Social Investment Agency
Toi Hau Tāngata

Standardised Cohort Analysis

Methodology Overview

New Zealand Government
Te Kāwanatanga o Aotearoa

Author

Simon Anastasiadis

Acknowledgements

Andrew Webber, Charlotte Rose, Craig Wright, Tahia Eaqub

Data Lab and Integrated Data Production teams at Stats NZ

Eric Krassoi Peach, Eyal Apatov, Lisa Meehan, and Marc de Boer for their review and feedback

Citation

Social Investment Agency 2026. Standardised Cohort Analysis – Methodology Overview. Wellington, New Zealand.

ISBN 978-1-997329-03-9 (online)

Published in June 2026 by

Social Investment Agency
Wellington, New Zealand

Contents

Repeatable analyses support social investment

Social investment involves making effective use of limited resources. In the context of social services, this includes ensuring that organisations seeking to help people understand what works and for whom. Organisations with this knowledge, including government and service providers, are better able to offer effective support and to reach the people who need them most.

The Social Investment Agency (SIA) has designed a repeatable analysis that uses data the government already collects to contribute to this understanding. The analysis is designed to provide descriptive and comparative insights into the experiences of different groups of people to support operational decision-making and service design. As a repeatable analysis it provides a standardised process that can be applied to different groups with minimal marginal effort, enabling the delivery of more insights for the same effort.

Data for this analysis is drawn from the Integrated Data Infrastructure (IDI): a large research database containing de-identified microdata about people and households developed and maintained by Stats NZ. By using existing data, we reduce the need for additional data collection and increase the value derived from established collections.

Our repeatable analysis is designed to analyse cohorts of individuals. These cohorts might be defined by a range of personal characteristics or experiences – for example, children who have had at least one parent incarcerated – or by a list of identities – for example, from linking a social service provider's client list into the IDI. As many insights arise from comparisons between groups, the repeatable analysis includes a matching methodology designed to construct a group from the general population against whom it is meaningful to compare the cohort.

This paper provides an overview of our repeatable analysis and the insights that it provides. We describe our analytical process including the key cohorts, matching methodology, and the sequence for automating the analysis.

Provides cross-section & time-series insights

The repeatable analysis is designed to provide insights at a point in time (cross-section) and insights into changes over time (time-series). By default, the point in time is set as the most recent date for which there is data, allowing it to approximate current state.

When presenting results from the repeatable analysis we tend to divide the insights into three parts. These parts would be presented in separate sections of a report or on separate pages of a dashboard. The three parts are:

1. **Demographic overview** – This includes overall counts of people in the input cohorts (see the next section for more details), demographic descriptors, and interactions between the cohorts where applicable.
2. **Current experiences** – This includes the prevalence of a range of experiences at a point in time (cross-section). Depending on the measure these may be recent experiences, experiences at

any point in a person's lifetime, or the experiences of family members. As the majority of the data in the IDI is government administrative data, most of the experiences we can observe are interactions with government or service receipt from government.

3. **Patterns over time** – This includes how the average experience has changed over several time periods around a date (time-series). Deliberate choice of this date allows for different before-after comparisons. For example: a policy implementation date allows us to consider how a policy change has impacted people's lives, while a programme enrolment date might let us examine the effect of receiving support from a community organisation.

For organisations seeking to make a difference in the lives of New Zealanders, these three parts might be described as "Who am I reaching?", "What were their experiences?", and "What were their outcomes?" By making careful choices about comparisons, these insights can both show how participants differ from the wider eligible population and provide an indication of the effectiveness of a service. Examples of these insights are shown below in Figure 3 to Figure 5.

A key challenge to generating the results needed for these three sets of insights is that the available data in the IDI is not real time. For some data sources, there can be significant delays between the occurrence of an event and the event being included in IDI data (an overview of these delays as at mid-2025 is provided in Table 2 in the appendix). This challenge is handled by our repeatable analysis in two ways:

- We use the most recent data available. Where the goal is to understand people's current experiences, but current data is unavailable, then the next best alternative is the latest data that is available. While people's experiences do change over time, in general it is preferable to have information that is a bit out-of-date than to have no information. Furthermore, because we only present average results for groups, and groups change more slowly than individuals, data delays have less of an impact on the insights we produce than they would for other applications.
- We annualise values where data for an entire twelve-month period is not available. For example: suppose we only have six months of data for a group, if members of this group had a total of 10 hospitalisations over these six months, then we would report the annualised number of hospitalisations as 20. In practice, date ranges and data availability may differ between individuals so the calculation is done at a daily resolution to allow for this nuance.

  Annualising values in this way requires us to assume that the patterns of experiences we can observe in available data continue beyond the available data. While this is a less robust assumption for individuals, it is more robust for groups of people, and we only present insights from grouped results. It is also a less robust assumption when there are seasonal patterns in the data, but only some measures exhibit significant seasonal trends.

The aim of our repeatable analysis is to provide useful insights, even where those insights are imperfect. Use of the IDI enables our repeatable analysis to summarise a wide breadth of experiences and to do so in a consistent and efficient manner. Although there are limitations, in practice we find they do not diminish the overall value the repeatable analysis provides.

Six cohorts are used in analysis

Central to our repeatable analysis is the concept of a cohort: a group of people who are examined together. A total of six cohorts are used in our repeatable analysis. An overview of each cohort is given in Table 1, with more detailed descriptions in the following sections.

The original motivation of our repeatable analysis was to provide insights to support organisations and services seeking to make a difference in the lives of New Zealanders. This application is the focus of this paper. However, we have also found the repeatable analysis to be effective for applications that are not oriented around an organisation or service. For such applications we would use a primary and secondary cohort as defined by our customer, rather than a participant and potential participant cohort as defined by a service or organisation.

Table 1: Overview of six cohorts with examples

Cohort	Description	Fictional example
Participant	Current and past participants of the organisation	People who TuitionNZ (a fictional NGO based in Christchurch providing tutoring support) have supported
Potential participant	People that an organisation has stated it aims to work with	People living in Christchurch, at the time TuitionNZ is operating, aged 17-18, not enrolled in study, and without qualifications
Possible comparison	People whose current experiences are comparable to the participant cohort such that they are candidates for matching	People living in the South Island, at the time TuitionNZ is operating, aged 17-18, not enrolled in study, and without qualifications
Possible forecast	People whose past experiences are comparable to the current experiences of the potential participants cohort such that they are candidates for matching	People living in the South Island, age 17-18, 5 years ago, not enrolled in study, and without qualifications
Matched comparison	The subset of the possible comparison cohort who are most similar to the participant cohort	People living in the South Island, at the time TuitionNZ is operating, aged 17-18, not enrolled in study, and without qualifications who are most similar to people TuitionNZ have supported
Matched forecast	The subset of the possible forecast cohort who are most similar to the potential participant cohort	People living in the South Island, at the time TuitionNZ is operating, age 17-18, 5 years ago, not enrolled in study, and without qualifications who are most similar to people in Christchurch with the same circumstances this year

The first four cohorts in Table 1 are provided by the user as inputs to our repeatable analysis. These must be provided each time the analysis is run. Rather than omitting an input if a cohort is not required for an application, a user instead provides a cohort that contains no members. This is useful if an organisation has no participants yet, or we do not need results for a matched population.

