

Care to Custody

Incarceration Rates

August 2022



Abuse in Care
Royal Commission of Inquiry



Contents

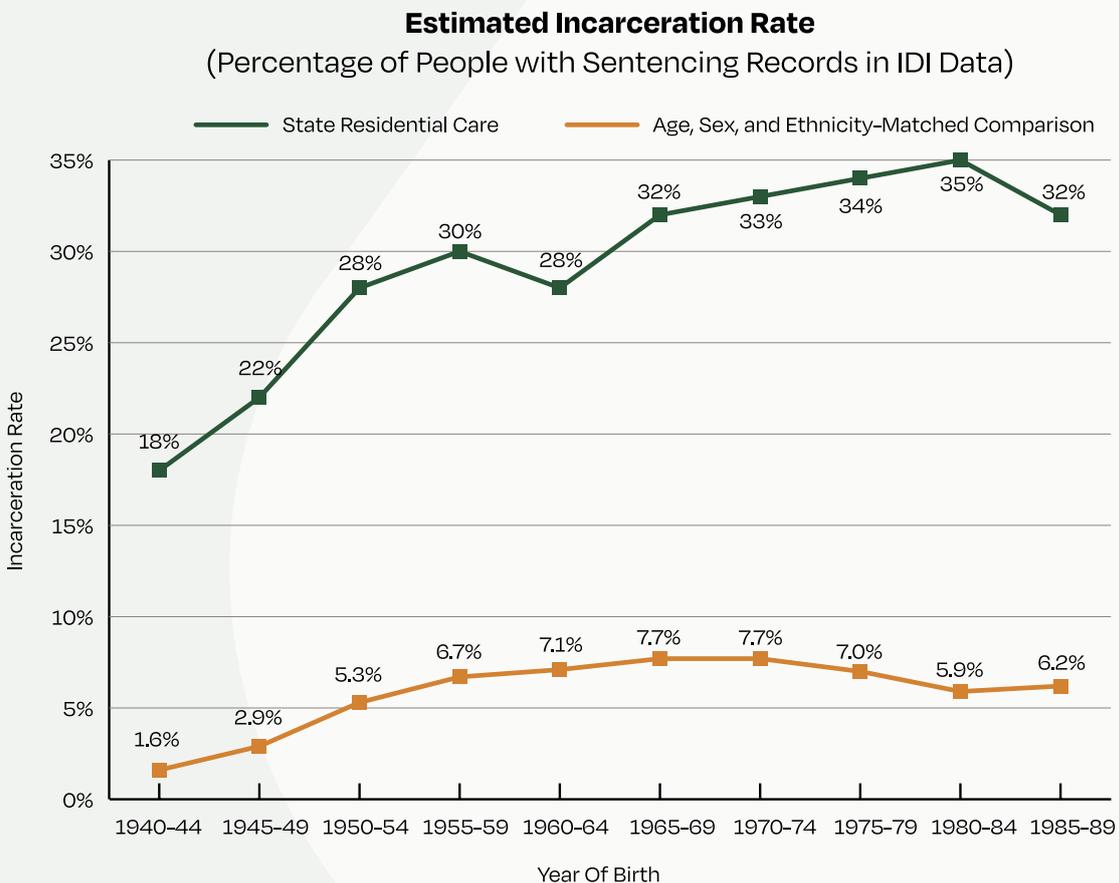
Executive Summary.....	4
Overview of Research Method	5
Key Limitations and Caveats.....	6
Results.....	7
Basic Demographics of Children in State Residential Care.....	7
Overall Incarceration Rates.....	8
Māori versus Non-Māori Incarceration Rates.....	9
Matched Cohort Construction and Interpretation.....	10
Summary of Key Findings	12
APPENDIX: FURTHER RESEARCH DETAILS	14
Research Question	14
Population of Interest	14
The State Residential Care Dataset	14
Definition of Residences	15
Data Cleaning	15
MoH Format of Dataset	15
Identity Matching	17
Measurement Methods.....	18
Matched Cohort.....	18
State Residential Care Dataset: Treatment of Unknowns	19
Key Caveats.....	20
Potential Issues with Data Quality and Matching.....	20
Data Availability	21
Representativeness.....	21
Limitations of Matched Cohort.....	22
Interpreting Comparisons	22
Effect on Main Conclusion.....	22
Data Table.....	23
Issues Relating to Māori Data Governance and Sovereignty	24
Disclaimer	24
Acknowledgements.....	24

Executive Summary

As part of the Royal Commission's inquiries, we wanted better insight into the link between children and young people who were placed in State residential care between 1950 and 1999 then going on to serve a custodial sentence through the criminal justice system.

A list of more than 30,000 names and birthdates of people who were in care during this time was provided by Oranga Tamariki. A data match, including the Ministry of Health, Department of Corrections, and Stats NZ, matched almost three quarters of those people to known identities in the Integrated Data Infrastructure (IDI), a combined database of government records managed by Stats NZ. For each person, a search was done to find any custodial sentencing records within the IDI.

