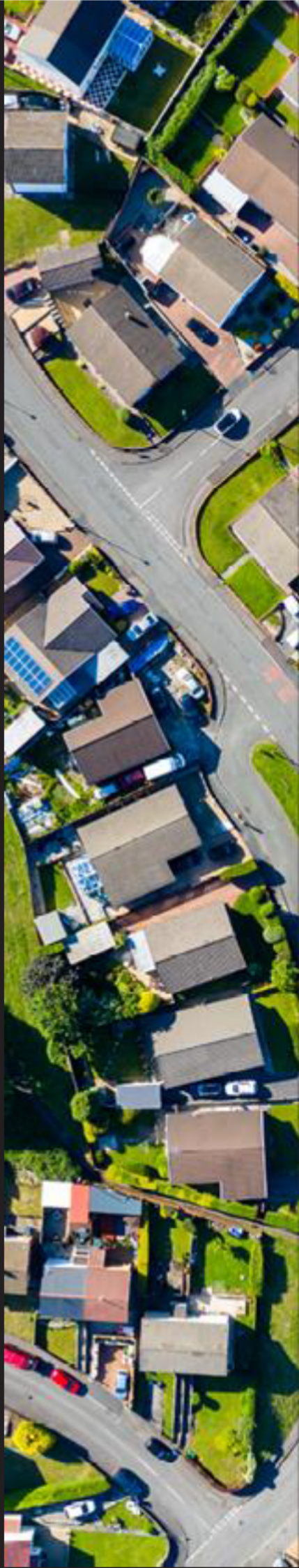
TAYLOR FRY

We find the patterns that matter.



Independent review of Kāinga Ora

Population comparison and outcomes research

8/03/2024

Document classification: Client use

GST 109 891 150
NZBN 942 903 052 9700

www.taylorfry.co.nz

Taylor Fry Pty Ltd

ISO 27001 INFO SEC Certified System

Climate Conscious Certified carbon offsets by South Pole

TAYLOR FRY

Table of contents

IDI disclaimer

These results are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI) which is carefully managed by Stats NZ. For more information about the IDI please visit https://www.stats.govt.nz/integrated-data/.

The results are based in part on tax data supplied by Inland Revenue to Stats NZ under the Tax Administration Act 1994 for statistical purposes. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

Confidentiality

To protect confidentiality, data is rounded to a multiple of 3. Results with small totals are suppressed. As a result, data in tables and figures may not add up exactly to population totals and table totals may differ slightly throughout the report.

1 Executive summary

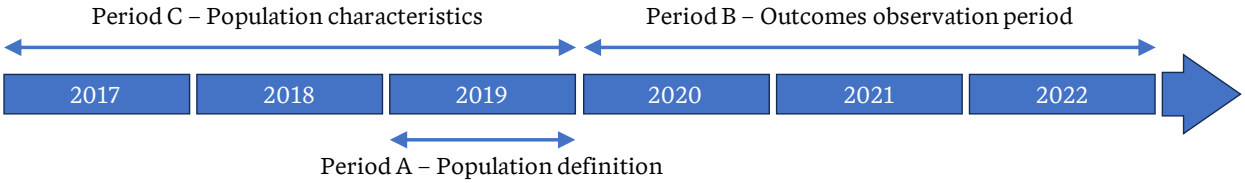
The Government has commissioned an independent review of Kāinga Ora to look into the Crown agency's financial situation, procurement and asset management.

To support the review, Taylor Fry have been commissioned by the Ministry of Housing and Urban Development to perform research to:

- Compare, with respect to a range of demographic and government service use factors, the Kāinga Ora population ('KO'), the Community Housing Provider population ('CHP') and the Accommodation Supplement population ('AS').
- Understand how outcomes differ for these populations controlling for, as best as possible with the available data, differences in the underlying populations.

Figure 1 illustrates three time periods which have been used to define and construct the research:

Figure 1 – Data and outcomes timeline



- Period A – The year to 31 December 2019 – This has been used to define the KO, CHP and AS populations.
- Period B – The three years to 31 December 2022 – This defines the observation period over which we observe outcomes for people in the three populations.
- Period C – The three years to 31 December 2019 – This defines the time window over which we define characteristics that describe people in the three populations.

1.1 Population comparison

Key points

- **The KO and CHP populations are materially similar**. Across almost all factors, the populations are, on average, very similar.
- A notable exception is region. The CHP population are more heavily weighted towards Auckland than the KO population, reflecting the location spread of CHP and KO public housing stock.
- **The AS population is notably different to the CHP and KO populations** in several ways:
  - Demographics – The CHP and KO populations are more weighted towards Auckland than the AS population and are much more weighted towards Pacific People. These two points are somewhat related – A relatively high proportion of Pacific People reside in Auckland.
  - Government service use – **The AS population appears to have more complex needs than the CHP and KO populations**. On average, the AS population:
    - Spends a higher proportion of time on the JobSeeker Support benefit
    - Has lower average income
    - Has a lower proportion who have attained NCEA level 3

- Has a higher proportion who interacted with child protection services as a child
- Has a higher proportion accessing mental health and additions services.

The exception to this observation is that the KO population has a higher proportion of people with a criminal conviction in the past three years.

1.2 Outcomes analysis

In this section, we describe the results of exploratory modelling, which attempts to identify which factors explain variation in outcomes for the AS and KO populations and, specifically, the extent to which housing status (AS or KO) explains variation.

Three separate models were developed initially, to predict taxable income, conviction rates, and mental health service use over a three-year observation period to 31 December 2022. Note that the modelling cannot be used to estimate causal effects; i.e. even if a factor explains a high proportion of variation in an outcome, it does not necessarily mean that one causes the other.

Key points

- **The model used in this research suggests that the variation in taxable income, conviction rates, and mental health service usage can be primarily explained by the equivalent variables of the preceding three years** (to 31 December 2019). For example, the taxable income earned in the three-year period to 31 December 2019 can explain almost all of the income variation in the three-year period to 31 December 2022.
- **The model used in this research suggests that housing status (AS or KO) explains very little variation in taxable income, conviction rates, and mental health service use.** Or expressed in a different way, whether a person receives AS for a full year or is in a KO public house for a full year appears to explain very little variation in income, conviction rates and mental health service use over the subsequent three-year period.
- More importantly, the research findings suggest that there are differences between the KO and AS populations, and they experience materially different outcomes. These differences can be explained by the underlying differences in the populations and their prior outcomes.

Due to the limited scope imposed by a short timeframe, this research did not disaggregate sub-cohorts with different characteristics in the KO and AS populations and explored a limited set of outcome variables. Therefore, variation in outcomes experienced by different sub-cohorts of the KO and AS populations are not presented in this research.

Note that this research does not rule out the possibility that there is a link between housing status and these outcomes, nor does it rule out the possibility that there is a link between housing status and other outcomes not considered in this research, e.g., specific health conditions, or housing quality.

Nevertheless, this research has identified interesting indications that warrant further investigation that will lead to insights on better aligning housing support and people. We recommend that the further work be carried out.

Note 2019 research by the Social Wellbeing Agency exploring the impact of public housing on people's wellbeing[1]. The research identified two key findings:

- Housing conditions generally improve for people placed in public housing
- Life satisfaction improves for people placed in public housing.

[1] Social Wellbeing Agency (2019) *Measuring the impact of social housing placement on wellbeing*

2 Scope and research approach

The Government has commissioned an independent review of Kāinga Ora to look into the Crown agency's financial situation, procurement and asset management.

The review is being supported by the Ministry of Housing and Urban Development (the 'Ministry') and The Treasury.

To support the review, Taylor Fry have been commissioned by the Ministry to perform research to:

- Compare, with respect to a range of factors, the:
  - The Kāinga Ora population – The public housing population in Kāinga Ora ('KO') properties. We define somebody who resides in a KO property as being in the KO state.
  - The Community Housing Provider population – The public housing population in Community Housing Provider ('CHP') properties. We define somebody who resides in a CHP property as being in the CHP state.
  - The Accommodation Supplement population – The population receiving Accommodation Supplement ('AS'). Note that about two-thirds of AS recipients are renters[2]. We define somebody who is receiving AS as being in the AS state.
- Understand how outcomes differ for these populations controlling for, as best as possible with the available data, differences in the underlying populations.

The research is also intended to support broader policy considerations of who public housing is best targeted to and its role within the continuum of housing supports (including monetary supports such as AS). While this document is for public release, its original purpose was to provide technical information to a specialised audience.

2.1 Outline of approach

The research incorporates two phases covered by Parts A and B in this report:

- Population comparison – Descriptive statistics comparing the three populations with respect to a range of factors. The populations incorporate people aged 16 and above.
- Outcomes analysis:
  - Descriptive statistics to compare outcomes for the three populations
  - Exploratory modelling techniques to compare outcomes for the three populations after controlling for differences in the underlying populations.

We used Stats NZ's Integrated Data Infrastructure ('IDI') to perform the research.

Figure 1 illustrates three time periods which have been used to define and construct the research:

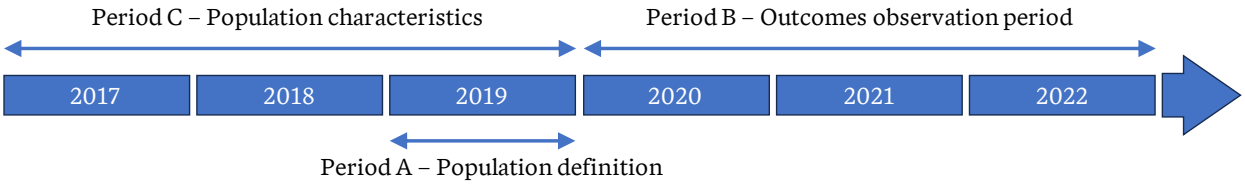
- Period A – The year to 31 December 2019 – This has been used to define the KO, CHP and AS populations. Specifically, people who are in the KO state for the whole of period A are defined to be in the KO population for this research. And equivalently for the CHP and AS populations.
- Period B – The three years to 31 December 2022 – This defines the observation period over which we observe outcomes for people in the three populations.

[2] Hyslop D and Rea D (2018) *Do housing allowances increase rents? Evidence from a discrete policy change*

- Period C – The three years to 31 December 2019 – This defines the time window over which we define characteristics that describe people in the three populations.

The periods have been chosen to balance the need for recency of analysis with the need to have a long enough observation period for outcomes to materialise over.

Figure 1 – Data and outcomes timeline

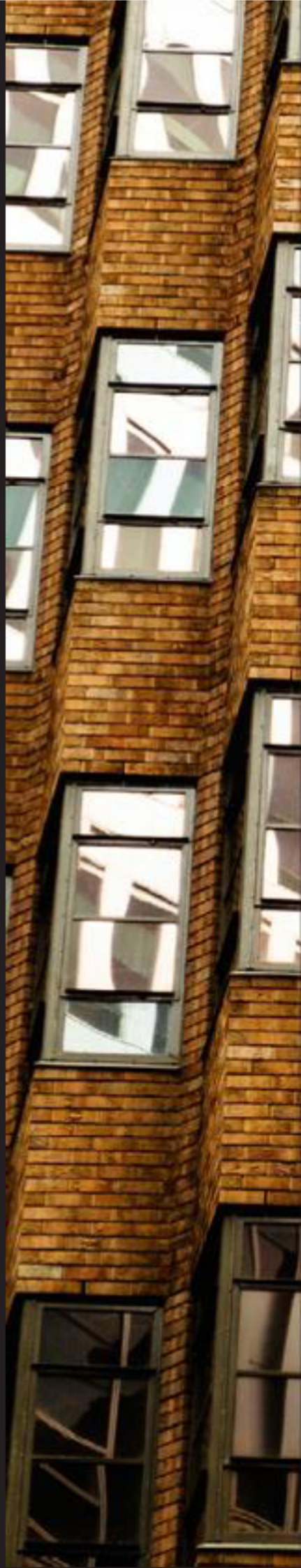


The scope of our commissioned work does not include a full technical write-up of the research approach. However, we provide more detail in Section 4.6. Also, the research code in the IDI is well organised and commented.