1. Participant cohort

The participant cohort is current and past participants of the organisation. Depending on the organisation, these people might also be called customers, clients, or service users. This cohort is provided as an input to the repeatable analysis and is used to produce both cross-section and time-series results.

One way we define this cohort is from organisations sharing participant records with Stats NZ for linking into the IDI. These linked identities then become the cohort. Another way we might define this cohort is to use existing data in the IDI. For example, if an organisation provided support to all families with truant children who live in Southland, then because the IDI already contains information about geography and school truancy, it would be possible to construct this cohort from existing data.

Several attributes are required when defining the participant cohort. In addition to an IDI-specific ID number, each member of the participant cohort must also have a reference date and an indicator for whether they are a current or past participant.

The reference date is most often the date at which the person became a participant. For example: a referral date, appointment date, or first service interaction date. This value tends to be specific to each individual. The reference date is essential for producing insights into patterns over time as it allows us to study the experiences of participants before they received the service and to observe how those experiences change during and after participating.

2. Potential participant cohort

The potential participant cohort reflects the people that an organisation has stated it aims to work with. Depending on the organisation, these people might also be referred to as the target population, service-eligible group, or people at risk. This cohort is provided as an input to the repeatable analysis and is used to produce both cross-section and time-series results.

The most common way we define this cohort is from an organisation describing who these people should be. Our IDI researchers then construct the best match to this description that can be done using characteristics observed in the IDI. Because this is done using existing data, this cohort is often the starting point when seeking to provide insights for organisations or services that are planned but not yet operating.

Like the participant cohort, the potential participant cohort also requires a reference date attribute. This reference date tends to be one of three values: the date the organisation or service operation began, the date the person became eligible for support, or the date of the latest available data. The choice between these three options depends on whether it is most meaningful to compare the participant population against potential participants when the service began, when people became eligible, or against people who could become participants today.

To avoid double counting of people as both participants and potential participants, our methodology removes people in the participant cohort from the potential participant cohort. If we assume that all participants are drawn from the potential participant cohort, then the two groups can be combined after the analysis to produce insights for the full population of interest.

3. Possible comparison cohort

The possible comparison cohort includes people who are comparable to the participant cohort such that they are candidates for matching to the participant cohort. These people are considered in similar years to those covered by the reference dates of the participant population. This cohort is provided to the repeatable analysis as an input but is only used during calculations, not to produce results.

This cohort is most often defined in a similar way to the potential participant cohort. However, the definition may need to be broader in order to ensure sufficient overlap in matching characteristics. For example, consider an organisation working with at-risk youth in a specific town. If most at-risk youth in the town are already participants of the organisation, then this cohort will need to include neighbouring towns or in the wider region so there are enough at-risk youth for matching. As a pragmatic guideline, we often aim for this cohort to be 20-100 times larger than the participant cohort. This ratio is not a strict requirement – what matters most is adequate overlap in key characteristics to ensure robust matching. Because the IDI covers the entire New Zealand population, there are almost always enough people from which to construct an appropriate comparison cohort.

Instead of a reference date attribute, this cohort is created with a reference year attribute. In contrast to the above cohorts, individuals in this cohort may be included several times – each time with a different reference year. Our methodology uses the reference year to construct reference dates as needed (often as the end date of each quarter). As the reference years are chosen to cover the entire date range of the participant cohort, this allows for matching to occur on both the best person and the best timing (using end of quarter dates as reference dates means we can ensure that the reference date for the participant cohort and the matched comparison cohort are within three months of each other).

To ensure a valid matching process, our methodology removes people in the participant cohort from the possible comparison cohort. This ensures that comparisons are made against similar, but different, people.

4. Possible forecast cohort

The possible forecast cohort includes people who are comparable to the potential participants cohort such that they are candidates for matching to the potential participants cohort. These people are considered several years before the reference date used for the potential participant population. This cohort is provided to the repeatable analysis as an input but is only used during calculations, not to produce results.

This cohort is defined using a similar process as the possible comparison cohort. Except for the difference in timing, it has the same attributes, and it handles intersection with the participant and potential participant cohorts in the same way.

The key difference between this cohort and the possible comparison cohort is that where the possible comparison cohort covers the same date range as the participant cohort, the possible forecast cohort is drawn from a past time period. This allows the matched forecast cohort resulting from the matching methodology to act as a form of forecast.

5. Matched comparison cohort

The matched comparison cohort is the subset of the possible comparison cohort identified by the matching methodology as being most similar to the participant cohort. This cohort is generated during the repeatable analysis and is used to produce time-series results.

This cohort can only be generated when both the participant cohort and the possible comparison cohort are available. This is because the matching methodology needs both a cohort to match to and a cohort to select matches from.

The intention is for this cohort to approximate what the participant cohort would look like if its members had not become participants. This means we can use the difference between the experiences of the participant and matched comparison cohorts to infer what effect the organisation has had on its participants.

6. Matched forecast cohort

The matched forecast cohort is the subset of the possible forecast cohort identified by the matching methodology as being most similar to the potential participant cohort. This cohort is generated during the repeatable analysis and is used to produce time-series results.

Like the matched comparison cohort, this cohort can only be generated when both the potential participant and possible forecast cohorts are available.

The intention is for this cohort to approximate what the potential participant cohort will look like in the future. For example, suppose we want to forecast the experiences of people leaving college without qualifications this year. We can estimate this by finding very similar people who left college without qualifications five years ago and observing their experiences since.

Standard matching methods are used

As part of our repeatable analysis, we use a matching methodology to select the two matched cohorts from the two possible cohorts. Matching methodologies are a well-established statistical technique used to create comparable groups. They are also transparent and relatively easy to communicate to stakeholders, especially when the aim is to compare trajectories rather than credibly claim causal effects. We use matching to provide an estimate of the effect of a service by considering the difference between the people who received it and the matched group of similar people who did not.

Standardised, not “the standard”

Matching is only one option for assessing the effect of a service or programme. In designing our repeatable analysis, we have prioritised having a standard process over fine-tuning the matching methodology for each application or maximising the robustness of each set of results. Our process is standardised – it is not intended to define “the standard”. While we hope our repeatable process can accelerate more robust analyses, we recognise that in some cases researchers will consider it insufficient. Our claim is not that a repeatable analysis provides robust evidence in every application, but that it can provide at least some evidence in many applications including ones that might otherwise go unexamined.

Standardising and automating much of the process means we can deliver results for more cohorts, organisations, outcomes, and time periods within the same resource limits. Something that would be difficult to achieve using more bespoke evaluation methods. There will also be cases where matching may be the most practical method available and where refinements to the methodology are unlikely to materially change the results.