The research found that one in five and, sometimes, as many as one in three of those children and young people who had been in State residential care, went on to serve a criminal custodial sentence later in life. This is a much higher rate than those who had not been in State care.¹ The graph below shows incarceration rates were far higher than for a comparison cohort matched in age, sex, and ethnicity.



1 Limitations and caveats around these analyses are described on pages 6 and 20

Overview Of The Research Method

Oranga Tamariki provided as comprehensive a list as possible of people who had been in State residential care during the Inquiry period of 1950–1999. Using the Integrated Data Infrastructure (IDI), a combined database of government records managed by Stats NZ, the aim was to find out how many of the people in the State care list had records of having received a custodial sentence.

Briefly, the steps taken were:

- › **Data cleaning.** The data was cleaned manually and duplicates were identified.
- › **Data matching.** The full list was sent to the Ministry of Health and Department of Corrections to see if they could match them with known identities in their data.
- › **Integration into IDI.** Stats NZ used the unique identifiers provided by the Ministry of Health and Department of Corrections, and used their own matching process, to match people to known identities in the IDI.
- › **Querying sentencing data.** For those people that could be matched to known identities, IDI data was searched to find any custodial sentencing records.
- › **Calculating incarceration rate.** For different age groups, and for Māori and non-Māori separately, the rates of incarceration were calculated.
- › **Comparing to a matched cohort.** A cohort of people in the IDI whose age, sex/gender, and ethnicity matched those who had been in State residential care was created and their corresponding incarceration rates calculated.²

2 Further methodological detail is provided in the Appendix. There are important limitations and caveats around the data to consider.

Key Limitations and Caveats

Key limitations and caveats around the data include:

- › There were significant quality issues with the names and dates of birth in the supplied list of those who had been in State residential care.
- › Other identifying information was absent in the supplied list.
- › The list may contain only a fraction of those who were actually in State care.
- › Matching to known identities will produce some false-positive matches.
- › Only around three out of four distinct individuals could be matched to a known identity. This is likely to bias the matched individuals towards those known to Department of Corrections, so it was necessary to calculate results in a way that accounted for this, resulting in a conservative estimate of incarceration rates for those who were in State residential care.
- › Data in the IDI is, itself, subject to false matches between linked datasets.
- › Data in the IDI may be incomplete or not available at all, particularly for older data.
- › If the incarceration rate differs between those who were in State residential care and those who were not, it does not mean that being in State care caused that difference. Nevertheless, as the results below show, the differences were clear and the general pattern was consistent across year and ethnic groupings.
- › Highlighting the experiences of disabled survivors is an important part of this Inquiry's work. Unfortunately, it was not possible to analyse disability in this research. Over the time period examined, disability information was either not available or tended to be recorded poorly or ambiguously.
- › Reporting on other ethnicities has not been included due to insufficient numbers. We intend to present data for each ethnicity as part of the technical report, currently being finalised.

Results

Basic Demographics of Children in State Residential Care

Almost two thirds of the children who had been in State residential care and who could be identified in the IDI were male. Figure 1 shows the ethnicities of those children according to core IDI data. Almost half were identified as Māori. This is a substantial over-representation of Māori compared to the general population: thirteen per cent of the population identified as Māori in the 2001 census, an increase from six per cent in the 1951 census. (Note, however, that there are considerable challenges and inconsistencies with how Māori are counted.)