Some limitations of the research are also described in Section 4.2, noting that the research approach has had been designed to ensure it was achievable in a timeframe for it to inform the Kāinga Ora review.

PART A

Population comparison



3 Kāinga Ora, Community Housing Provider and Accommodation Supplement populations

In this section, we compare the KO, CHP and AS populations in relation to a range of demographic and government service use factors.

Key points

- The KO and CHP populations are materially similar. Across almost all factors, the populations are, on average, very similar.
- A notable exception is region. The CHP population are more heavily weighted towards Auckland than the KO population, reflecting the location spread of CHP and KO public housing stock.
- The AS population is notably different to the CHP and KO populations in several ways:
  - Demographics – The CHP and KO populations are more weighted towards Auckland than the AS population and are much more weighted towards Pacific People. These two points are somewhat related – a relatively high proportion of Pacific People reside in Auckland.
  - Government service use – The AS population appears to have more complex needs than the CHP and KO populations. On average, the AS population:
    - Spends a higher proportion of time on the JobSeeker Support benefit
    - Has lower average income
    - Has a lower proportion who have attained NCEA level 3
    - Has a higher proportion who interacted with child protection services as a child
    - Has a higher proportion accessing mental health and additions services.

    The exception to this observation is that the KO population has a higher proportion of people with a criminal conviction in the past three years.

In this section, we present results of descriptive analysis to compare the KO, CHP and AS populations. We show analysis by:

- Demographic variables
- Variables based on Government service use data.

The analysis is purely descriptive (no modelling or standardisation is involved). The three populations are defined as having been in the relevant state for the whole of the year to 31 December 2019.

Note that while descriptive analysis can be informative, it also has the potential to be misinterpreted. Descriptive analysis charts that appear to show correlations between variables may actually be highlighting correlations with other variables. For example, Pacific People are over-represented in the KO and CHP populations. However, this is at least partly because a high proportion of Pacific People live in Auckland, where housing affordability issues are acute, and a relatively high proportion of public housing is situated.

3.1 Population totals

Table 1 shows the size of the three populations (defined as being in that state for the whole of the year to 31 December 2019).

Table 1 – Population totals over year ending 31 December 2019

Kāinga Ora	Community Housing Provider	Accommodation Supplement
85,836	7,374	220,155

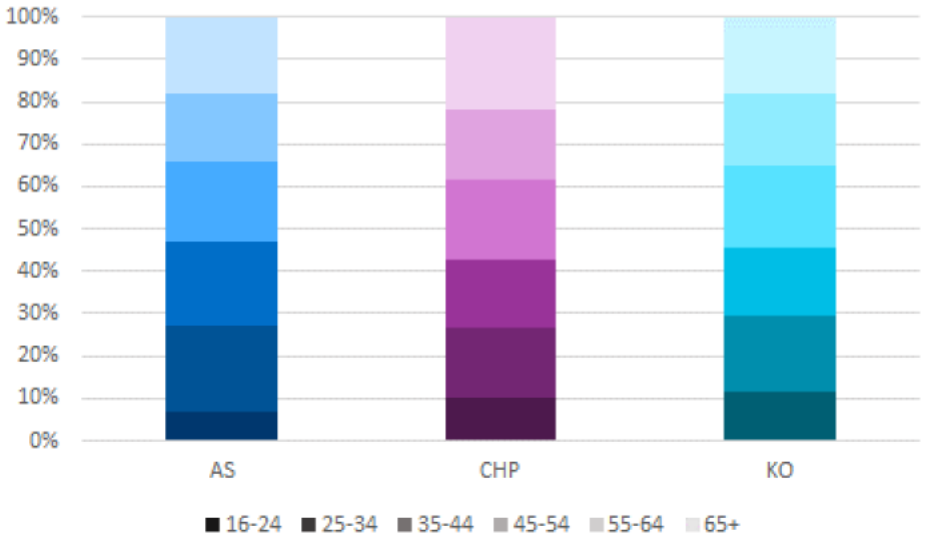
Note that:

- The KO and CHP POPULATIONS include all people aged 16 and over who are named on the tenancy, which notionally includes all people aged 16 and over residing in the house (excluding any boarders)
- The AS population includes all people aged 16 and over and receiving AS directly. There may be other people aged 16 and above in the same household who are not captured by this definition.
- 300 data points are omitted out of 313,665 because individuals did not link to the IDI spine.

3.2 Demographic variables

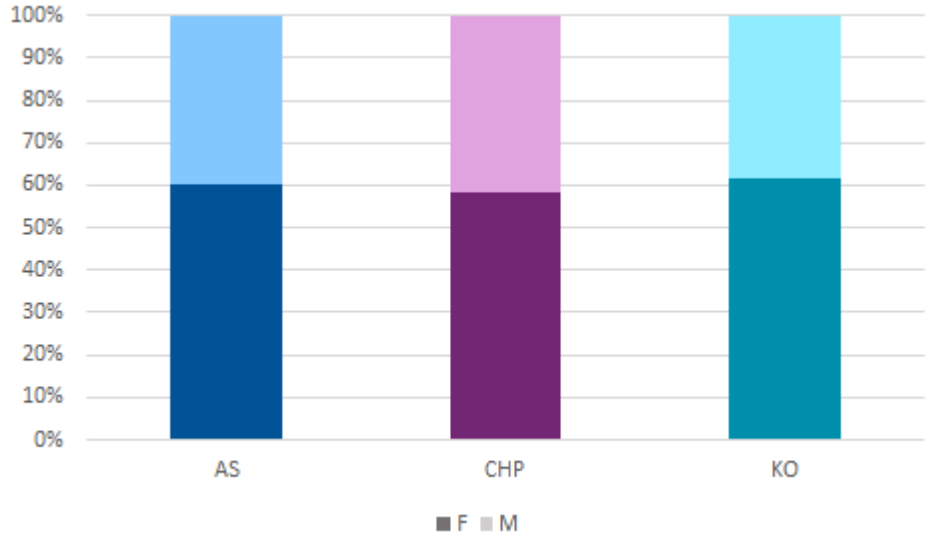
In this section, we describe and compare the populations with respect to demographic variables. Age (Figure 2), gender (Figure 3) and ethnicity (Figure 4) variables are sourced from the IDI personal details table. The region variable (Figure 5) is sourced from the IDI address notification table. These tables are derived from a variety of data sources in the IDI.

Figure 2 – Age



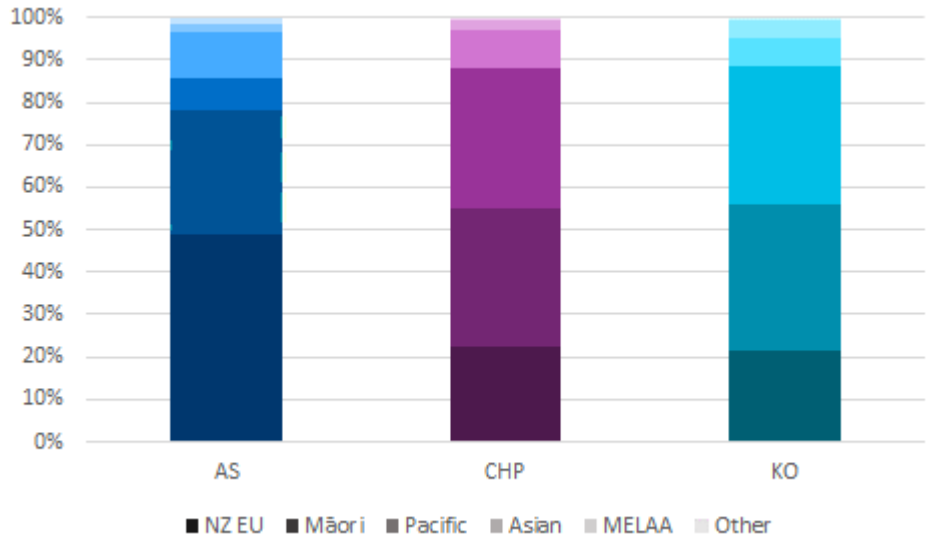
- The age profiles of the three populations are broadly comparable, with some minor differences:
  - The AS population has a lower proportion of 16-24-year-olds, but a higher proportion of 25-44-year-olds
  - The CHP population has a higher proportion of over-65-year-olds.

Figure 3 – Gender



- The gender profiles of the three populations are broadly comparable, with about 60% females.
- The reason this is materially higher than 50% is because all three populations, and the welfare system more generally, contain a large number of sole parents. The vast majority of sole parents are female.

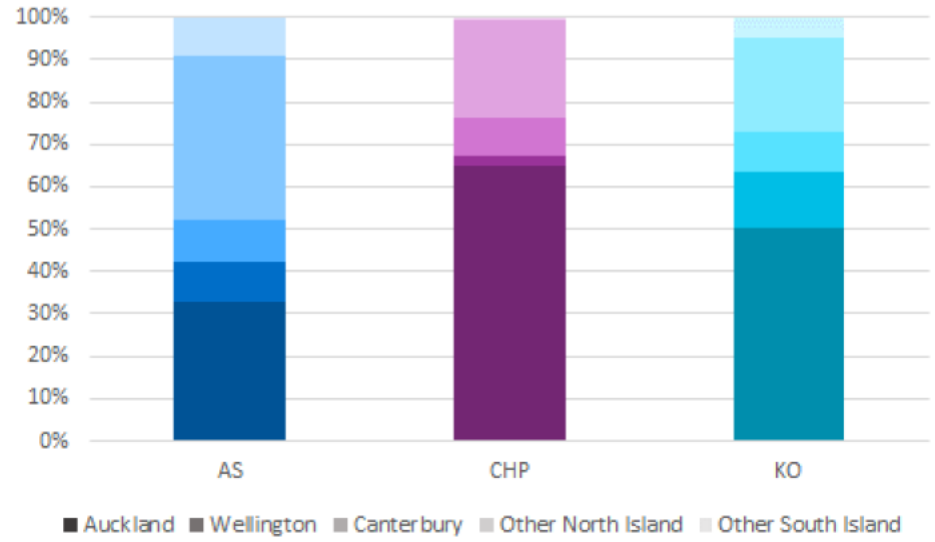
Figure 4 – Prioritised ethnicity[3]



- The ethnicity profile of the CHP and KO populations are materially different to the AS population. A significantly higher proportion of the CHP and KO populations are Pacific People, and a significantly lower proportion are NZ European.
- This partly reflects the regional spread of different ethnicities and, in particular, the high concentration of Pacific People residing in Auckland, where housing affordability issues are acute, and a high proportion of public housing is situated.

[3] People are allocated to a single ethnic group in an order of priority, even if they identify with more than one ethnicity. Our priority ordering is Māori, Pacific, Asian, MELA, Other, NZ European.

Figure 5 – Region



- The CHP and KO populations are heavily skewed to Auckland, reflecting public housing demand and supply factors in the region.
- The CHP population is more skewed to Auckland than the KO population, reflecting the historical supply of public housing in New Zealand and the concentration of community housing providers in Auckland.
- There are very few CHP public housing places in the South Island.

3.3 Government service use

Variables defining government service use in Table 2 are either:

- Defined over a 3-year period before the observation date of 31 December 2019. For example, how many convictions an individual has in the 3 years before 31 December 2019.
- An indicator which describes whether the event has ever occurred in an individual's life, before 31 December 2019. For example, whether an individual has experienced an Oranga Tamariki intervention before 31 December 2019.

Note that some variables are only defined for an age-based subset of the population, reflecting how far back in time the associated data source goes.

In Figure 6 to Figure 25, each variable is shown in charts by population:

- For the total population
- In most cases, by gender and age band for each population.