However, standardisation also means that there are limits to the standard of evidence provided by the results. Our approach allows for descriptive comparisons not causal inferences. Use of this kind of approach for causal inference (a true quasi-experimental propensity score matching approach) would require strong assumptions about the real world that are unlikely to hold: unobserved confounding, measurement error, and endogenous selection into treatment. The main threat to a causal inference is that there could always be features in the lives of participants that the IDI researcher does not know but that the participant themselves or practitioners working with them do know which are relevant both to their participation and outcomes. Hence, results should be interpreted as powerful structured comparisons between similar cohorts that are best triangulated with other sources of evidence and insights from practice.

The intention of our repeatable analysis is to enable the responsive creation of insights not to define the standard for matching or comparison. When assessing the impact of a programme (for example, by organisations who partner with the Social Investment Fund), we expect a formal evaluation process will be undertaken. Such an evaluation should be designed to use the most effective approach for assessing each programme. In this context, our repeatable analysis might be considered more timely evidence that supplements targeted impact evaluation.

Matching process is used twice

Our repeatable analysis uses matching twice: first to construct the comparison cohort for actual participants, and second to construct the forecast cohort for potential participants. Propensity scores are estimated using a logit model, with exact matching on age and quarter for the comparison cohort and exact matching on age for the forecasting cohort. The key steps within our matching methodology can be summarised as follows:

1. The treated cohort we want to match to (for example, participant cohort) and the control cohort we want to match from (for example, possible comparison cohort) are gathered into a single dataset.
2. A range of measures are calculated for both cohorts – including standard demographics and a range of experiences in the two years prior to the reference date. These measures are used both during the scoring of the two cohorts and for assessing the quality of the match (steps 3 and 6 below). A complete list of measures can be found in Table 3 in the appendix.
3. A logit model is fit to predict the likelihood that an individual originated from the treated cohort. This likelihood provides a propensity score for how similar the members of the control cohort are to the members of the treated cohort.

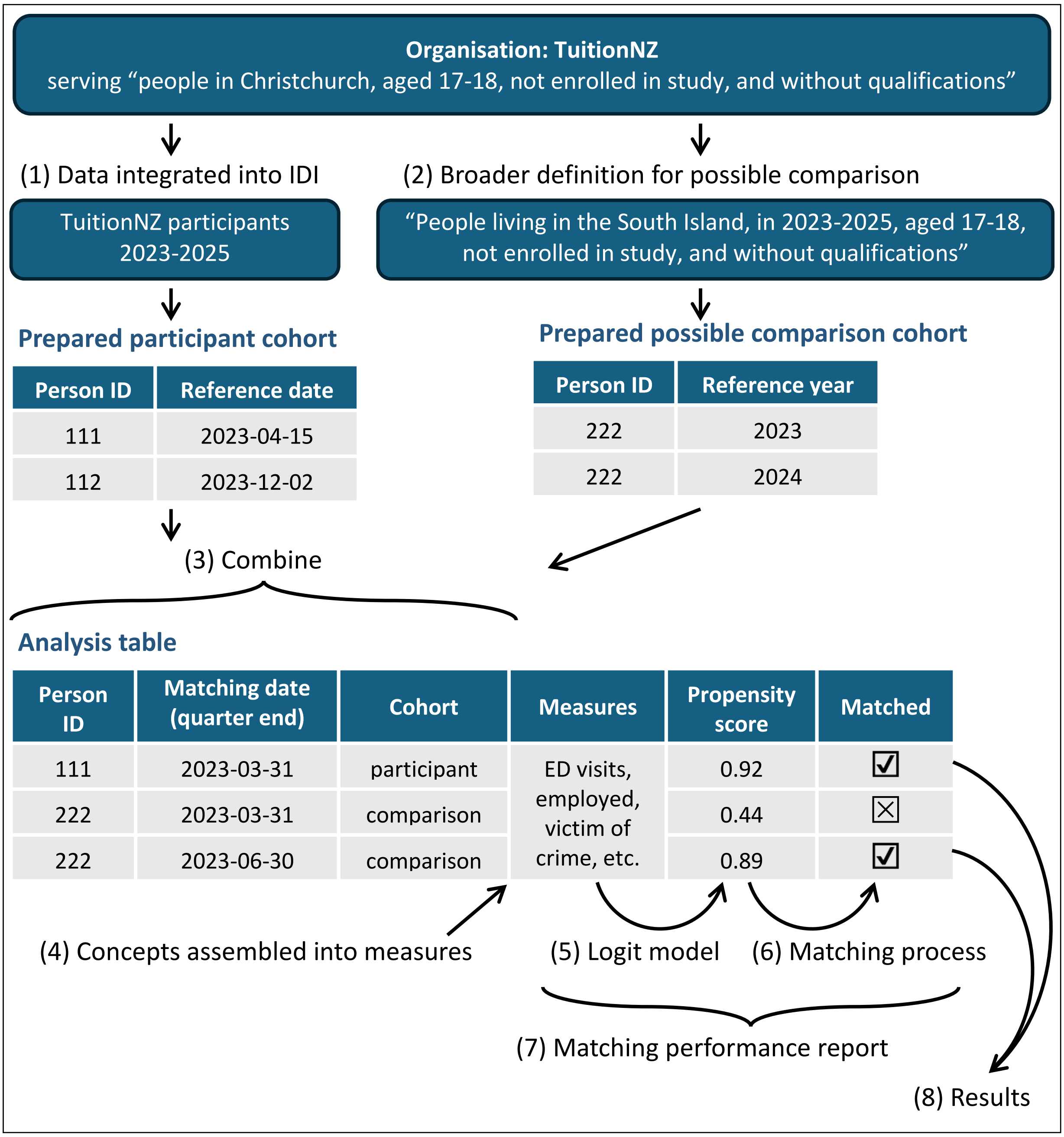
4. The matched cohort is created by selecting records from the control cohort with scores closest to the treated cohort – subject to an additional requirement for exact matches by age (and quarter for the comparison cohort).
5. The quality of the match is assessed and reported on. SIA staff can review this report to confirm there have been no significant variation from regular performance of the matching methodology.
6. Results are then generated for both the matched cohort and the treated cohort.

The next section provides an illustration of this process in Figure 1 using the fictional organisation TuitionNZ from Table 1. Technical readers interested in the detailed design choices that shaped our matching process can find these in the appendix.

Illustration of matching

Figure 1 provides an example of the matching process for TuitionNZ, a fictional NGO based in Christchurch. The numbered steps set out the order in which tasks are undertaken. The tables demonstrate of kinds of values that appear in the data structures (all data values are fictional).

Figure 1: Illustration of matching process with fictional data



Use of matching for forecasting

Recall that by design the current experiences of the potential participant cohort are similar to the past experiences of the matched forecast cohort. This lets us use the subsequent experiences of the matched forecast cohort as an estimate for the future experiences of the potential participant cohort. This is a form of forecasting.

Every form of forecasting has strengths and limitations. Compared to microsimulation or time series approaches, our use of matching has two significant strengths: First, because our forecasts are based on observed experiences for actual individuals, rather than modelled experiences, they avoid any need to construct causal links (specifying how earlier experiences lead to later ones). Second, because our forecasts use only a relevant subset of historic records, our method better accounts for changes in group composition over time (for example, if a group used to be 20% male but is now 40% male, then matching to potential participant cohort will account for this).