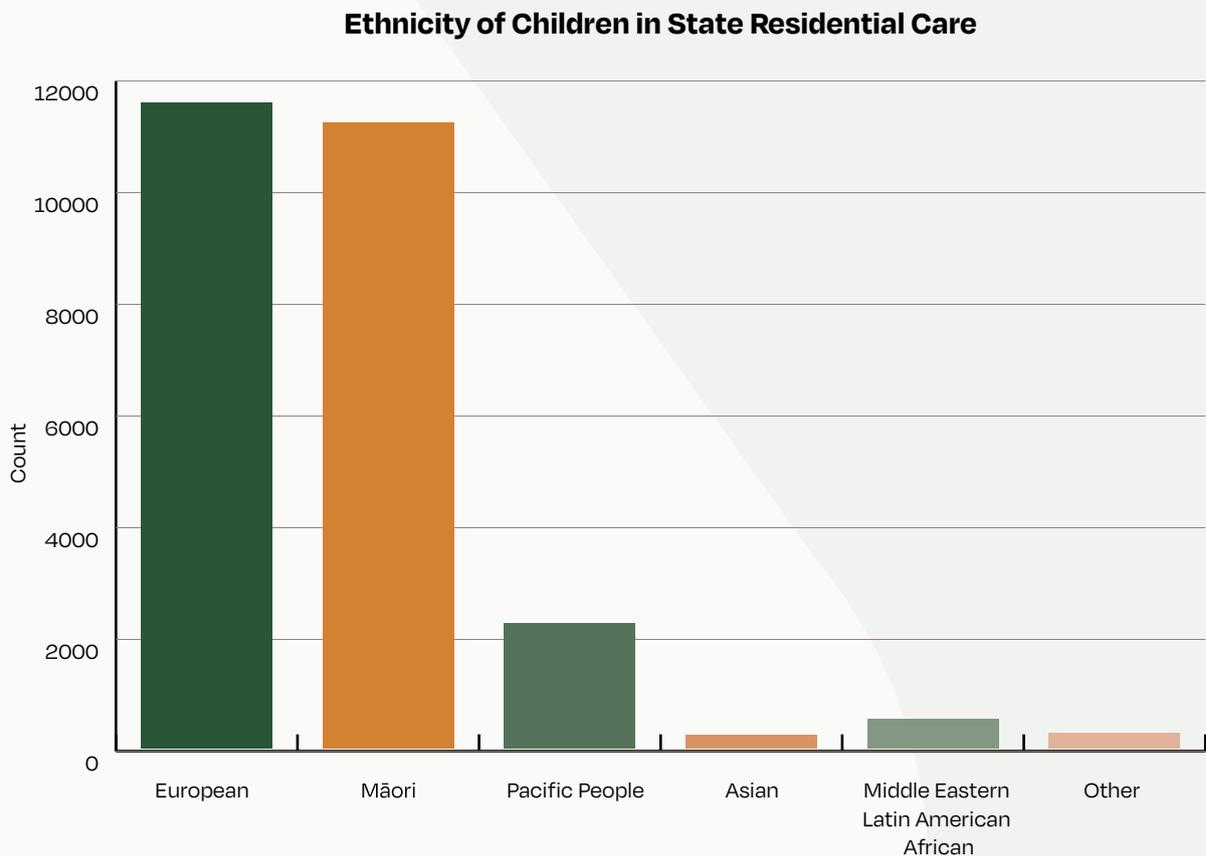


Figure 1. Ethnicity of children who were placed in State residential care and subsequently identified in the IDI. Ethnicity is based on standard categorisations in core IDI records. A single person may identify with multiple ethnicities.

Overall Incarceration Rates

The list of people in State care was grouped into five-year cohorts and the proportion of people who had a corresponding record in the IDI of receiving a custodial sentence was calculated. The same was done for a comparison cohort matched by age, sex/gender, and ethnicity. Figure 2 shows that anywhere from around one in five (20%) to more than one in three (33%) of people in State care went on to serve a custodial sentence. For the matched cohort, the incarceration rate was no more than around one in thirteen (7.7%). People who had been in State residential care were far more likely to have received a custodial sentence than the matched cohort.