Table 2 – Government service use variables

Variable	Data source	Time window	Population restriction
JobSeeker Support benefit	Ministry of Social Development	Preceding 3 years	NA

Variable	Data source	Time window	Population restriction
Supported living payment benefit	Ministry of Social Development	Preceding 3 years	NA
Sole parent support benefit	Ministry of Social Development	Preceding 3 years	NA
Taxable income (not including benefit payments)	IRD	Preceding 3 years	NA
Convictions (any)	Ministry of Justice	Preceding 3 years	NA
Highest NCEA level achieved	Ministry of Education	Ever	Age 20 to 25 only (not available for those with schooling outside NZ)
Oranga Tamariki interactions (from notice of concern up to placement)	Oranga Tamariki	Ever	Age 19 or 20 only
Mental health pharmaceutical service use	Ministry of health (pharmaceuticals table)	Preceding 3 years	NA
Mental health acute service use	Ministry of health (PRIMHD database)	Preceding 3 years	NA
Hospital discharges	Ministry of health	Preceding 3 years	NA

3.3.1 Benefit receipt

In Figure 6 to Figure 11, for JobSeeker Support, Sole Parent Support and Supported Living Payment, we show the average percentage of the preceding three years (to 31 December 2019) that people in each population received that benefit, among people who received that benefit at least once in the period.

- In general, the average percentage for the three populations is materially similar for Sole Parent Support (Figure 8 and Figure 9) and Supported Living Payment (Figure 10 to Figure 11), across all age bands and genders
- For Jobseeker (Figure 6 to Figure 7), the pattern is different, with a higher percentage for the AS population, particularly for younger ages and for males
- The proportions that received each benefit at any point in the preceding three years is fairly consistent across the three populations:
  - JobSeeker Support – AS 41.5%, CHP 41.5%, KO 40.2%
  - Sole Parent Support – AS 24.2%, CHP 17.4%, KO 20.6%
  - Supported Living Payment – AS 26.4%, CHP 24.4%, KO 22.3%.

Figure 6 – Average % of preceding 3 years on JobSeeker Support benefit given received Jobseeker benefit in that period and aged under 65

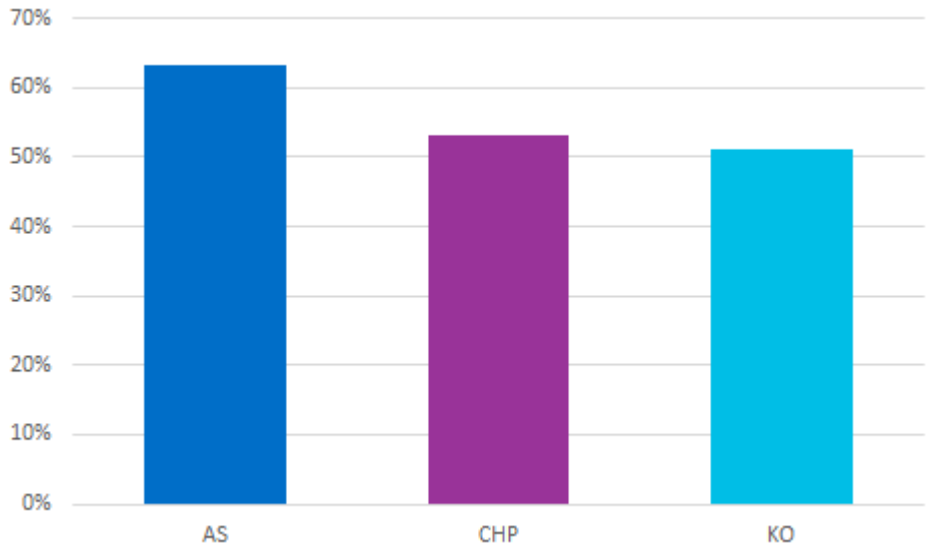


Figure 7 – Average % of preceding 3 years on JobSeeker Support benefit given received Jobseeker benefit in that period – by gender and by age band

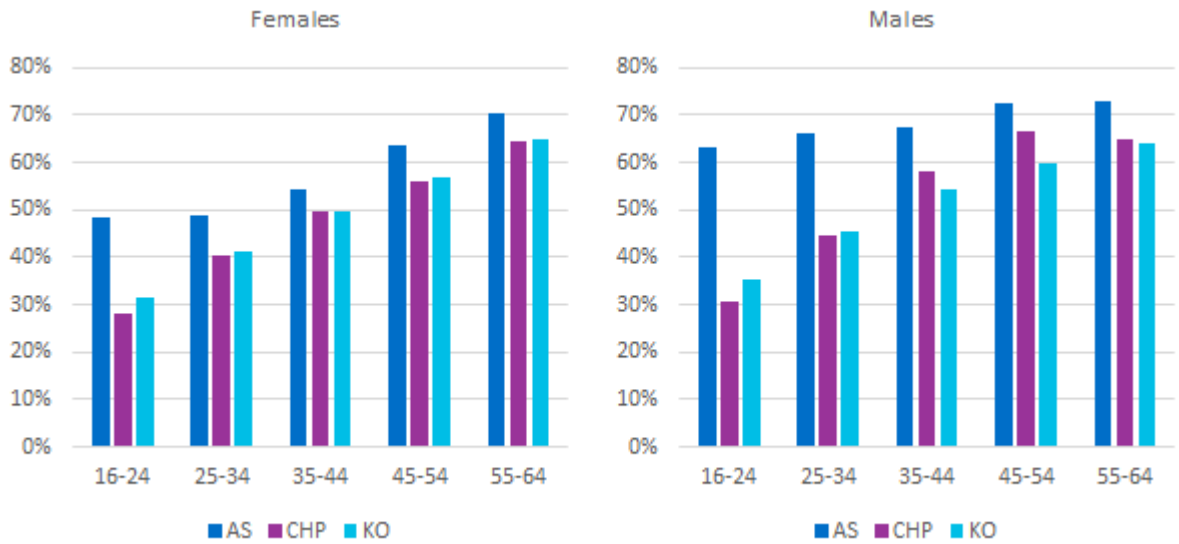


Figure 8 – Average % of preceding 3 years on Sole Parent Support benefit given received Sole Parent Support benefit in that period and aged under 65

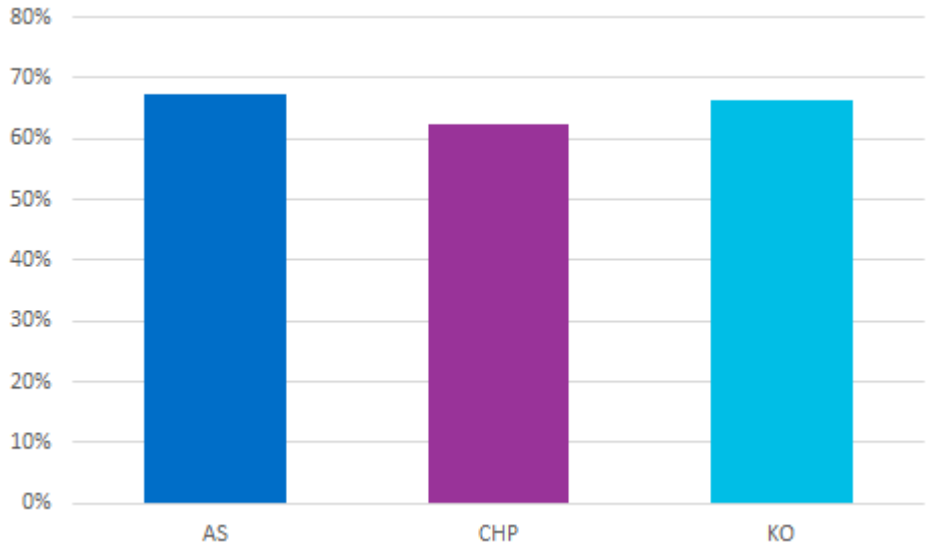


Figure 9 – Average % of preceding 3 years on Sole Parent Support benefit given received Sole Parent Support benefit in that period and under 65 – females by age band

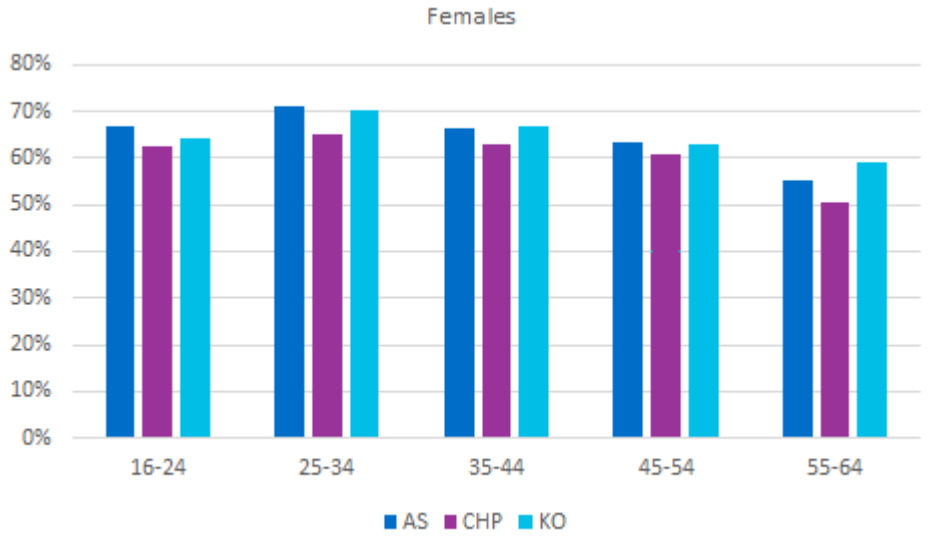


Figure 10 – Average % of preceding 3 years on Supported Living Payment benefit given received Supported Living Payment benefit in that period and aged under 65

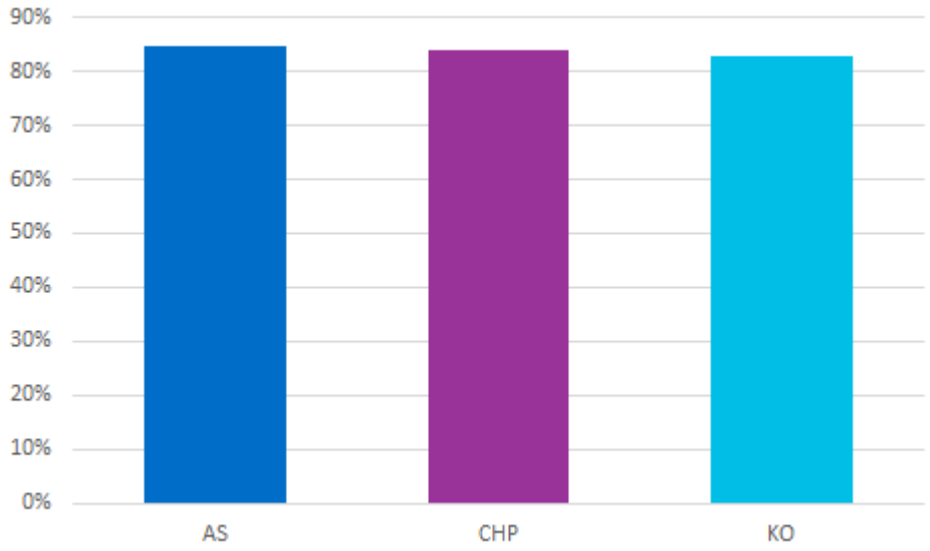
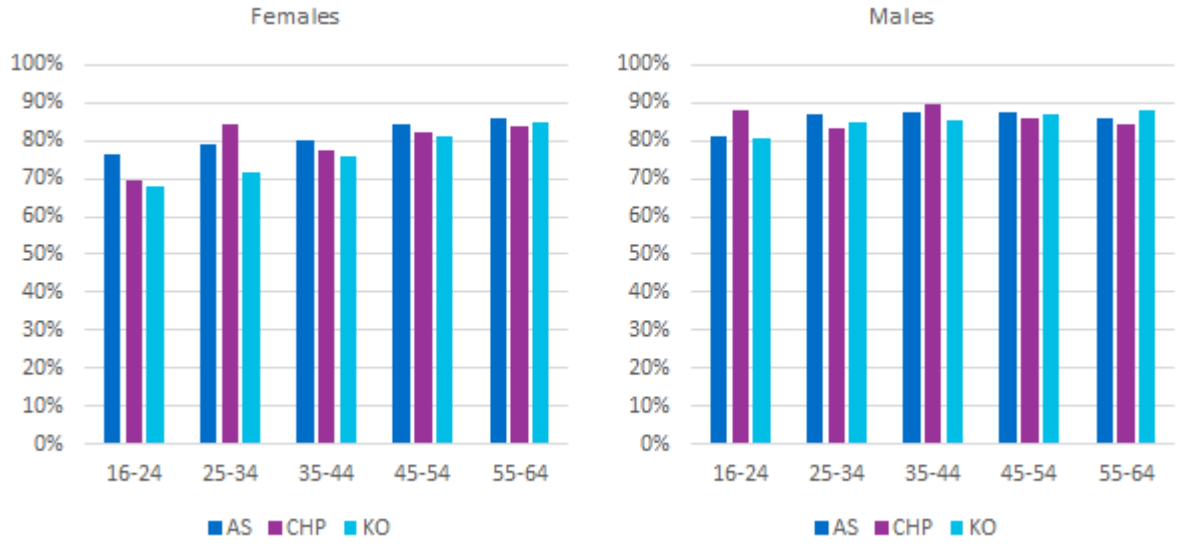


Figure 11 – Average % of preceding 3 years on Supported Living Payment benefit given received Supported Living Payment benefit in that period and aged under 65 – by gender and by age band



3.3.2 Income

In Figure 12 and Figure 13, we show the average taxable income for each population (excluding benefit payments), among people who receive taxable income.