There are also two significant limitations of using matching for forecasting. First, our method is less able to account for trends over time. It relies on historical examples instead of modelling changes and therefore assumes that outcome patterns remain stable across cohorts and time. In contrast, time series approaches can explicitly project trends. Second, our method requires that comparable people to current potential participants existed in the past. In contrast, modelled methods like microsimulations can create synthetic comparison cohorts with any combination of characteristics even if people with those characteristics never existed. However, such models require assumptions about people's behaviours and social dynamics, especially when extrapolating beyond observed data.

Running the analysis is automated

Repeating a process can ensure consistent outputs, but by itself does not guarantee efficiency. In developing our repeatable analysis, we have isolated the components that require user input allowing us to automate almost all the execution. This means that our repeatable analysis is both consistent and performant.

Execution of our repeatable analysis can be divided into three phases as discussed in the next three sections. Code for our repeatable analysis can be found on the SIA GitHub page (https://github.com/nz-social-investment-agency). Researchers seeking to run the analysis themselves are also advised to familiarise themselves with the Accelerated Data Analysis & Pipeline Toolkit (ADAPT) as these tools are used throughout the analysis. User guidance for ADAPT is available on the SIA website.

1. Refresh concept setup

Data in the IDI is updated three times a year. Each of these updates in called a refresh and can involve significant changes to the data. Rather than depending on the data in its raw form, our repeatable analysis depends on a collection of concept definitions.

Each refresh, these concept definitions need to be updated and rerun so that the latest data can be used. Running all the concept definitions has been automated with a definitions-specific pipeline. This pipeline attempts every definition, reporting for each definition whether it was completed or errored.

For staff, much of the refresh setup is done by updating references from the previous refresh to the new refresh. In almost all cases, this is trivial to do in the global settings file. Once references have been updated, the definitions-specific pipeline is executed. Out of seventy-five concept definitions, based on previous experience, we would expect at most five to error during this pipeline. Such errors tend to be caused by changes to the source data that require staff investigation to adjust for.

Once the refresh setup is complete, it does not need to be run again until the source data changes.

2. Specifying cohort input

User input is required to specify each set of results that will be generated. We have described above how our repeatable analysis uses six cohorts, of which four are provided as inputs by the user. These are the primary inputs provided by a user and are the best way we have found both to minimise the effort required and to give users suitable control over the process.

To ensure these inputs are in the format required by the main pipeline, an initialise script is provided. This creates all the necessary templates and folders, including files for the four cohorts for the user's input. After initialising a cohort, it is expected that a user will edit these four files. A test cohort exists to guide demonstrate this process.

When the pipeline is run to produce results, all the files specific to a set of results – including the outputs – are saved into its folder alongside its input scripts. Keeping the inputs and outputs together in a single location is a deliberate design feature. This simplifies the user experience and ensures that all results depend upon the same repeatable analysis (rather than localised copies with small variations).

3. Pipeline execution

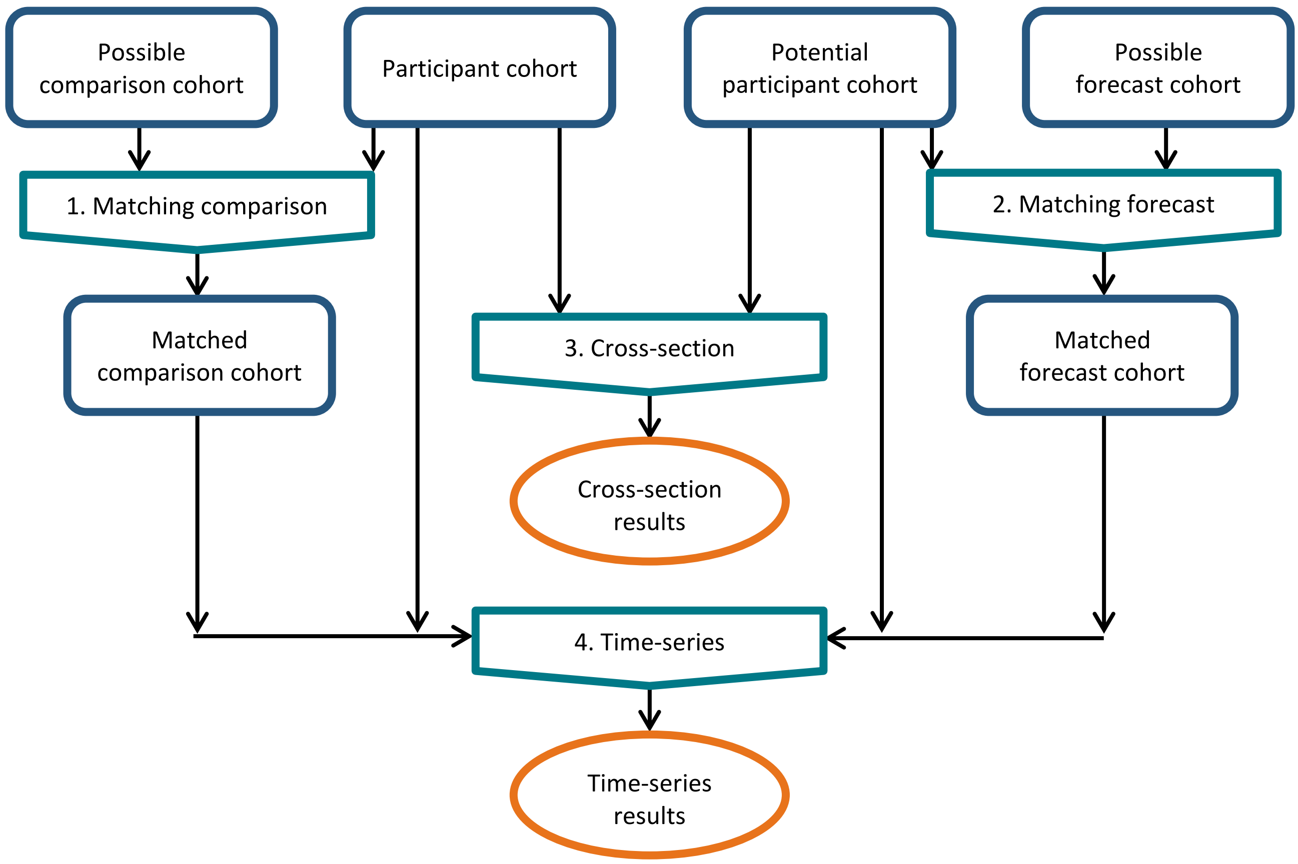
Once a refresh is setup and the input cohorts are specified, the automated pipeline responsible for our repeatable analysis is ready to be executed. The pipeline can execute any of the results for which there is user input. It can produce multiple sets of results at once, in series (recommended) or in parallel. A set of results takes 3-4 hours on average, allowing us to produce 3-5 complete analyses overnight. Figure 2 provides an overview of how our repeatable analysis executes.