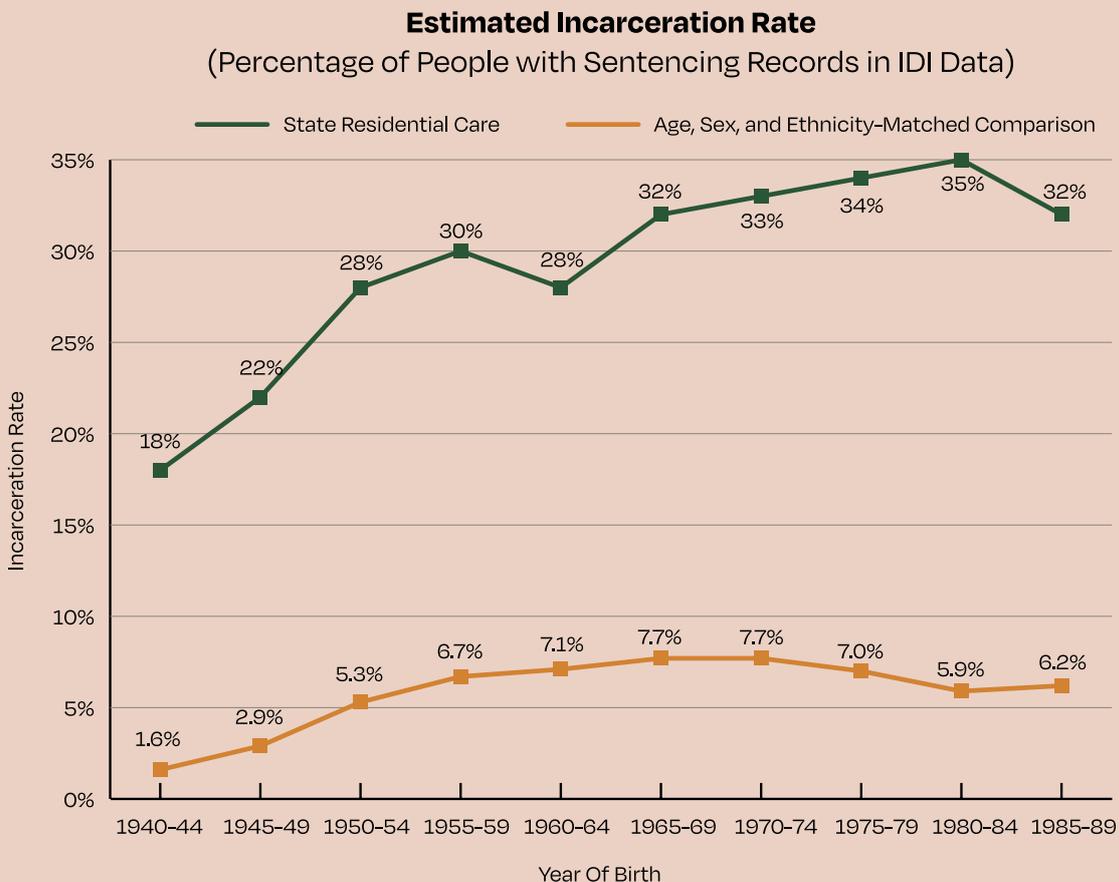


Figure 2. Estimated incarceration rates for individuals who had been in State residential care (upper line), and a comparison cohort matched by age, sex/gender, and ethnicity (lower line). Estimated incarceration rates are based on Department of Corrections and Ministry of Justice data within Stats NZ's Integrated Data Infrastructure. See below for further details about how the incarceration rates were calculated. Incarceration rates shown for those who had been in State residential care are conservative and likely to be lower than their true value.

Māori versus Non-Māori Incarceration Rates

We looked at data relating to Māori versus non-Māori in State care separately. Previous research has established that Māori are disproportionately represented in State care. The make-up of the Māori population has also changed considerably over the period examined. Figure 3 shows that the proportion of Māori who had been in State residential care and subsequently received a custodial sentence was much higher than for corresponding non-Māori.

State Care Māori vs Non-Māori Estimated Incarceration Rate
(Percentage of Non-Māori with Sentencing Records in IDI Data)

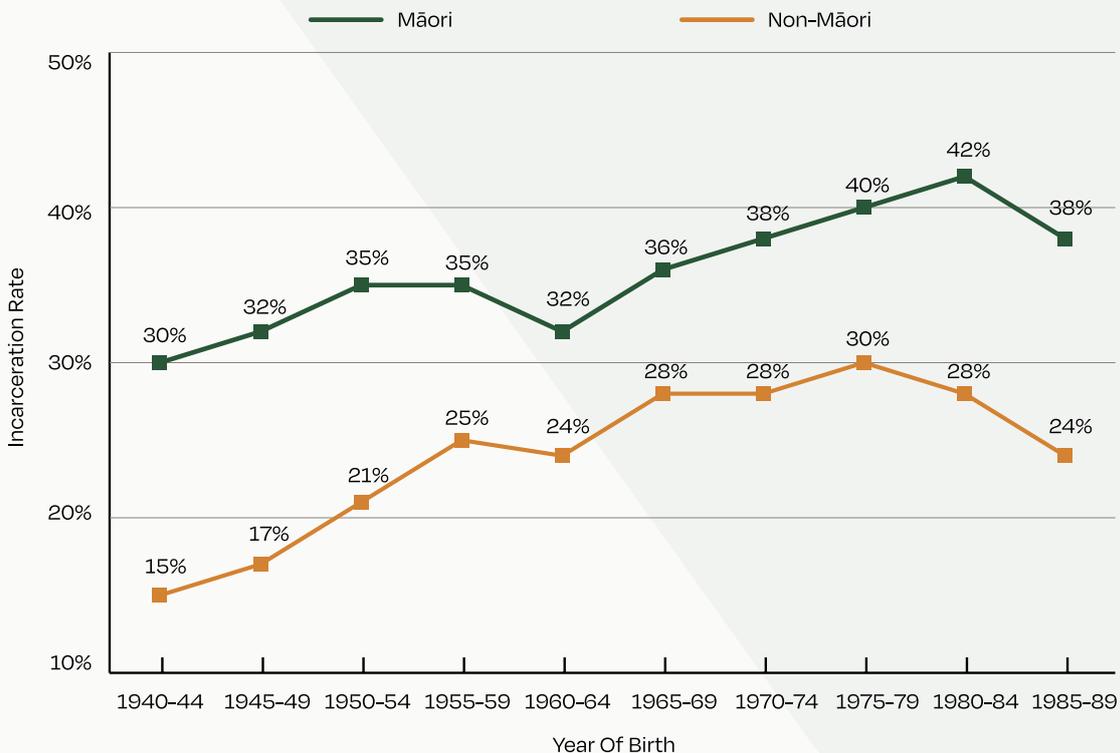


Figure 3. Estimated incarceration rates for Māori versus non-Māori, all who had been placed in State residential care as children. Estimated incarceration rates are based on Department of Corrections and Ministry of Justice data within the IDI. Note that earlier prison sentencing records are more likely to be missing and that incarceration rates shown are likely to underestimate their true value.

Matched Cohort Construction and Interpretation

To create the matched cohort, the process started with every person in the IDI (several million people) and excluded those who had been identified as being in State residential care. The process then found those whose age, sex/gender, and ethnicity matched those who had been in State residential care. Finally, the process weighted them to ensure they were in the same proportions as those identified as having been in State residential care.

To understand what a comparison to this cohort means – and what it does not mean – it is essential to grasp that there are additional factors that might have been used to create a matched cohort. Poverty, location and educational attainment are possible examples amongst many. The more imprisonment-related factors that are used to create the matched cohort, the closer its incarceration rate is likely to be to those who were in State residential care. Figure 4 illustrates this through the different factors that the process used.