- Incomes for people in the CHP and KO populations are higher than for the AS population. This is particularly true for males.
- Note that the proportions that received taxable income at any point in the preceding three years is fairly consistent across the three populations – AS 66.0%, CHP 71.1%, KO 70.5%.

Figure 12 – Average taxable income (excluding benefit payments) given taxable income in the period and aged under 65

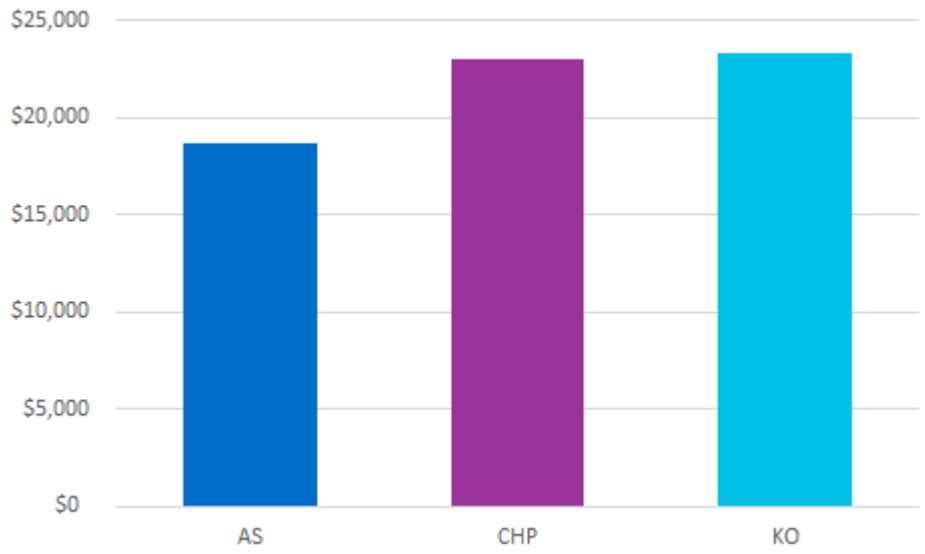
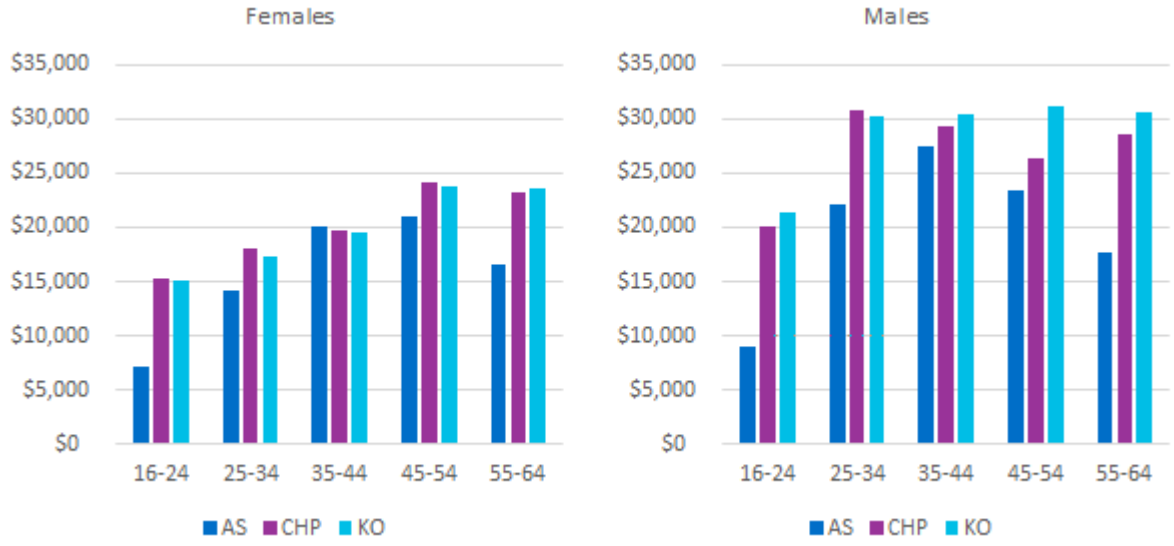


Figure 13 – Average taxable income (excluding benefit payments) given taxable income in the period and under 65 – by gender and by age band



3.3.3 Convictions

In Figure 14 and Figure 15, we show the proportion of people in each population who were convicted of a crime in the preceding three years.

- The proportion for the KO population is higher than for the AS and CHP populations, though by age band and gender the relativities are less clear and consistent. Note that a higher proportion of the KO population are under the age of 35, compared to the AS and CHP populations.

Figure 14 – Proportion with a conviction in the preceding 3 years

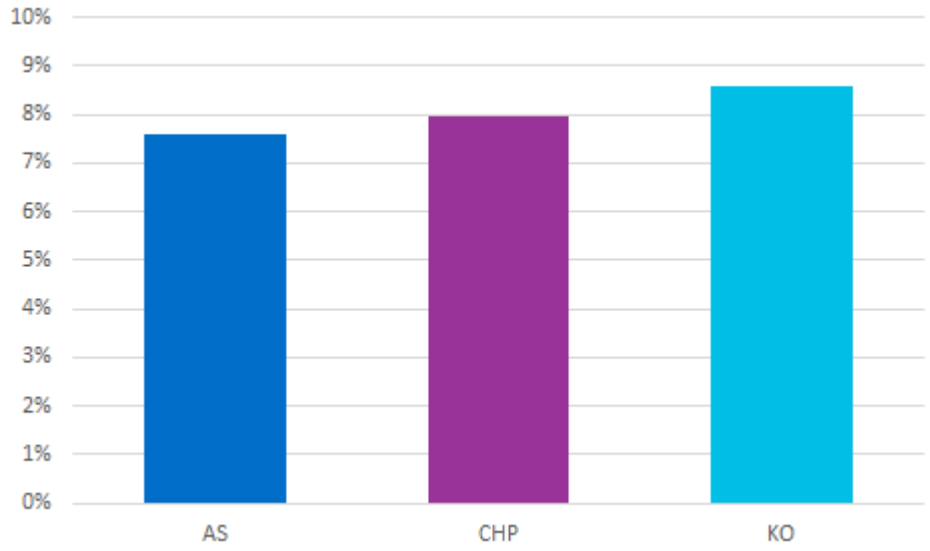
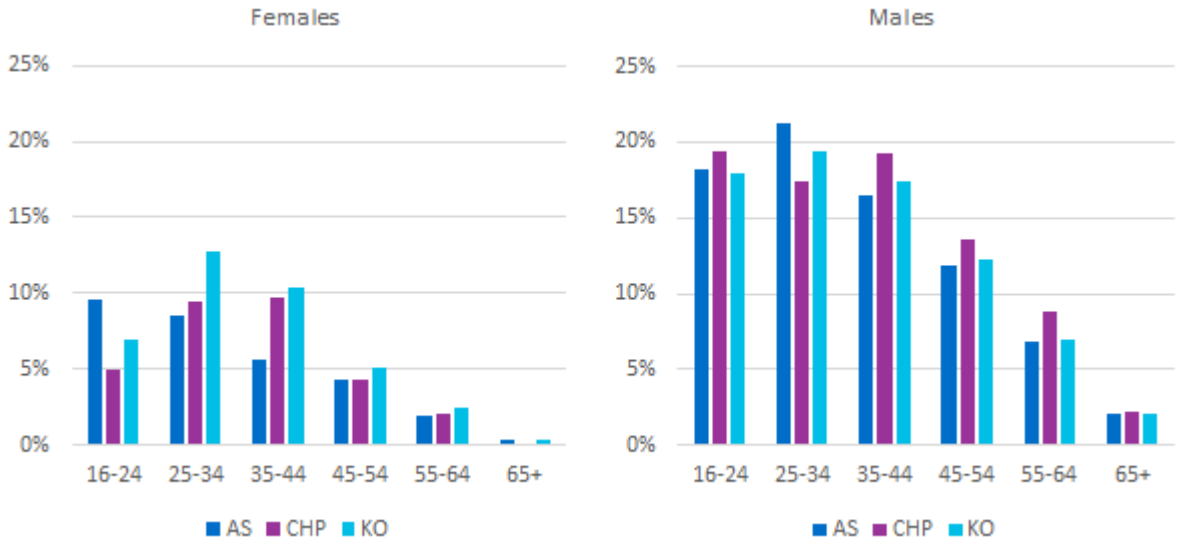


Figure 15 – Proportion with a conviction in the preceding 3 years – by gender and by age band



3.3.4 NCEA level 3 attainment

In Figure 16 and Figure 17, we show the proportion of each population who have attained NCEA level 3 by 31 December 2019. This is restricted to people aged 20 to 25 at 31 December 2019.

- Attainment rates for the CHP and KO populations are substantially higher than for the AS population.

Figure 16 – Proportion with NCEA level 3 or higher (age 20 to 25 at 31 December 2019)

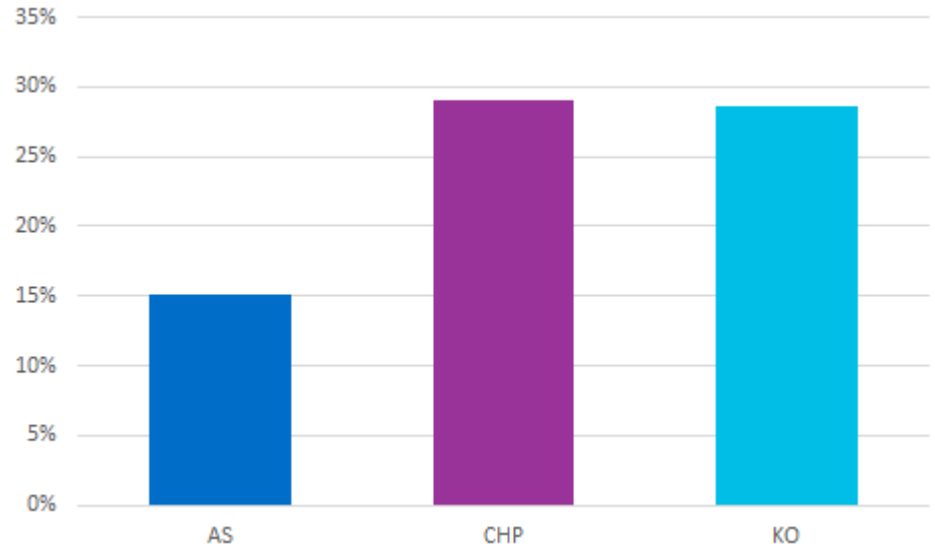


Figure 17 – Proportion with NCEA level 3 or higher (age 20 to 25 at 31 December 2019) – by gender

3.3.5 Oranga Tamariki interactions

In Figure 18 to Figure 20, we show the proportion of each population who interacted with Oranga Tamariki as a child. This is restricted to people aged 19 and 20 at 31 December 2019

- A higher proportion of the AS population interacted with Oranga Tamariki as a child, than the CHP and KO populations.