Execution of the pipeline is governed by a series of control files and a series of code files. The control files are a structured way to provide instructions to our code. Such instructions include what code to run and in what order, how concept definitions should be combined into datasets, the summaries that should be produced from each dataset, and the confidentiality protections that should be applied to every set of results.

Robust error tracking and recovery have been built into the pipeline. Error tracking makes it easy to determine when and why an error has occurred and hence to fix it. This is done by regular reporting of progress to the user during execution, and by recording the success or failure of each step within the corresponding control file.

Error recovery means that if an error occurs, the pipeline will continue to run where appropriate. So, where several tasks are independent, an error in one task will not stop other tasks from completing. It also means that after an error is fixed execution can be resumed from the last successfully completed step. This reduces wasted effort repeating tasks that are already complete.

Figure 2: Overview of repeatable analysis execution



The way our repeatable analysis produces consistent results from the IDI with minimal staff effort has created new options for how SIA can generate and share insights. We expect that organisations seeking to improve the lives of New Zealanders will use these insights to better understand what works and for whom, leading to more effective support that reaches the people who need it most.

Insights support operational improvement

The repeatable design of our analysis is intended to lower the effort to generate results. This means the analysis can be run several times during a service or programme, producing insights that can be used to refine operational activities at low cost.

Alongside the design of the repeatable analysis, SIA has developed a dashboard template for presenting its results. Figure 3 to Figure 5 show examples from this dashboard template with fabricated data to illustrate the kind of insights the analysis produces and how they can be applied.

Demographic overview

The demographic results from our analysis include overall counts of people in the participant and potential participant cohorts. For organisations seeking to make a difference in the lives of New Zealanders, these can give insights into coverage and reach of their services or programmes.

Figure 3: Example results: demographic overview

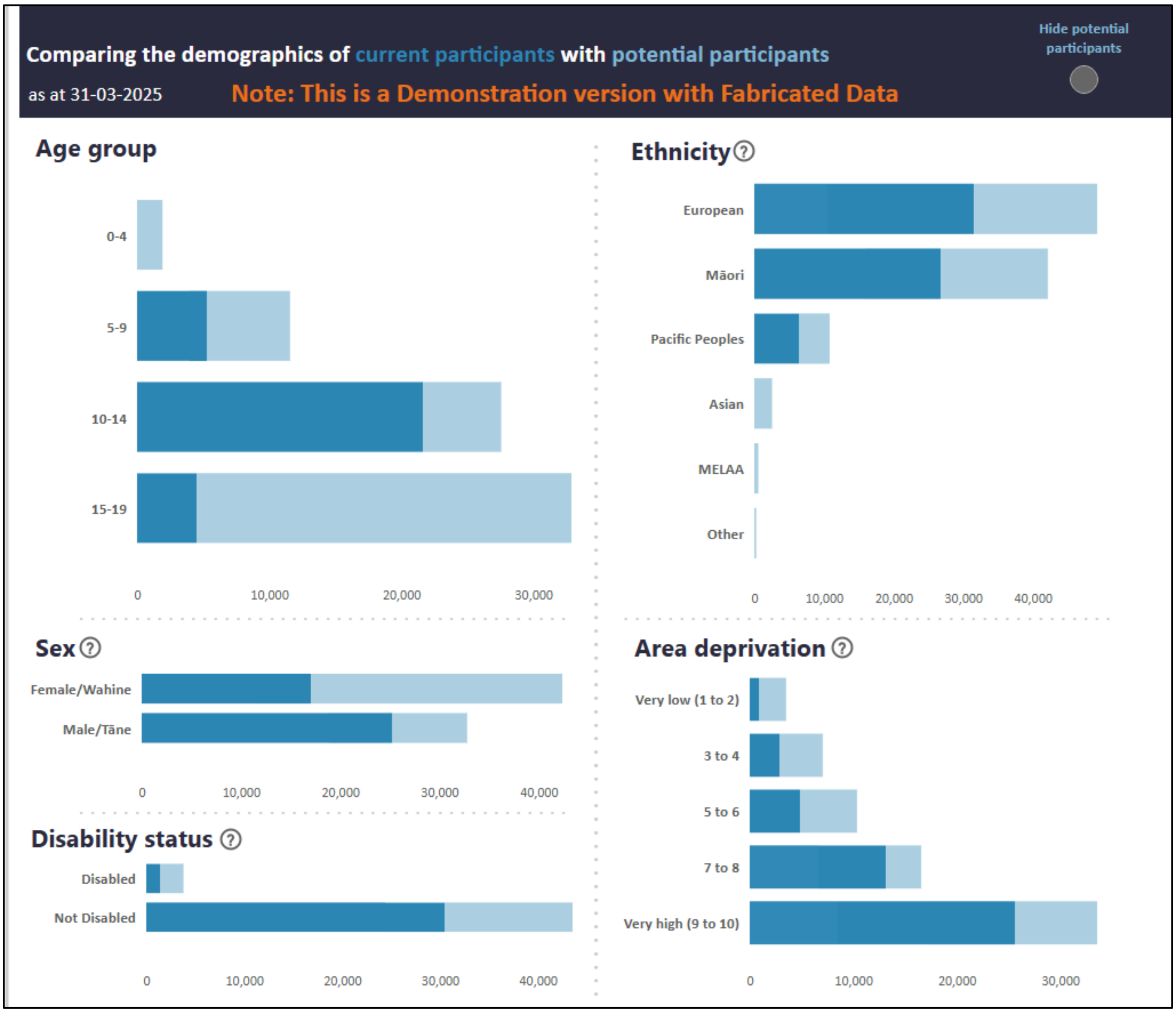


Figure 3 gives an example presentation of these demographic results. The coloured bars help make clear the relative sizes of the participant and potential participant cohorts. This presentation helps users answer questions like:

- Are we reaching the people we set out to reach?
- Where might there be subgroups of people who we are not reaching as intended?
- How could we adapt our approach to reach more potential participants?

Answering these questions involves a comparison of the results for participants and potential participants with the organisation's intentions and resources.

Current experiences

The current experience results from our analysis include the prevalence of a range of experiences for people in the participant and potential participant cohorts. For organisations seeking to make a difference in the lives of New Zealanders, these can give insights into the broader context of people's lives, enabling more effective responses to interconnected experiences.

