following sub-appendices provide brief comments only.

B.5.2 Model of duration on main benefits

There are 10 models of duration on main benefits: one for each LLTBR group (banded in 10% ranges). Survival analysis is used to fit models to the probability of remaining on benefit during a 4 year data window. The survival curves are Weibull and various explanatory variables are used including factors relating to employment and unemployment rates, age, period since selection and seasonal factors.

Use of survival analysis is an appropriate technique. However, extrapolation of the survival curves based on the fit of only 4 years is likely to produce inaccurate results. Other techniques may be able to use more of the past experience to enable duration modelling over a longer span which would reduce the need for such a long extrapolation period. If survival analysis was retained a technique for doubling the effective period for fitting can be used. The technique involves continuing each survival curve for another 4 years based on the updated characteristics of clients in the cohort and the observed survival curves for the initial 4 year period. For example, for LLTBR group 90, assuming the group had 80% LLTBR=90 and 20% LLTBR=80 at the end of the initial 4 year period, the curve would continue for durations 49 to 96 as a weighted average of 80% of the LLTBR = 90 curve and 20% of the LLTBR = 80 curve. This reduces the amount of extrapolation required.

Another area of concern is the adjustment used for deaths. It appears that the adjustment for past deaths has been made on the basis of modelled deaths during the period. If possible it would be more accurate to remove their impact by removing the actual deaths from the data to be analysed.

The use of LLTBR as a single score obscures the information underlying the LLTBR model, in particular current benefit type. While further exploratory analysis will be required to determine the most appropriate structure for the models of liability for cohorts we believe it will be necessary to develop models separately by benefit type. LLTBR could then be used in 33% or 50% bands within benefit type to keep the number of models manageable.

B.5.3 Model of time in receipt of supplementary assistance

The model of time in receipt of supplementary only assistance is derived as a proportion of the clients on main benefits. The structure of this model seems counterintuitive. Given that the number of clients on supplementary benefits depends on the number of clients not on main benefits it would seem preferable to model the proportion of clients on supplementary benefits as a proportion of clients that are off main benefits. While this might be more intuitive and therefore simpler to understand and use it is unlikely to have any impact on the results.

Our only significant area of concern for this model is the extrapolation of 4 years data to, in some cases, 40 years or so. Given the uncertainty involved it would be preferable to develop a technique that required less extrapolation.

B.5.4 Model of income support entitlement

The model of income support entitlement is based on average payments observed during the 4 year observation period (all amounts adjusted to 2011 dollars based on NZ CPI). The average amounts are summarised according to:

- Benefit type;
- Number of children;
- Age group; and
- Partner status.

Projections of benefit payment amounts are derived using the starting averages described above with transitions to different benefit amounts based on:

- Age (as clients age they transition from one benefit related age group to the next);
- Cessation of dependent children based on age of youngest child;
- Observed transitions to other benefit types over the 4 year data period;
- Extrapolation of the change in benefit type to retirement age.

It is difficult to comment on the exact approach given our limited understanding of the exact manner of model fitting and projection. However, we can comment that there are some simplifying assumptions that have been taken that may have a material impact on the results:

- No transitions are assumed after the first 4 years;
- Benefit rates based on the number of children assume clients have no more children; and
- Change in benefit rates appears to occur when the youngest child reaches age 18 rather than changes in benefit as each child reaches 18.

B.5.5 Future liability reference table

As noted we have not checked the calculations underlying the results in the report and specifically not those used to determine the reference table. However, we have reviewed the formula stated for discounting of future benefit payments. In relation to the discounting we have comments on two matters.

Discount rate assumption

The use of the Treasury's 2010 assumed 3.5% risk-free real discount rate seems high based on current New Zealand consumer price inflation (current expectations of say 4% p.a.) and market yields (3% to 4½% p.a.). In the insurance context in Australia, companies are required to use current expectations of inflation and market based risk-free yields in determining appropriate rates for discounting. Under such a regime the appropriate discount rate would currently be about 0%. However, the use of market based discount and inflation rates results in fluctuations in the estimated liability from balance date to balance date despite all other factors remaining constant. This may not be desirable for the financial management of the New Zealand welfare system and a more static view using long term inflation and discount rates may be appropriate.

Thus, in the current context, the use of the Treasury assumption of 3½% p.a. risk-free real discount rate is probably reasonable. This assumption is presumably based on target long term consumer price inflation of the order of 3% and Government bond yields of 6% to 7% p.a.

Discount rate formula

The formula stated for discounting future benefits may have a small error. It would normally be assumed that benefit payments occur on average half way through a month and therefore the discount to present time would require a discount factor of:

$1 / (1 + DR)^{(i + ½)/12}$ instead of $1 / (1 + DR)^{i/12}$.

B.6 Reporting standards

We recognise that the DFL report provided for our review was in draft, and that the tool is in the early stages of its development. Therefore our comments relate only to the general format and level of reporting. Furthermore the report is not intended as a widely distributed valuation report but rather as an internal document describing the development of the DFL tool.

We provide the following comments in the knowledge that the DFL report is not meant to achieve the reporting standards required of actuaries. However, the comments may be useful to inform MSD in regards to the level of reporting that would likely be provided for an actuarial valuation of the social welfare system liability.

As it stands the DFL report falls well short of actuarial professional standards. The general requirement for an actuarial report is that other qualified professionals with access to the same data would be able to reproduce the analysis in its entirety. This does not seem possible with the DFL Report. For example, it is not explicitly stated how the adjustment for deaths is carried out on data prior to fitting nor is the formula for the combination of benefit duration and amount of income support payment provided. There are many such areas which appear to be left open to interpretation.

A standard actuarial report would follow a logical sequence describing in detail:

- Background and scope
- Data
- Methodology
- Comparisons of previous model and experience
- Assumptions
- Summary of results
- Comparisons of current results to previous results
- Commentary

In addition to the report sections outlined above further details would be provided in Appendices covering:

- Summaries of data

- Details of the fitting of each model including data, fitting procedures, results (e.g. parameter estimates) and diagnostics
- Detailed results including comparisons to previous models

There are also specific standards which apply to certain actuarial valuations which define required content for reports.