Figure 18 – Proportion with Oranga Tamariki interactions (age 19 and 20 at 31 December 2019)

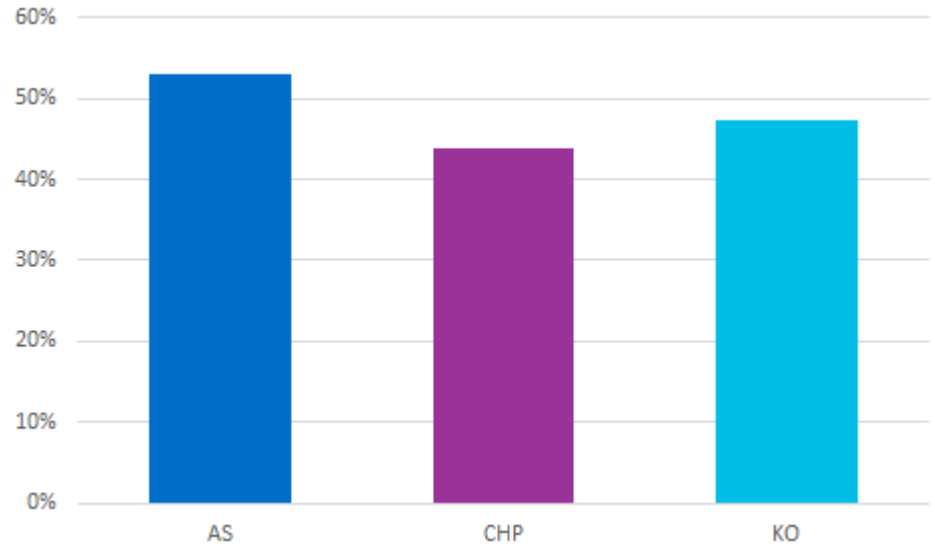
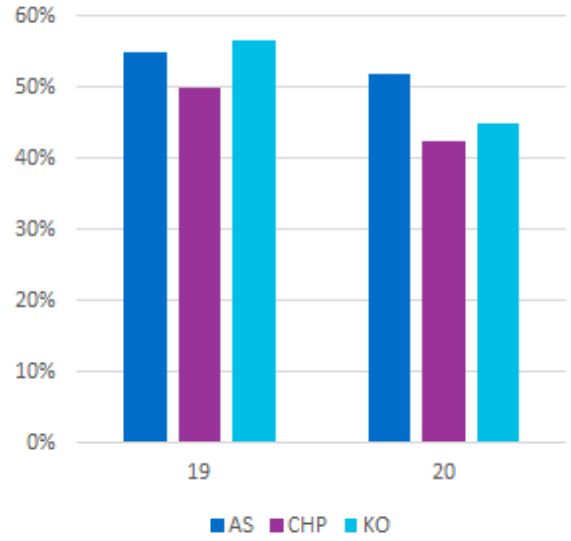


Figure 19 – Proportion with Oranga Tamariki interactions (age 19 and 20 at 31 December 2019) – by age



3.3.6 Mental health and addiction service use

In Figure 20 and Figure 21, we show the proportion of each population who accessed mental health and addiction related pharmaceuticals in the preceding three years.

- A significantly higher proportion of the AS population accessed mental health and addiction related pharmaceuticals, than the CHP and KO populations. This is particularly true at younger ages.

Figure 20 – Proportion with mental health/addiction pharmaceutical use in the preceding 3 years

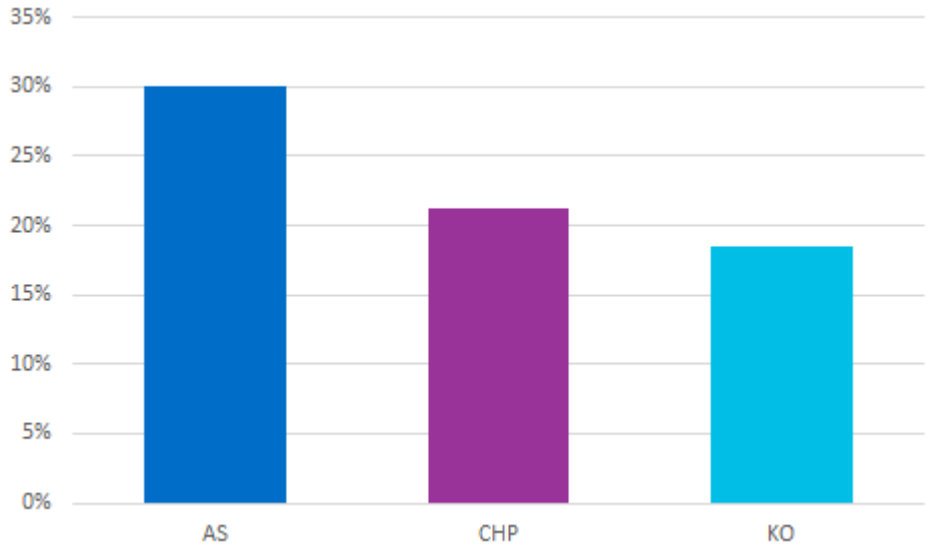
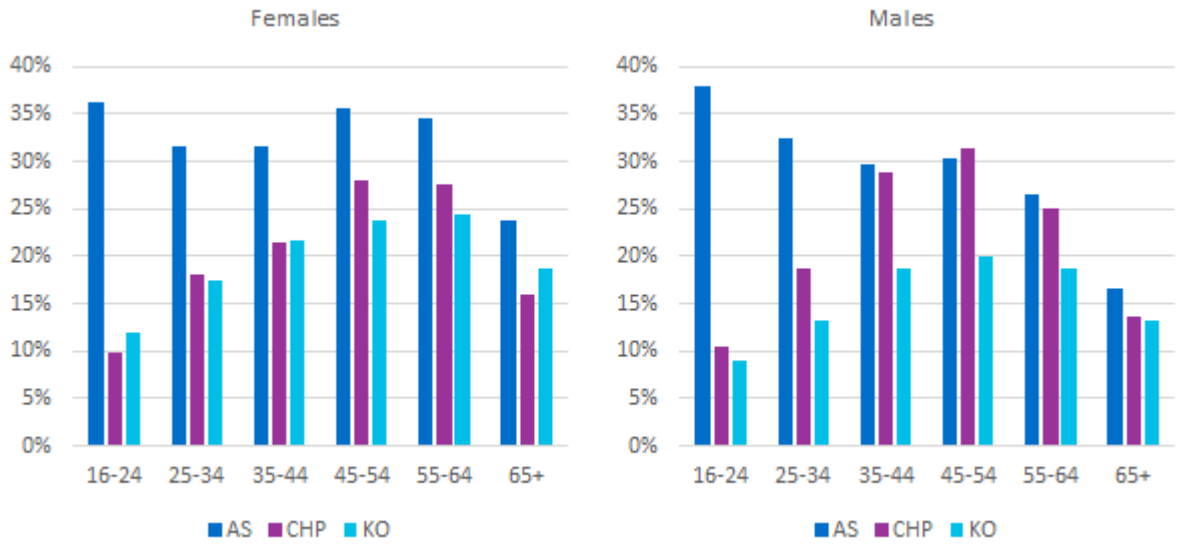


Figure 21 – Proportion with mental health/addiction pharmaceutical use in the preceding 3 years – by age and by gender



In Figure 22 and Figure 23, we show the proportion of each population who accessed acute mental health and addiction services (inpatient and outpatient) in the preceding three years.

- A relatively high proportion of 16-24-year-olds in the AS population access acute mental health services
- For ages 35+, a relatively high proportion of the CHP population access acute mental health services.

Figure 22 – Proportion with acute mental health/addiction service use in the preceding 3 years

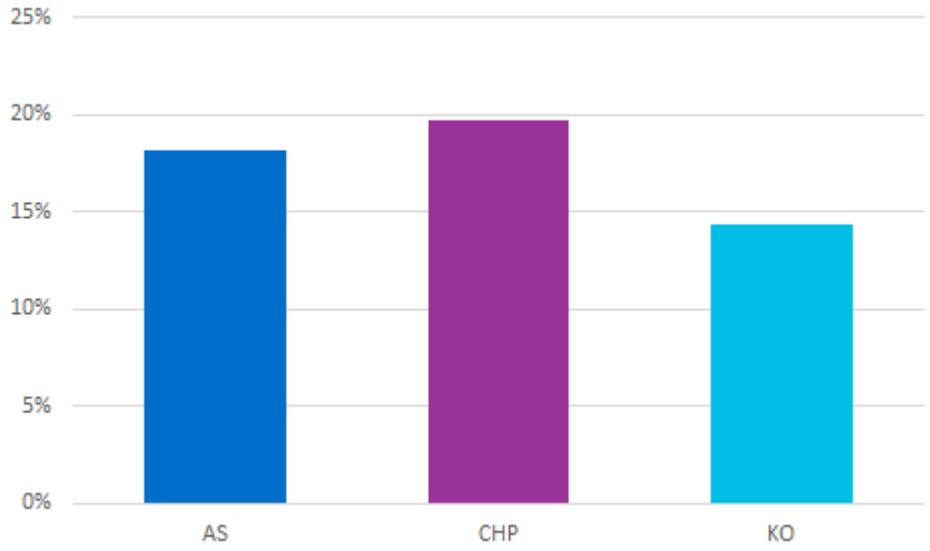
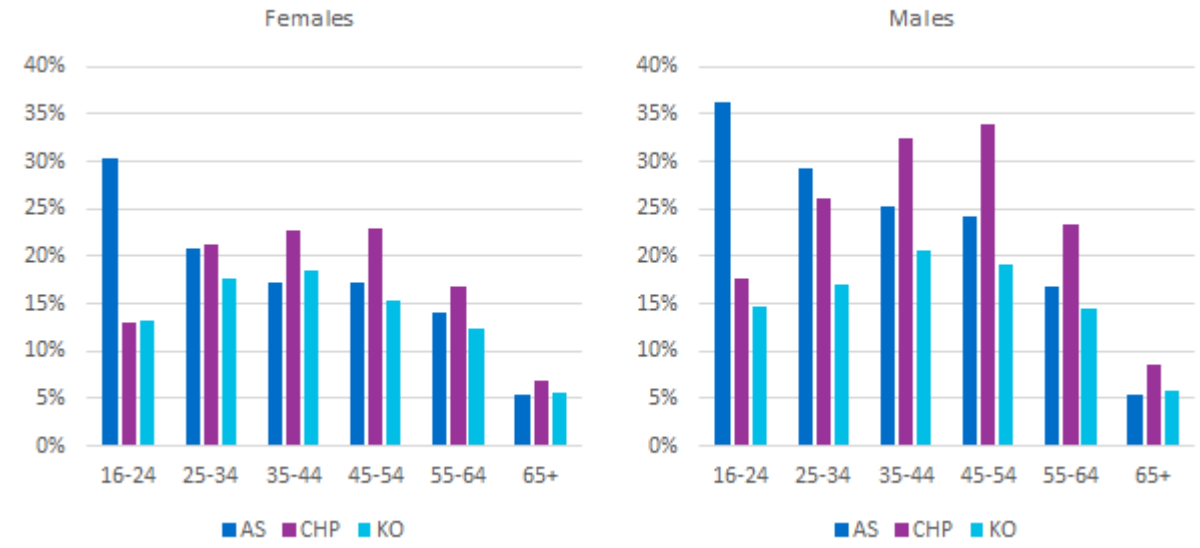


Figure 23 – Proportion with acute mental health/addiction service use in the preceding 3 years – by age and by gender



3.3.7 Hospital discharges

In Figure 24 and Figure 25, we show the proportion of each population who was discharged from hospital at least once in the preceding three years.

- The proportions are similar for the three populations and there are no clear and consistent differences in patterns by gender and age.