Figure 4: Example results: current experiences

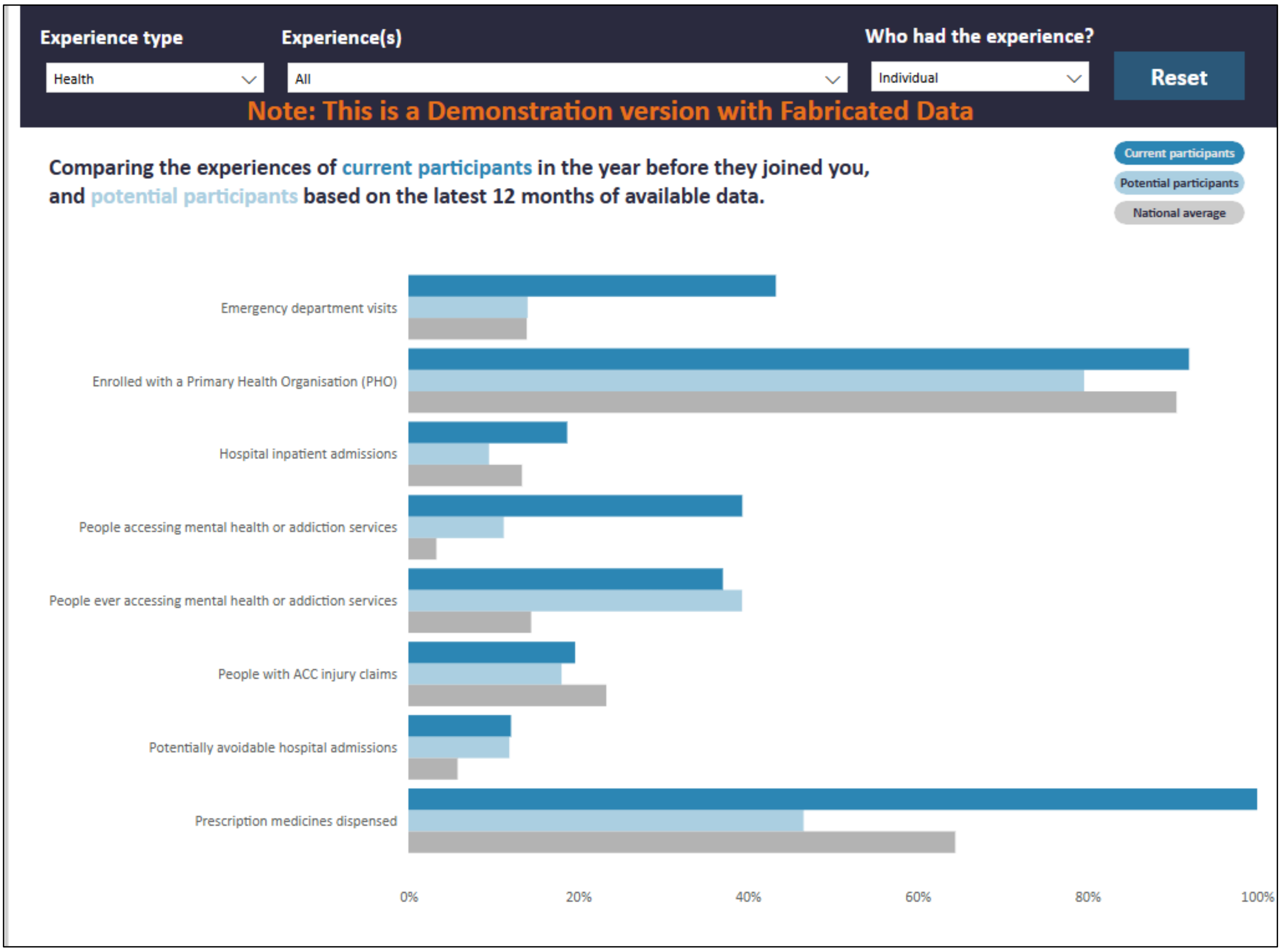


Figure 4 gives an example presentation of these current experience results. Note the selection filters along the top, mean it only shows a subset of all the experiences considered in our analysis. The coloured bars show the prevalent of different types of experiences for the participant cohort, potential participant cohort, and the national average. This presentation helps users answer questions like:

- Are we reaching the people whose needs we are best placed to respond to?
- Where might our referral pathways be refined to better reach people?
- What additional details would it help to collect from participants during onboarding to understand their circumstances?
- How might other experiences be interconnected with the experiences of most concern to our participants or community?

Answering these questions involves a comparison of the results for participants and potential participants with the organisation’s practice knowledge and frontline insight.

Patterns over time

The patterns over time results from our analysis include the average experience for people in each cohort. For organisations seeking to make a difference in the lives of New Zealanders, these can give insights into whether their work is having the intended effect.

Figure 5: Example results: patterns over time

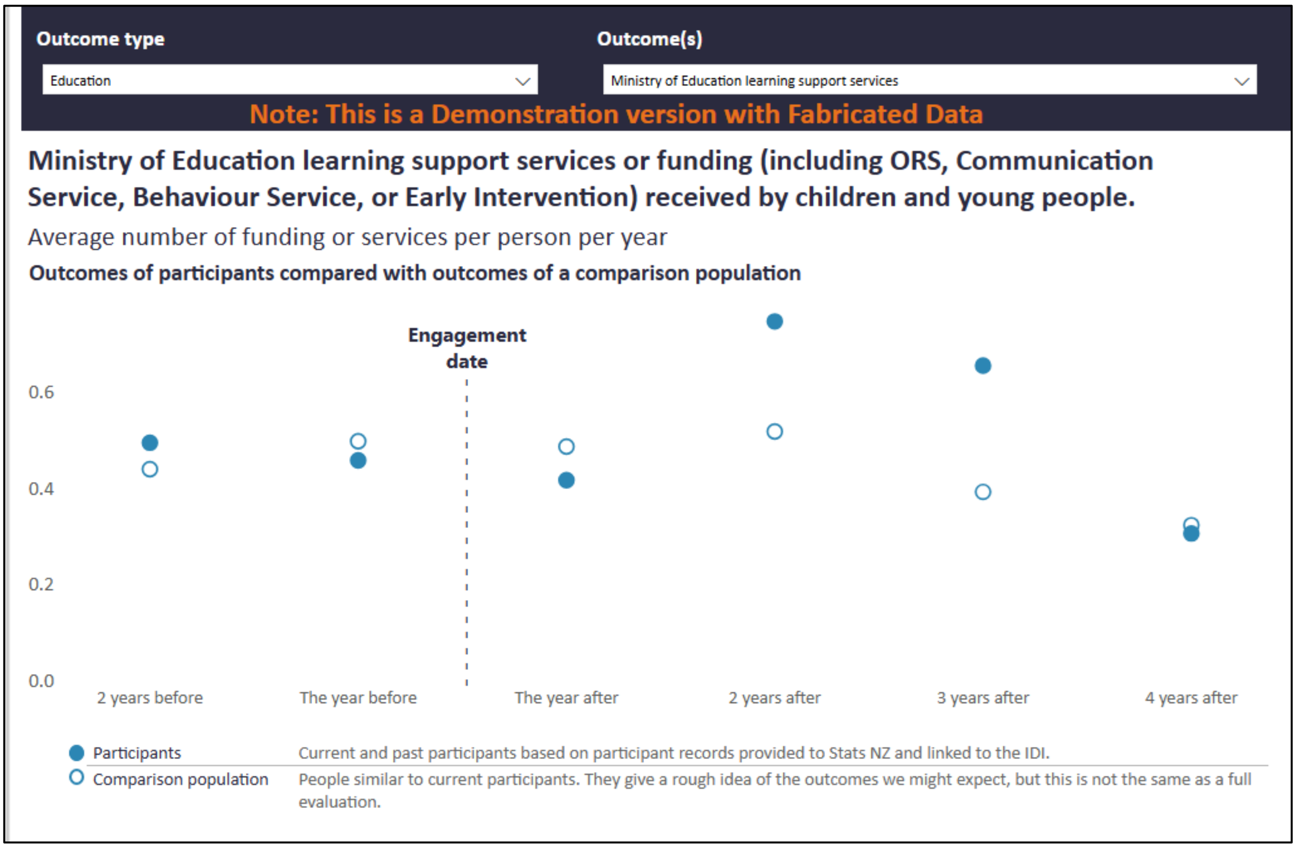


Figure 5 gives an example presentation of these time series results. Note the selection filters along the top, mean that only one experience is shown at a time. The coloured and hollow circles show the average experience for the participant and comparison cohorts. A similar chart is produced showing the potential participants and the forecast cohort. This presentation helps users answer questions like:

- Are we seeing patterns consistent with what we would expect if our work was effective?
- Where is our work more effective and others might learn from what we do? Or less effective and we might improve by learning from what others are doing?
- How well are the desired improvements in participants' lives being sustained?

Answering these questions involves a comparison of the results for the participant and comparison cohorts with the organisation's expectations and goals.

Note that these results depend on the cohorts constructed by our matching process. Because this process is not designed to meet the standards for formal causal evaluation, the findings should be treated as associative rather than causal. In addition, while the presentation allows for before-and-after comparisons, we are not using a difference-in-differences design nor testing for parallel trends. Hence these results are best interpreted as early indications of progress, pending more rigorous evidence from a formal evaluation.

Appendix 1: Data timeliness

The delay range in Table 2 shows the minimum and maximum difference between an event being recorded and that event appearing in the IDI, measured in months. The minimum delay occurs when events happen just before a data update. The maximum delay occurs when events happen just after a data update and must wait for the next one.

There is ongoing work by Stats NZ and organisations that supply data into the IDI to improve the timeliness of data provided to the IDI. The update frequency and delay range shown in Table 2 are as at mid-2025. These improved during 2025 and should improve further over time.

Table 2: Delays between event recorded and data availability within the IDI

Data collection	Update frequency	Delay range (months)
ACC	Three times a year	3-7
Corrections	Two times a year	3-9
Customs	Three times a year	4-8
Department of Internal Affairs	Three times a year	3-7
Early Start	Annual	10-22
Housing NZ	Three times a year	3-7
Inland Revenue	Monthly	3-4
Education	Annual	6-18
Health Before School Checks	Annual	8-20
Health Cancer	Annual	9-22
Health INTERRAI	Annual	4-16
Health Maternity	Annual	22-34
Health Mortality	Three times a year	48-52
Health NES enrolment	Three times a year	3-7
Health immunisation register	Annual	12-24
Health non-admitted patients	Annual	9-21
Health pharmaceuticals	Three times a year	6-10
Health PRIMHD	Three times a year	9-13
Health Private hospital	Annual	27-39
Health Public hospital	Three times a year	9-13
Ministry of Justice	Annual	6-18
Ministry of Social Development	Monthly	3-4
NZTA	Three times a year	3-7
Oranga Tamariki	Three times a year	3-7
Police	Three times a year	4-8
Student Loans	Annual	9-21
Working for Families	Monthly	3-4

Appendix 2: Current measures

Table 3 lists the concepts from which the measures used in our repeatable analysis are derived. Each concept generates multiple measures across time windows and family links (for example: mother's and father's experiences). The indicators used for the time-series results are also used for the matching process – in total, over 130 variables. This list was accurate at time of writing but may expand over time.

Each of these measures is calculated over some period. For cross-section results this is most often the previous twelve months, but it can also be an indication of any past experience ('ever occurred'). For time-series results the default design uses six time periods: two 12-month periods before the reference date (1-12 months before, and 13-24 months before), and four 12-month periods after the reference date.

Most concepts are used to generate measures for both the cross-section and the time-series results. However, the concepts may be converted to measures in different ways: cross-section results tend to focus on whether a person had an experience, time-series results tend to count the number of events, days, or dollars associated with an experience. This difference reflects the kinds of insights each set of results is designed to support.

Table 3: List of measures used in repeatable analysis

Compact name	Cross-section	Time-series	Parent	Explanation
Age or age grouping	x	x		Age is used for matching processes. Five-year age groupings (0-4, 5-9, 10-14, 15-19, …, 75-79, 80+) are used for summaries
Deprivation	x	x		Lives in meshblock with a deprivation score of 1-2 (least deprived), 3-4, 5-6, 7-8, or 9-10 (most deprived)
Disability	x	x		The present/absence of an indication of disability
Ethnicity	x	x		Total response ethnicity (European, Māori, Pacific, Asian, MELAA, or Other), people can have more than one ethnicity
Sex	x	x		Males or Female (may be sex or gender depending on the source, most sources record sex)
REGC	x			Regional council area of residence
TALB	x			Territorial or Local Board (for Auckland) area of residence
JSWR benefit	x	x	x	On a jobseeker work ready benefit
SPS benefit	x	x	x	On a sole parent benefit
main benefit	x	x	x	On a main benefit
employed	x	x	x	Employed for those age 15+
NEET	x	x	x	Not in Employment, Education, or Training (NEET) for those age 15-24
dl holder full current	x		x	Has a full Driver licence

Compact name	Cross-section	Time-series	Parent	Explanation
dl holder current	x		x	Has any Driver licence
hospitalisations	x	x	x	Admitted to Hospital
pharmaceutical	x	x	x	Pharmaceutical dispensing (number of pharmaceuticals or dispensing occasions)
pho enrolment	x		x	Enrolled with a Primary Health Organisation (PHO)
ACC injuries	x	x	x	Injury claims to ACC
MHA service	x	x	x	Interaction with a specialist mental health service
ED visits	x	x	x	Visited the emergency department of a hospital
avoid hospitalisation (ASH & PAH)	x	x	x	Avoidable hospitalisations (ASH & PAH)
below med income	x		x	Individual had income below the median
income period		x	x	Total dollars of income
W&S income	x	x	x	Income from wages and salary (W&S)
OT report concern ever	x			Ever had a report of concern to Oranga Tamariki (OT)
OT investigation ever	x			Ever has an investigation by OT
OT FGC	x	x		Ever had Family Group Conference with OT
OT placement ever	x			Ever in OT care / on placement with OT
YJ FGC	x	x		Ever had Family Group Conferences via Youth Justice
younger sibling mother	x			Does the person have a younger sibling via their mother
younger sibling father	x			Does the person have a younger sibling via their father
crime victimisations	x	x	x	Recorded as victim of a crime
FVSV user	x	x	x	Recorded as a user of family or sexual violence (known to be an under count)
FVSV victim	x	x	x	Recorded as a victim of family or sexual violence (known to be an under count)
offences	x	x	x	Prior offences proceeded by Police
alcohol offences	x	x	x	Alcohol offences proceeded by Police
drug offences	x	x	x	Drug offences proceeded by Police
violent offences	x	x	x	Violent offences proceeded by Police
non-violent offences	x	x	x	Non-violent offences proceeded by Police
police interactions any	x	x	x	One or more Police interactions in any capacity
court charges	x	x	x	Court charges
corrections any	x	x	x	Under management by Corrections
corrections incarcerated	x	x	x	Incarcerated or held on remand by Corrections
social housing	x	x	x	Living in social housing or community provided housing
emergency housing	x	x	x	Primary applicant for an emergency housing grant
highest qual	x		x	Current highest qualification