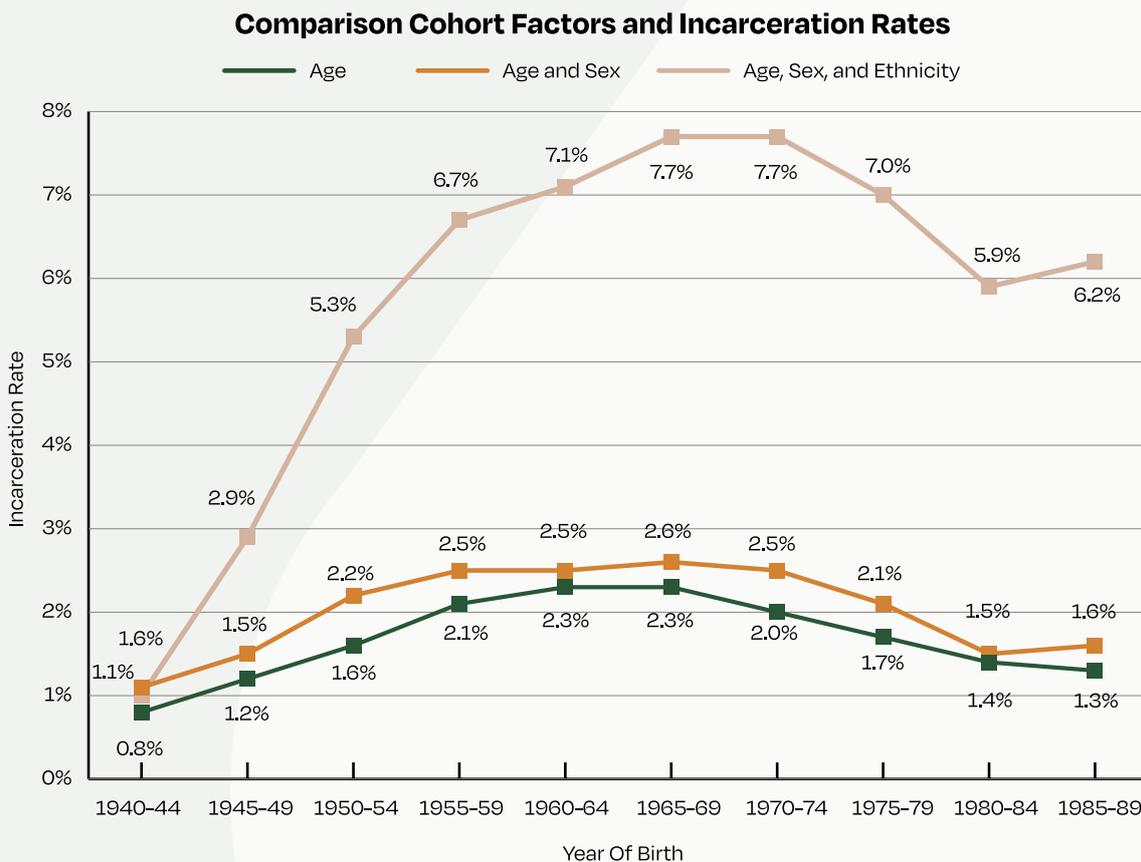


Figure 4. Demonstration of how the comparison cohort's incarceration rate changed when more factors were used to mirror the make-up of the people placed in State residential care as children. The incarceration rate for the comparison cohort was highest when more demographic factors were considered. Note that earlier prison sentencing records are more likely to be missing.

Figure 4 shows that a comparison cohort matched to the State residential care group by age only had the lowest incarceration rate. When it was matched by both age and sex/gender, the incarceration rate was higher. Finally, when it was also matched by ethnicity, the incarceration rate was substantially higher. It is not possible to know all of the factors linked to incarceration. More importantly, it is not known if adding all such factors would result in an incarceration rate as high as those who had been placed in State residential care as children, or if there would still be a gap. Consequently, it is not possible to tell if being placed in State residential care is a cause of imprisonment. It is only possible to tell from this data that there is a clear association. Deeper contextual research is needed to understand why.



Summary of Key Findings

All results are shown in Table 2. This includes likelihood ratios, indicating how much more likely the people in State care were to receive a custodial sentence than their matched comparison cohort. These likelihood ratios indicate that, over the different five-year groups.

- › People who had been in State residential care were usually about five to nine times more likely to be incarcerated than people who had not.
- › Māori who had been in State care were usually around four to seven times more likely to receive a custodial sentence than their matched cohort.
- › Non-Māori who had been in State care tended to be around 15 to 24 times more likely to receive a custodial sentence than their matched cohort.
- › The differences between Māori and non-Māori described here result mostly from a higher incarceration rate for Māori overall, regardless of whether they had been in State care or not.



Appendix: Further Research Details

Research Question

What proportion of children and young people who were placed in State residential care between 1950 and 1999 went on to receive a judicial custodial sentence?

Population of Interest

The population of interest was approximately 31,000 distinct individuals, identified by name and date of birth, who were placed into State residential care facilities as children or young people between 1950 and 1999. Further details about this data set appear below.