Figure 24 – Proportion who have been discharged from hospital at least once in the preceding 3 years

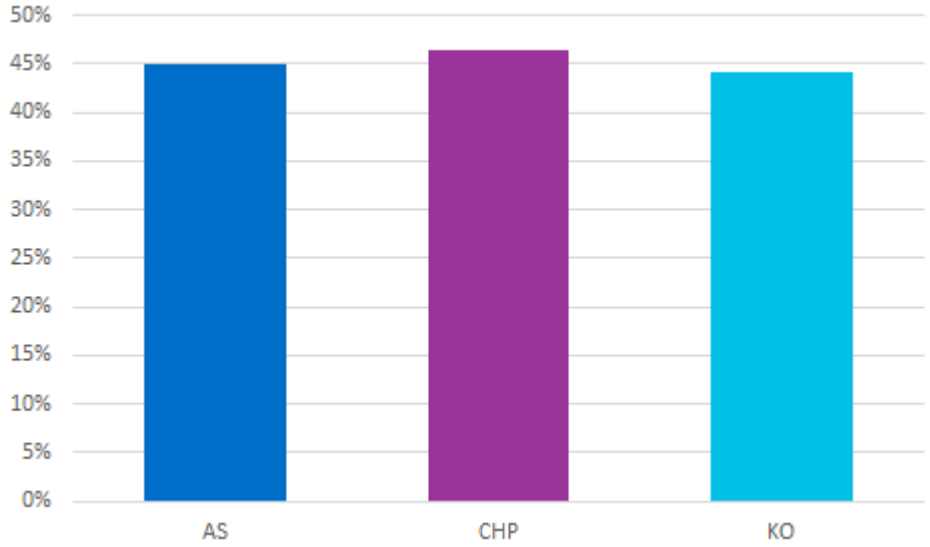
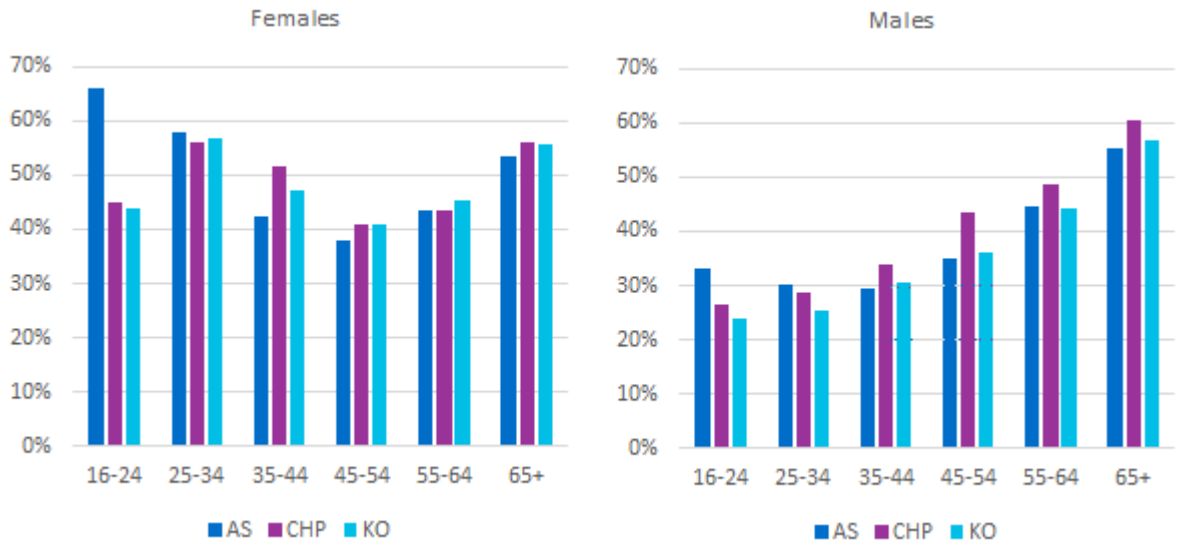


Figure 25 – Proportion who have been discharged from hospital at least once in the preceding 3 years – by age and by gender



PART B

Outcomes analysis

4 Exploratory modelling

In this section, we describe the results of exploratory modelling, which attempts to identify which factors explain variation in outcomes for the AS and KO populations and, specifically, the extent to which housing status (AS or KO) explains variation.

Three separate models were developed initially, to predict taxable income, conviction rates, and mental health service use over a three-year observation period to 31 December 2022. The modelling cannot be used to estimate causal effects i.e. even if a factor explains a high proportion of variation in an outcome, it does not necessarily mean one causes the other.

Key points

- The model used in this research suggests that the variation in taxable income, conviction rates, and mental health service usage can be primarily explained by the equivalent variables of the preceding three years (to 31 December 2019). For example, the taxable income earned in the three-year period to 31 December 2019 can explain almost all of the income variation in the three-year period to 31 December 2022.
- The model used in this research suggests that housing status (AS or KO) explains very little variation in taxable income, conviction rates, and mental health service use. Or expressed in a different way, whether a person receives AS for a full year or is in a KO public house for a full year appears to explain very little variation in income, conviction rates and mental health service use over the subsequent three-year period.
- More importantly, the research findings suggest that there are differences between the KO and AS populations, and they experience materially different outcomes. These differences can be explained by the underlying differences in the populations and their prior outcomes.

Due to the limited scope imposed by a short timeframe, this research did not disaggregate sub-cohorts with different characteristics in the KO and AS populations and explored a limited set of outcome variables. Therefore, variation in outcomes experienced by different sub-cohorts of the KO and AS populations are not presented in this research.

Note that this research does not rule out the possibility that there is a link between housing status and these outcomes, nor does it rule out the possibility that there is a link between housing status and other outcomes not considered in this research, e.g., specific health conditions, or housing quality.

Nevertheless, this research has identified interesting indications that warrant further investigation that will lead to insights on better aligning housing support and people. We recommend that the further work be carried out.

Note 2019 research by the Social Wellbeing Agency exploring the impact of public housing on people's wellbeing[4]. The research identified two key findings:

- Housing conditions generally improve for people placed in public housing
- Life satisfaction improves for people placed in public housing.

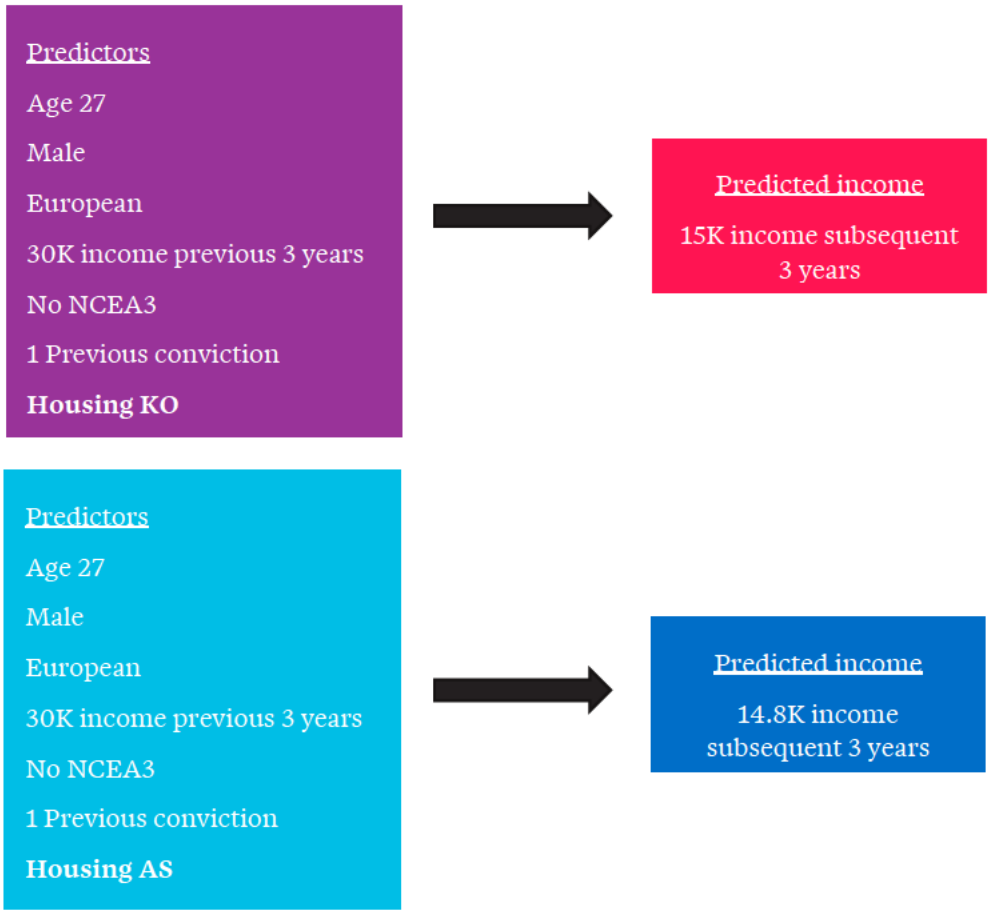
In this section, we present results of exploratory modelling to look at how individuals in different housing populations experience different outcomes over the observation period (2020-2022). For ease of comparison, and to focus on the groups with largest sample size, we have restricted our analysis in this section to comparing people in the KO and AS populations.

[4] Social Wellbeing Agency (2019) *Measuring the impact of social housing placement on wellbeing*

4.1 Outline of modelling

To understand how previous government service use and housing status help predict future outcomes, we have built models that predict a range of outcomes over the observation period. These predictions are based on all the government service use described in Table 2 and demographic factors. These predictions are also based on the housing status of an individual during 2019 (either KO or AS) and duration in that state over the three years to 31 December 2019. This allows us to compare predictions for people who are similar, based on the data available to us, except for their housing status – see Figure 26.

Figure 26 – Indicative example



For this research, the purpose of the modelling isn't explicitly to produce predictions of outcomes. Rather, the purpose is to disaggregate the relationships between factors that explain variation in outcomes, and specifically understand the extent to which housing status explains the variation. This can't be achieved from the descriptive analysis in Part A, because the populations vary with respect to other factors.

Further detail on the modelling process is contained in Appendix A.

4.2 Limitations

Before discussing the results of the exploratory modelling, limitations should be noted, as they impact the conclusions that can be drawn from the research.

- **Differences in population** – When we compare the KO and AS populations, we only have access to information about their demographics and government service use. For example, there is no information about untreated mental health. There are almost certainly differences between the two populations which are not visible in our data. Differences in prediction may be at least partly explained by differences in such unobserved factors, rather than a difference in housing status.
- **Correlation vs. causality** – The analysis we have undertaken cannot be used to infer causal relationships. Rather, the exploratory modelling we have performed considers correlations between variables.
- **Population definition** – Housing status for the purpose of defining the populations has been defined based on being in a housing state for the full year to 31 December 2019. The relationships between housing status and outcomes, as identified in the exploratory modelling, is limited to this definition of housing status. Rather than any broader definition of being in a Kāinga Ora public house or in receipt of Accommodation Supplement.
- **Time constraints** – Concessions in research design needed to be made to fit in with tight timescales. These include:
  - Limiting our research population to people aged 16 and above.
  - Limiting our research to considerations of individuals i.e. not households.
- **Timeframes** – The observation period used for this research covers the pandemic period. It is plausible that correlations between variables may be materially different during this period.
- **Data limitations** – There are known limitations to several of the IDI datasets, and many (such as education and Oranga Tamariki) are only available for certain age groups or are unavailable for immigrants to New Zealand. We have indicated possible data restrictions in Table 2.

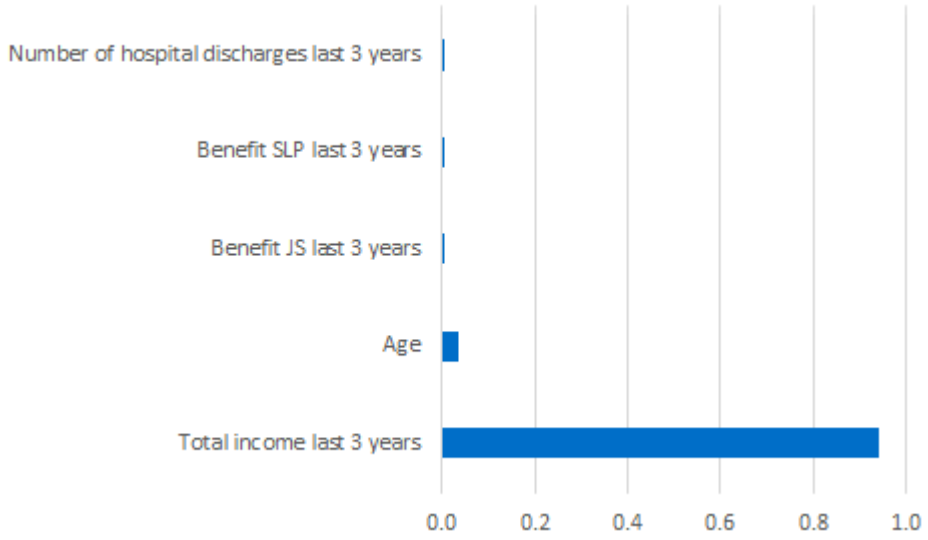
4.3 Taxable income

We built a model of the taxable income in the three years after 31 December 2019 to try and understand relationships to previous government service use and demographic factors. Note that the definition of taxable income in this analysis does not include benefit payments.