Compact name	Cross-section	Time-series	Parent	Explanation
tertiary study	x	x	x	Enrolled in tertiary education
school enrol	x	x		Whether enrolled in school (primary, intermediate, or secondary)
regular attendance	x			Student has regular attendance for at least one term
chronic absence	x			Student has chronic attendance for at least one term
SSEE exclude expel	x			Ever excluded or expelled from school
SSEE suspended	x			Ever suspended from school
SSEE standdown	x			Ever stood down from school
learning support	x	x		Receipt of learning supports via school
Alt-ed	x			Enrolment in Alternative Education
attendance service	x	x		Supported by Attendance Service
2+ non-structural school changes	x			At least two non-structural (e.g. primary to secondary) school moves, this tends to indicate inconsistent connection to education

Appendix 3: Detailed design choices for matching

Developing our matching methodology required a range of detailed design decisions. For those wanting to understand more of the nuance within our methodology, we document many of these decisions here:

- Because the participant and potential participant cohorts are defined not just by who, but also by when – the reference date or year – matching was also designed to handle a temporal component.
    - Recall that while the participant and potential participant cohorts have a reference date, the possible comparison and forecast cohorts have only a reference year. To match on timing, a consistent set of dates is easiest. We term these dates the matching dates and restrict them to be quarter end dates (quarters were chosen to balance the precision from adding more points in time against the effort to compute them).

      Each input record in the participant and potential participant cohorts has its matching date created by mapping its reference date to the end of the most recent quarter. For example, a reference date in April would be mapped to a matching date of 31 March in the same year.

      Each input record in the possible comparison and forecast cohorts becomes four records for matching: with matching dates at the end of each quarter in the reference year. Note that as input records are person-year pairs, one person can appear in multiple input years, each of which becomes four records for matching.
    - For matching the possible comparison cohort to the participant cohort, matching is then restricted to records that have the exact same matching date and the exact same age on the matching date. By requiring the same date, we ensure consistency in experiences of social and economic conditions. By requiring the same age, we ensure consistency in life stage.
    - For matching the possible forecast cohort to the potential participant cohort, matching is only restricted to records that have the exact same age on the matching date. This ensures the matched cohort has a range of ages that are consist with the range of ages in the potential participant cohort. Because the possible forecast cohort is drawn from a past time period, it does not make sense to require an exact match on date.
- By default, we do not restrict the matching methodology to ensure exact matches on measures other than age and matching date. This is consistent with expert advice we received to restrict the methodology only for the essential attributes, and to use all other attributes to access the quality of the matching.
    - The matching algorithm we use has options to further restrict which records can match. Most common among these are requiring exact matches on additional attributes and using callipers to restrict how far apart matched records can be on certain attributes. Such options could be activated as part of the repeatable analysis if required.

- While not imposed within the matching methodology, there are implicit restrictions on matching by how the possible comparison and forecast cohorts are defined. For example, if an organisation only works with people from a specific locality or ethnicity and this restriction is used to define the possible comparison and forecast cohorts, then we are implicitly requiring an exact match on locality or ethnicity.
- In a small number of cases a record may have no suitable matches. Such records go unmatched. For example, if an organisation focused on people 17-18 years old, then this age range might be used when defining the possible comparison and forecast cohorts. If the organisation later works with someone 20 years old as a special case, then there might be no records with a matching age and hence no match would be created. In our experience so far, this kind of non-matching affects less than 1% of participant records.

- Limited common support (of propensity scores) can also lead to cases where there are no suitable matches. This occurs where the treated cohort contains people for whom there is no one with similar characteristics in the control cohort.

  As we do not require exact or near matching of propensity scores, cases of limited common support to not automatically lead to non-matched records. Instead, we review the common support of the cohorts pre- and post-matching.

  Because the IDI covers the entire New Zealand population, it contains data on many people with similar experiences. So, a lack of common support most likely indicates that we have been too restrictive in defining the possible comparison or possible forecast cohorts, rather than that no suitable matches exist within the IDI. Hence, where a review identifies a lack of common support, we are most likely to reassess and relax how these cohorts were defined.
- It is common design for matching methods to select one matched record for each record in the (potential) participant cohort. As larger cohort sizes provide greater statistical power and confidentiality protection for cohort members, we increased this to two matched records for each record in the (potential) participant cohort.
- Our matching methodology selects matches without replacement. This means that each record in the possible comparison or forecast cohorts can only be selected once for inclusion in the corresponding matched cohort. Recall that a record is defined by both an individual and a matching date. So, even though each record can only be used once, a single individual can appear in the matched cohort more than once: in records with different matching dates.

  The extent to which this happens is reported on when assessing the quality of a match. In the cases we have observed so far, at least 90% of selected matches are unique individuals, so we have no reason to expect this to introduce significant bias into our results.
- As matching is an established technique, there are standard ways to assess the quality of a match. Our repeatable analysis includes the automated creation of a report containing several such quality measures. This makes it straightforward for the researcher to check that nothing unexpected has occurred during matching.

  The matching quality report includes counts of matched and unmatched records, counts of distinct underlying people, and plots of the distribution of likelihood/propensity scores before and after matching. Inspection of these plots should show common support and a much closer match in distributions following the removal of the unmatched records.

- The report also includes tables of standardized mean differences (SMD), with an overview visualised in a Love plot. While SMD less than 0.1 are often used as an indication of good balance, that our matching process includes over 130 measures means it is unrealistic to expect every measure to obtain an SMD less than 0.1.

  When reviewing these results, we look for two key things:

  - Simplest is a heuristic based on what distribution of SMD we might expect given the large number of measures we use. Testing by randomly assigning records to treat and control suggests that some differences from perfect balance are to be expected within normal variation. Given the large number of measures we use, for most applications we can be content to proceed with having at least 80% of measures with SMD less than 0.1, less than 15% of measures with SMD between 0.1 and 0.2 – indicating moderate imbalance, and less than 5% of measures with SMD over 0.2 – indicating poor balance.
  - More important, is that the measures of most relevance for the application are balanced and that there are obvious reasons for the measures with poor balance. For example, in the fictional case of TuitionNZ: the tutoring focus on the organisation means that we would want to see good balance on qualification and educational attainment measures. As TuitionNZ is focused on Christchurch, we might expect many of their participants to live in major urban locations. If a possible comparison cohort was used that included people outside of Christchurch, then it would be reasonable to expect measures related to urban-ness to be less balanced.