The State Residential Care Dataset

The State residential care dataset contained names and dates of birth of children, young people who were placed into State residential facilities, primarily between 1950-1999. The dataset was obtained in response to the Royal Commission's Notices to Produce No. 36 dated 17 February 2020. Schedule B of the Notice sought a complete list of all individuals admitted to Residences (as defined below) from 1950-1999 for most Residences, but also outside of this period for those placed at Whakapakari Youth Trust, Eastland Youth Rescue and Moerangi Treks.

Oranga Tamariki compiled the list based largely on handwritten admissions registers. In producing the data, Oranga Tamariki noted that a complete list does not exist, that admissions registers may not always be filled in accurately, that some Residences did not hold admissions registers and that some records may have been destroyed. However, the resulting list was "as close as possible to a complete list as is reasonably practical". The complete list of fields in the dataset is:

- > row number [1-36,046] – actual row count was 35,228
- > first name
- > middle / alternative names
- > surname
- > date of birth (dd/mm/yyyy) or age at admission.

Other identifiers were absent from this dataset. As such, there was no data about ethnicity or sex/gender, nor any iwi affiliation. Data about location and date of admission was also missing.

When the data was reviewed, it was found that names, especially first name and surname, were present for almost all records. Names were, however, subject to spelling errors for a number of reasons. One was that most data was sourced from handwritten records. Dates of birth appeared to be less reliable, with over 10% appearing to be missing or unusable. So, although there were examples of missing data in different fields, around 90% of rows appeared to be appropriately intact. However, the "Middle / alternative names" field was ambiguous, leading to different possible combinations of individuals' names, and that alternative name information was sometimes supplied in different fields.

Definition of Residences

Residences were defined as:

- › Children's homes and institutions providing care and protection residential facilities for children and young people, including social welfare and family homes;
- › Institutions that provided remand or secure facilities as well as care and protection residential and/or training facilities for children and young people; and
- › Residential programmes or facilities provided by Whakapakari Youth Trust, Eastland Youth Rescue, and Moerangi Treks, and explicitly excluding those provided by faith-based institutions.

Data Cleaning

To further prepare the data, cases of potential name or date ambiguity were identified and then cleaned manually. This included identifying potential alternative names and birthdate ranges. Although near-duplicates were identified, they were retained in the dataset described below, to assist with the matching process.

The dataset was then converted into a format specified by the Ministry of Health to support the names and birthdates being matched to known identities. This file was then supplied via secure means to the Ministry of Health and Department of Corrections for matching purposes.

Ministry of Health Format of Dataset

The State residential care dataset was then converted to a pipe-separated value text file consistent with the preferred format for the Ministry of Health for carrying out matching of names to known identities. The manual cleaning performed included separating names into explicit multiple alternative first name, middle name, and surname fields, and into likely date of birth ranges. The resulting list of fields, with additional description, is shown in Table 1, below.

Table 1. Fields included in data file submitted to the Ministry of Health and Department of Corrections for matching to known identities

Field Name	Description or Notes	
UniqueID	Positive, mostly sequential integer	
Old NHI	Blank in all cases	
DOB	Blank in some cases	
DOB Range Start	When a DOB was ambiguous, the start and end date of a likely range of dates was specified	
DOB Range End		
DOB Description	The original text provided as a description of the DOB. May be useful in a handful of ambiguous cases for manual matching only.	
Gender	Blank in all cases	
Surname	Alternate names were supplied when multiple names were given. Note that:	
First Name		
Middle Names		
Alternate Surname 1	<ul style="list-style-type: none"> > There may have been repeats within an individual if it was difficult to determine whether the name was a first name, middle name, or surname. > The index of an alternate name (e.g., "1") did not necessarily imply that it was associated with other alternate names of the same index. For example, "Alternate Surname 2" might be associated with "First Name", "Alternate First Name 1", and so forth. 	
Alternate First Name 1		
Alternate Middle Names 1		
Alternate Surname 2		
Alternate First Name 2		
Alternate Middle Names 2		
Alternate First Name 3		
Address line 1		Blank in all cases
Address line 2		
Address line 3		
Address line 4		
Email		
Phone		

Identity Matching

The Ministry of Health and Department of Corrections used a combination of automated and manual data matching techniques to match names and dates of birth within the State residential care dataset to known identities. The goal was for the Ministry of Health to augment the dataset with individuals' National Health Index (NHI) numbers, and for Department of Corrections to augment the dataset with Ministry of Justice's unique identifiers. This work was carried out following the Royal Commission's Notice to Produce No. 416, pursuant to sections 20 and 22 of the Inquiries Act 2013. As part of the identity matching exercise, and in accordance with the Notice to Produce, Department of Corrections and the Ministry of Health shared information with one another to improve their identity matching.

Once the Ministry of Health and Department of Corrections had completed their identity matching processes, they submitted the resulting augmented dataset to Stats NZ for integration into the IDI. Stats NZ matched NHI and Ministry of Justice unique identifiers to unique identifiers of individuals in the spine of the IDI through existing concordances.

Additionally, Stats NZ used their own fully-automated fast-match load process to match individuals to known identities in the IDI (June 2022 data refresh). The matches that were retained were the high-quality matches that were completed on the first pass, and that were an exact or near-exact match.