The first output from modelling is Figure 27 which shows the top 5 most important variables for predicting taxable income in the observation period, ranked by how much variation in taxable income they explain. In this scale, a value of 1 means that the variable explains all the variation in predicted future taxable income. Figure 28 shows that predicted future taxable income is almost entirely predicted by past taxable income, and other factors don't play a large role.

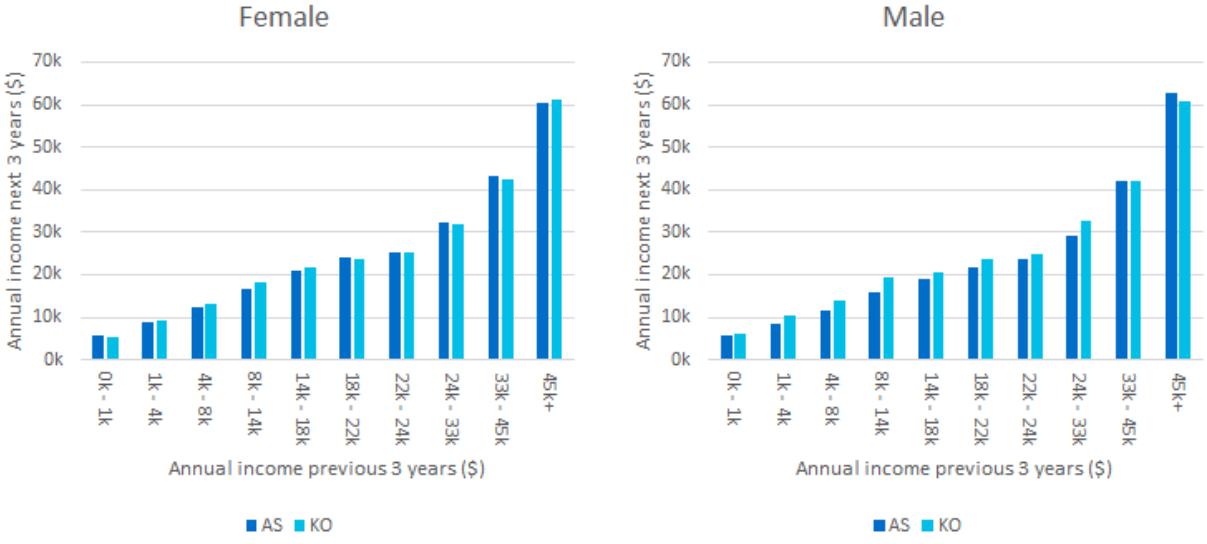
While this may seem like a very intuitive result, it is nonetheless an important one. It tells us that very little variation in taxable income is explained by factors other than past taxable income for the AS and KO populations. Or expressed differently, once we control for taxable income in the past three years there is relatively little variation in taxable income over the next three years.

Figure 27 – Top 5 variable importance for future taxable income model



Further confirmation of the strong relationship between past taxable income and future taxable income can be seen in Figure 28, which shows average total income in the three years after 31 December 2019, restricted to individuals who had any income. The data is shown by average total income in the three years to 31 December 2019. In these plots, average income in the three years after 31 December 2019 is closely aligned to average income in the three years to 31 December 2019.

Figure 28 – Average annual taxable income during the three years after 31 December 2019 (for individuals with > 0 income) – by average annual taxable income in the three years to 31 December 2019[5]



To analyse specifically the effect of housing status, when all other factors are kept the same, we look at average predicted taxable income for people with different housing status. Table 3 shows that altering housing status has little effect on predicting future taxable outcome when all other factors in the modelling are kept the same. Or, expressed differently, once we control for all other factors in the modelling (notably taxable income in the past three years), there is very little variation in taxable income by housing status.

[5] Each plotted previous income band is approximate to the nearest thousand. 10% of the entire population (males and females) are in each band. This equates to roughly 10% of males and 10% of females in each band.

Table 3 – Partial dependence of income on housing status

Housing status	Predicted taxable income (3 years)
AS	52,744
KO	53,084

4.4 Convictions

In this section, we model how conviction rates depend on characteristics of the population before the observation window. This will determine if the differences between conviction rates for AS and KO populations displayed in Figure 29 are explained by housing status or underlying differences in the populations.

Figure 29 – Rates of conviction over observation period

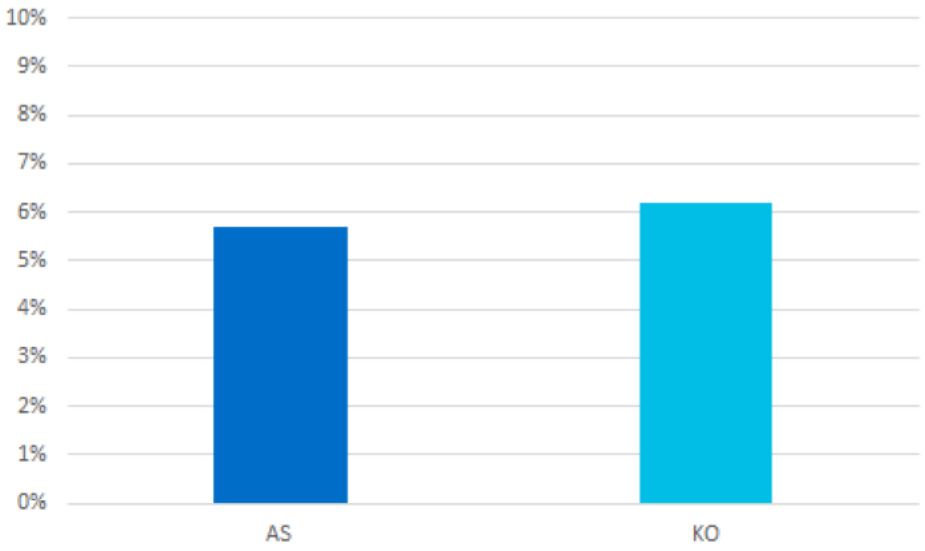


Figure 30 shows that previous convictions are the most important factor in predicting future convictions. Māori ethnicity is correlated with other factors that influence likelihood of conviction – this is not a causal factor in itself.

Figure 30 – Top 5 variable importance for convictions models

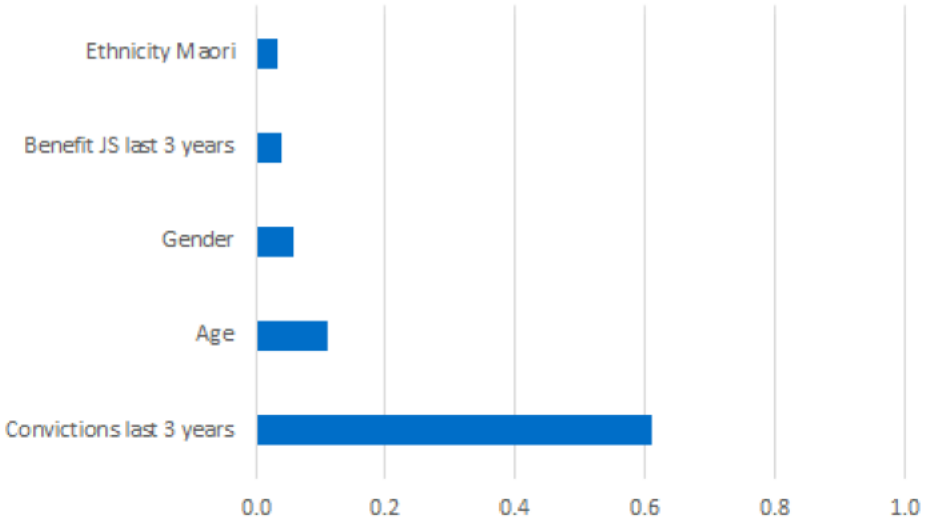
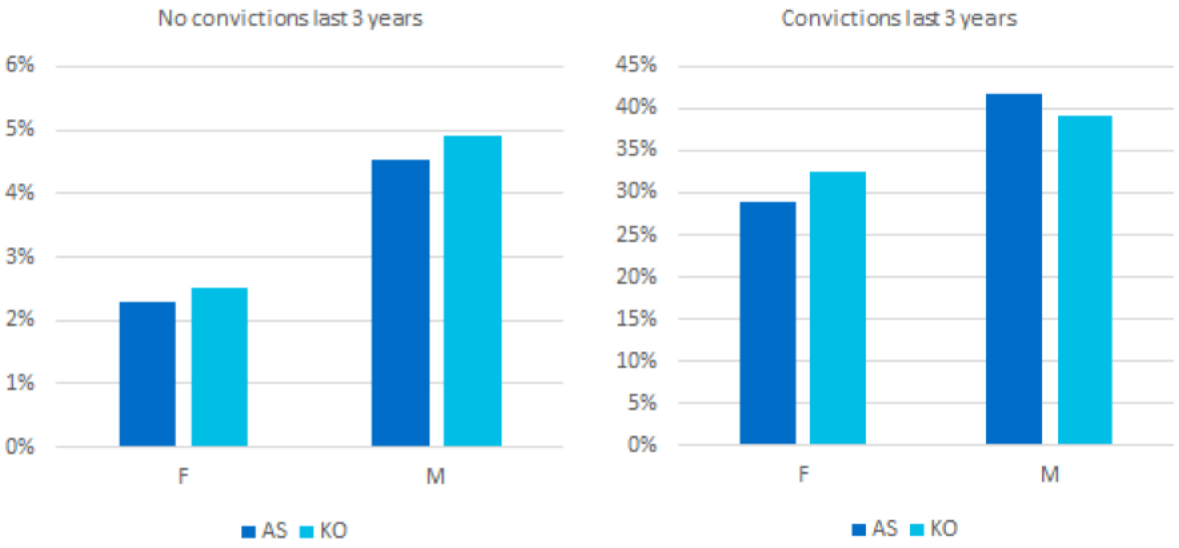


Figure 31 shows how rates of convictions over the observation period are strongly related to gender and whether someone has previous convictions, consistent with the results from the modelling.

Figure 31 – Rates of convictions over observation period (subsets)



To analyse the effect of housing status, when all other factors are kept the same, we look at average proportion of people with future convictions for people who are in KO, compared to people in AS. Table 4 shows that altering housing status has little effect on predicting future convictions, when all other factors are kept the same. Or expressed differently, once we control for all other factors in the modelling, there is very little variation in predicted conviction rates by housing status.

Table 4 – Partial dependence of future convictions on housing status

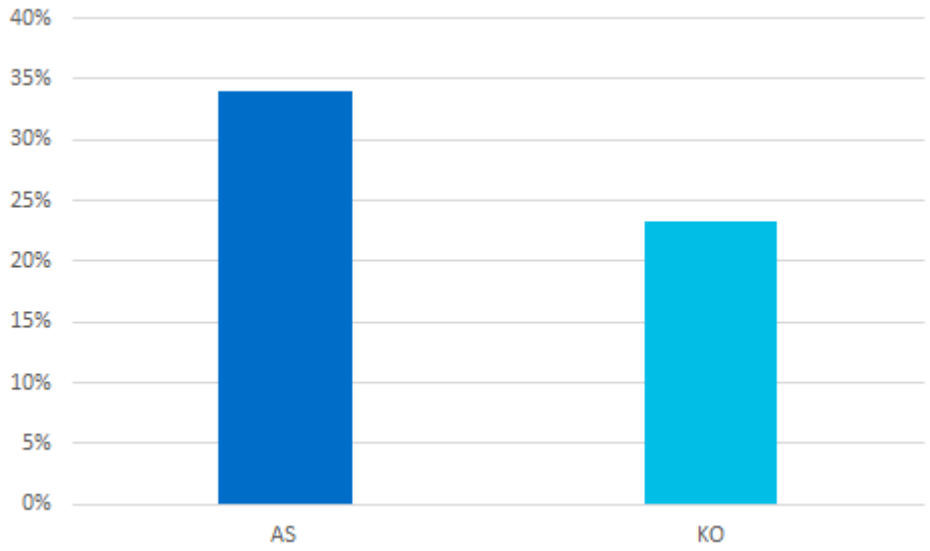
Housing status	Predicted rate of future convictions (3 years) – all other factors constant
AS	5.8%
KO	5.9%

4.5 Mental health and addiction service use

In this section, we look at the proportion of people who access any mental health/addiction services in the three years of the observation window. This incorporates pharmaceutical use, acute mental health and addiction services (inpatient and outpatient) and mental health and addiction related hospitalisations.

Figure 32 shows that there is a clear difference in mental health/addiction service use in the observation period – The AS population are more likely to use mental health services.

Figure 32 – Proportion with any mental health/addiction service use over the observation period



As in Section 4.3 and Section 4.4 we model how rates of mental health/addiction service use in the observation window depend on characteristics of the population before the observation window. This will determine if the differences between rates of mental health/addiction service use for AS and KO populations seen in Figure 32 is explained by housing status or underlying differences in the populations.

Figure 33 shows that previous mental health/addiction service use is the most important factor in predicting future mental health/addiction service use.

Figure 33 – Top 5 variable importance for mental health outcome model

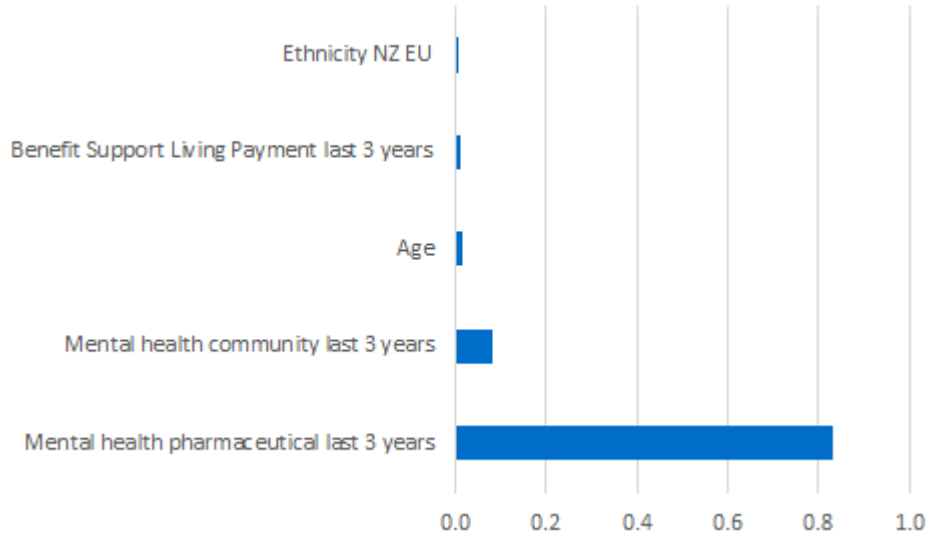
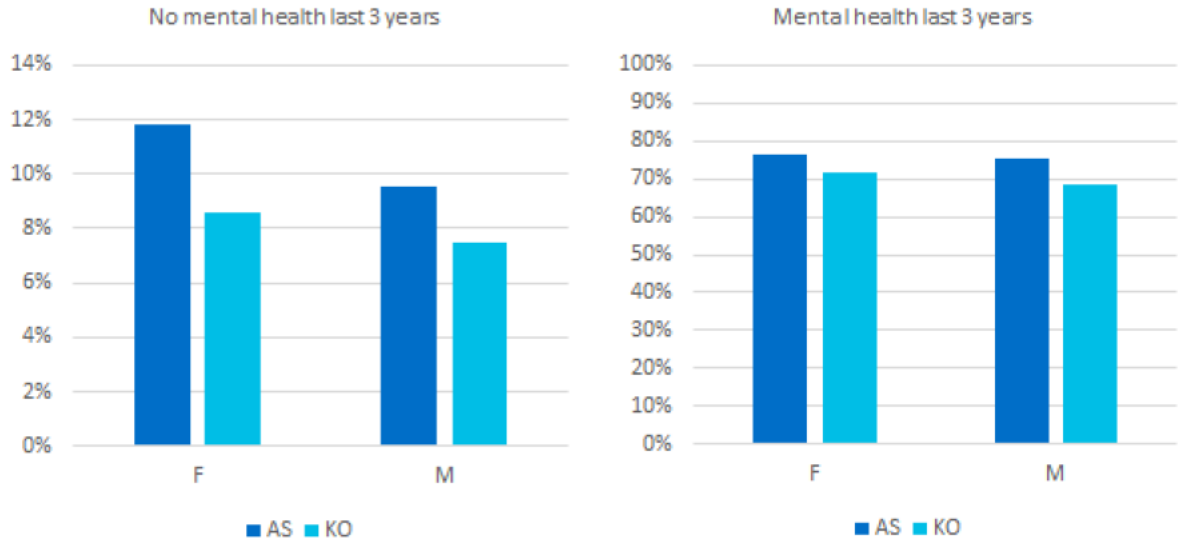


Figure 34 confirms that differing rates of mental health/addiction service use in the subsequent 3 years of the observation window are clearly related to mental health/addiction service use in the 3 years before the observation window.

Figure 34 – Proportion with mental health/addiction service use over the observation window



To analyse the effect of housing status, when all other factors are kept the same, we look at predicted average proportion of people with mental health service use, comparing people who are in KO with people on AS. Table 5 shows that altering housing status has little effect on predicting future mental health service use, when all other factors are kept the same. Or expressed differently, once we control for all other factors in the modelling, there is very little variation in mental health/addiction service use by housing status.

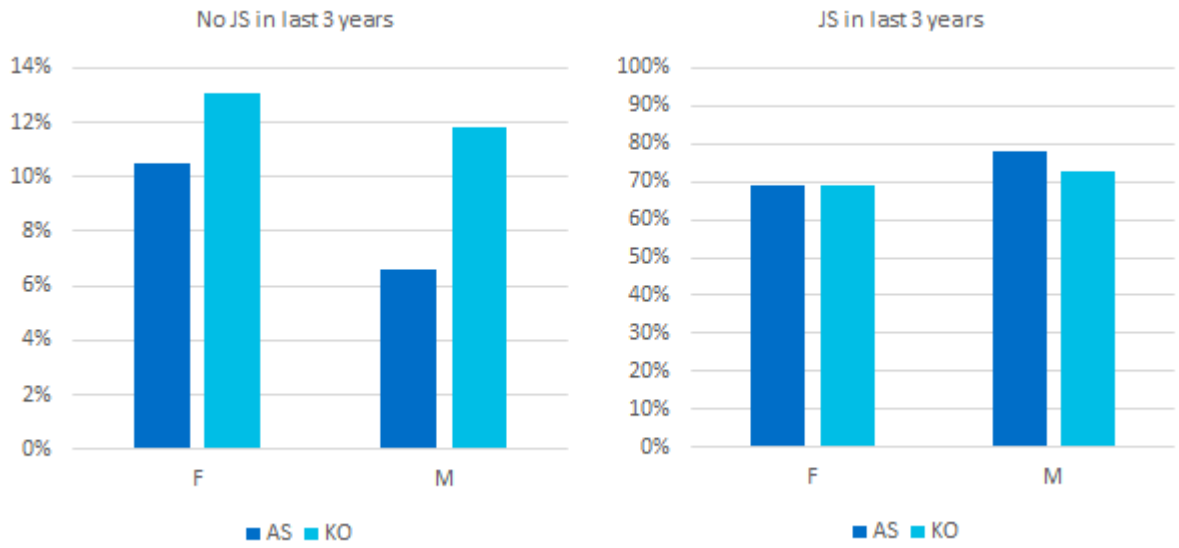
Table 5 – Partial dependence of mental health service use on housing status

Housing status	Predicted rate of future mental health service use (3 years) – all other factors constant
AS	31.0%
KO	30.6%

4.6 Further descriptive statistics

We also extracted descriptive statistics on jobseeker benefit status over the observation period, which is displayed in Figure 35.

Figure 35 – Proportion receiving JobSeeker Support at any point in the three years to 31 December 2022 (subsets)



Appendix A – Research approach

A.1 Population definitions

All data for this analysis is from the 202310 IDI refresh.

Table 6 – Population definitions

	KO	CHP	AS
Definition	All individuals with KO spell that starts before 1 January 2019 and ends after 31 December 2019. Spells that end within 2 days of the start of another spell are concatenated.	All individuals with CHP spell that starts before 1 January 2019 and ends after 31 December 2019. Spells that end within 2 days of the start of another spell are concatenated.	All individuals with a spell related to support codes 471, 470, 472, 474, 473 or 832 where the spell starts before 1 January 2019 and ends after 31 December 2019. Nested and overlapping spells are concatenated.
IDI Data sources	[hnz_clean].[tenancy_snapshot]	[hnz_clean].[tenancy_snapshot]	[msd_clean].[msd_second_tier_expenditure]
	[hnz_clean].[tenancy_household_snapshot]	[hnz_clean].[tenancy_household_snapshot]	
Total	85,836	7,374	220,155

In principle, there should be no overlap between the three categories in Table 6. Due to a small number of data inaccuracies, there are some individuals who are in different states at the same times. In our modelling dataset, overlaps have been eliminated – if an individual is both in CHP and KO state at the same time, they are set be in the KO state. If they are in the AS and KO/CHP state at the same time, they are set to be in the KO/CHP state.

Table 7 highlights the number of data overlaps.

Table 7 – Data overlaps

One person in multiple households	One person in multiple households (different categories)	Overlaps between KO and AS	Overlaps between CHP and AS
345	213	2,526	144

A.2 Exploratory modelling

The exploratory modelling in this analysis uses a type of machine learning called Gradient Boosting Machines (GBMs). Roughly speaking, GBM models are constructed by minimising an 'objective function' which measures the error between the actual target values and the predicted values of the model. This class of models have the following features which make them appropriate for this analysis:

- Ability to incorporate flexible nonlinear relationships and complex interactions
- Robust to the inclusion of multiple correlated features
- Simple to fit multiple models with large numbers of features.

Table 8 lists the variables used in the modelling.

Table 8 – Variables used in GBM models

Variable name	Transformations
Age	NA
Convictions (imprisoned) past 3 year	NA
Convictions past 3 years	Transformed to indicator
Days in KO in previous 3 years	NA
Days on benefit JS past 3 year	NA
Days on benefit SLP past 3 years	NA
Days on benefit SPS past 3 years	NA
Ethnicity	One-hot encoded
Gender	One-hot encoded
Highest secondary qualification	Restricted to Age 19 to 28 only
Mental health community service use	Transformed to indicator
Mental health pharmaceutical service use	Transformed to indicator
Number of hospital discharges past 3 years	NA
Observation period status (KO or AS)	Transformed to indicator
Oranga Tamariki intervention	Transformed to indicator for ages 19 and 20
Region	One-hot encoded
Total income past 3 years (not including benefits)	NA

Three GBM models were constructed for our analysis, as outlined in Table 9. In itself, the RMSE doesn't give an intuitive measure of how accurate the model is, we used Actual vs. Expected plots to verify the accuracy of the three models.

Table 9 – Modelling details

Model	Target type	Objective function	Optimised value of objective function
Total income	Continuous	Root mean squared error (RMSE)	36,529
Convictions	Binary	Log loss	0.161
Mental health	Binary	Log loss	0.367

Wellington
Level 6
22 The Terrace
Wellington 6011

Melbourne
Level 27
459 Collins Street
Melbourne VIC 3000

Sydney
Level 22
45 Clarence Street
Sydney NSW 2000

www.taylorfry.co.nz

TAYLOR FRY