Through these different approaches of identity matching, approximately three quarters of distinct individuals within the State residential care dataset were matched to known identities in the IDI. In instances where multiple identifiers were supplied, preference was given to Ministry of Health identifiers because their data represented a large, general dataset, and the matching process they used was thorough and included a significant manual matching component. The next preference was given to Ministry of Justice identifiers supplied by Department of Corrections, whose process also involved some manual supervision. The lowest preference was given to identifiers provided solely by Stats NZ's fast-match load process because it was fully automated.

Measurement Methods

Matched Cohort

Briefly, the method used to generate a matched cohort was as follows:

- › Start with every person in the IDI (several million people).
- › Exclude people who were in the State residential care dataset (as determined by the identity matching process described above).
- › Retain everyone else who, based on core IDI details, had a year of birth, month of birth, "sex/gender" value, and "ethnicity1"- "ethnicity6" combination ("biodata combination") that matched the biodata combination of at least one person in the State residential care dataset. Usually, several hundred or several thousand people per biodata combination were found.
- › Obtain sentencing information for each person and calculate the incarceration rate for each biodata combination.
- › Weight each biodata combination – and its associated incarceration rate – by how often that biodata combination occurred in the State residential care dataset for each slice examined.
- › Take the weighted average of those incarceration rates applicable to the slice being examined.

This produced the incarceration rate for a group of people whose biodata profile matched that of the relevant slice of the residential State care dataset, and in the same proportions, but based on every relevant person in the IDI.

One exception to the method was that there were some biodata combinations (around 90 of about 8,000) in the State residential care dataset that did not have a matching combination in the IDI data. For these occurrences only, which affected only 90 people, the biodata matching criteria were relaxed to two ethnicities only instead of six: "ethnicity1" and "ethnicity2", corresponding to New Zealand European and Māori respectively. Year of birth, month of birth, and sex/gender criteria were unchanged in these instances.



State Residential Care Dataset: Treatment of Unknowns

After removing likely duplicates, only around three quarters of the people in the State residential care dataset were matched to identities in the IDI. For the purposes of measuring the incarceration rate, a conservative approach of assuming that none of those people had received a custodial sentence was taken. This conservative assumption may make the incarceration rate for the State residential care dataset appear lower than it really is. This assumption does not affect the incarceration rate of the matched cohort.

To execute this conservative assumption procedurally, the counts of distinct people in each five-year group of the State residential care dataset, whether identity-matched or not, were taken and those counts used as the denominators of the incarceration rates. The main limitation of this approach was that the year group of the 10% of people listed without a date of birth was not known. Similarly, when the data was split into Māori and non-Māori, the ethnicity of the roughly one quarter of distinct people in the dataset who were not matched to an identity was not known. It was therefore assumed that the year-group distribution for missing dates of birth was the same as that of everyone else in the State residential care dataset. A similar assumption was made around the distribution of Māori and non-Māori. These assumptions may not be correct. If they are not, though, it would not affect the conclusion that those in State care were far more likely to receive a custodial sentence. The general pattern was consistent across years and ethnic groups.



Key Caveats

Potential Issues with Data Quality and Matching

The original list of names and dates of birth was sourced from historical handwritten records. The nature of the data source lends itself to numerous potential issues, including:

- › lost or missing handwritten records
- › spelling errors in original handwritten text
- › missing or unusable dates of birth in around 10% of cases
- › inaccurate or guessed dates of birth
- › lack of other identifying information such as ethnicity, iwi, location, or date of admission
- › transcription errors from handwritten text to electronic format
- › data entered into incorrect fields (e.g., first name entered as middle name).

Other issues noted with the data are that:

- › spelling errors appeared to be more common for non-European names, particularly Māori names
- › the ambiguity of the "Middle / alternative names" field meant that it was difficult to differentiate between middle and alternative names, leading to several plausible name combinations for a given individual.

It should also be noted that the identity matching process is not perfect. Some of the matches to Ministry of Health, Department of Corrections, and IDI identities will be false positives. This can result from issues like those in the population of interest having the same name and date of birth as one or more people outside of it. Department of Corrections provided a rough estimate that this false positive rate could be as high as 2%. Stats NZ's data linking process within the IDI targets false positive rates below 2%, and rates are commonly in the range of 1-2%³. It is also possible that the matching process might be more likely to find matches for particular groups, such as if Māori names were less likely to be matched because they were more prone to spelling errors.

3 Stats NZ (2022). Integrated data infrastructure (IDI) refresh: linking report. Methods and design, June 2022 refresh. Wellington: Stats NZ

Data Availability

There are limitations on the data available in the IDI. Older data is generally less available and/or poorer quality than newer data. Stats NZ provides information about the available datasets in the IDI, but it should be noted that data older than the periods specified by Stats NZ are available in some cases, but tend to be increasingly questionable the older they are.

Representativeness

Regardless of any biases brought about by the matching process, it is possible that the State residential care dataset may not itself be representative of those who had been in State care during the period of interest. From examination of the data, the overall number of individuals in the dataset might only be a fraction (roughly 1/5 to 1/3) of the cohort size estimated in research conducted by Martin Jenkins and Elizabeth Stanley, although their scope may not have been the same as this research. It is unknown whether the factors that led to people being missed from the list might have caused the dataset to be biased or unrepresentative.

Limitations of Matched Cohort

A matched cohort was generated for comparison purposes based on age, sex/gender, and ethnicity. These were derived from core data fields within the IDI and managed by Stats NZ. It is critical to note that there are many attributes that the matched cohort was not based on. Examples include location, economic deprivation, and education. Had such factors been included when producing the matched cohort, the differences in incarceration rates between the matched cohort and the State residential care group may have been smaller. No matter how many attributes were included, though, it would not be possible to include all other factors that may be associated with imprisonment.

Interpreting Comparisons

It is important to interpret comparisons with care. First, consider the key limitations and caveats around the data noted above. Second, remember that conditions and the availability and quality of records changed considerably over the fifty years being examined. This consideration limits the conclusions that can be drawn comparing across years, and merging data between year groups is not recommended. Third, and importantly, it is not appropriate to conclude from this data alone that the differences in incarceration rates were caused by being in State residential care. The research was correlational in nature, and correlation by itself should not be inferred as causation. Last, as noted above, the cohort matching process did not account for other factors that might make incarceration more likely.

Effect on Main Conclusion

Although there are caveats and limitations to consider, the principal results are nevertheless clear. The incarceration rate of people who were in State residential care is high. It is much higher than that of a cohort matched in age, sex/gender, and ethnicity.

Data Table

Table 2. Data from the research study for Māori (left), non-Māori (middle), and overall (right). See legend below.

Year of Birth	Māori			Non-Māori			Overall					
	N	Incarceration Rate		N	Incarceration Rate		N	Incarceration Rate				
		Matched Cohort	State Care		Likelihood Ratio	Matched Cohort		State Care	Likelihood Ratio	Matched Cohort	State Care	Likelihood Ratio
1940-1944	96	4.47%	29.85%	9.10	324	0.78%	14.97%	22.39	420	1.64%	17.85%	13.03
1945-1949	189	6.24%	31.75%	6.99	372	1.23%	17.37%	16.88	561	2.92%	22.22%	9.50
1950-1954	435	9.33%	34.97%	5.23	516	1.84%	21.36%	14.49	948	5.26%	27.68%	6.89
1955-1959	981	10.99%	35.02%	4.36	924	2.07%	25.17%	15.91	1905	6.66%	30.13%	6.04
1960-1964	2034	11.63%	32.03%	3.58	1848	2.05%	23.58%	14.74	3882	7.07%	28.01%	5.11
1965-1969	2112	12.62%	35.95%	3.89	1914	2.35%	28.45%	16.52	4023	7.74%	32.41%	5.71
1970-1974	1926	12.48%	37.78%	4.26	1737	2.30%	28.25%	16.72	3663	7.65%	33.20%	6.00
1975-1979	624	12.16%	39.52%	4.72	657	2.02%	30.03%	20.82	1281	6.97%	34.49%	7.03
1980-1984	1521	10.16%	42.28%	6.48	1482	1.54%	27.70%	24.50	3006	5.91%	35.05%	8.59
1985-1989	735	9.75%	38.29%	5.74	591	1.84%	24.22%	17.05	1326	6.23%	31.83%	7.03

Legend:

N	The number of people in the State residential care dataset who were matched to known identities in the IDI. The process found a match for about three quarters of distinct people within that dataset.
Incarceration Rate	The proportion of people who have records of having received a custodial sentence based on data supplied to the IDI by the Department of Corrections and Ministry of Justice. For the matched cohort, we weighted results so that the biodata make-up matched that of the State residential care dataset. For the State care group, we made the conservative assumption that people who could not be found in the IDI had not received a custodial sentence.
Likelihood Ratio	The likelihood ratio of receiving a custodial sentence versus not is given by $p / (1-p)$, where p is the incarceration rate or proportion. The increased likelihood of receiving a custodial sentence for those in the State residential care dataset compared to the matched cohort is given by the likelihood ratio of the State care set divided by that of the matched cohort. This produces the likelihood ratios shown.

Issues Relating to Māori Data Governance and Sovereignty

The Commission is committed to the culturally appropriate treatment of data.

Please see:

- <https://www.abuseincare.org.nz/about-us/the-royal-commission-of-inquiry-into-abuse-in-cares-commitment-to-te-tiriti-o-waitangi/>
- <https://www.abuseincare.org.nz/about-us/working-with-maori/>
- <https://www.abuseincare.org.nz/about-us/te-taumata/>

This specific piece of research is aligned with and subsumed under the overarching Royal Commission of Inquiry into Abuse in Care.

It is noteworthy that Māori are substantially over-represented in the residential care population. The Royal Commission has a dedicated Māori Investigation Team aimed at ensuring appropriate treatment of data relating to Māori.

Disclaimer

The results presented 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/>

Acknowledgements

The Royal Commission acknowledges the work of Synergia who was retained to complete this research on behalf of the Inquiry.



Abuse in Care
Royal Commission of Inquiry