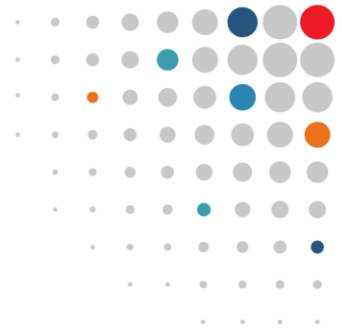




4 December 2024

File Ref: 2024013

9(2)(a)

On 4 November 2024 the Education Review Office transferred your Official Information Act request to the Social Investment Agency. You requested:

1. *The [Left Behind: How do we get our chronically absent students back to school?] report notes that chronically absent students are:
“Four times as likely to have a recent history of offending,” with 4 percent of chronically absent students having such a history, compared to less than 1 percent of all students. Could you clarify how this link was established? Specifically:*
 - *What data from schools was used to link students to their offending history?*

The analyses for this study were carried out within the Integrated Data Infrastructure (IDI). The IDI is a large research database run by Stats NZ. It includes a range of information about people’s interactions with government agencies – all of which is anonymised. It is a collection of linked datasets that allows evaluation and research on the pathways, transitions, and outcomes of people. There is one specific dataset, called the ‘spine’ which all of the other datasets link to. Attendance data is supplied to the Ministry of Education (MoE) by schools and loaded into the IDI by MoE. Data on offending is supplied to the IDI by Police NZ. Using the IDI, we are able to link records for anonymised individuals to understand the overlap between datasets.

- *Which agency provided the “history of offending” information, and what specific data points did it contain?*

The indicator, a proxy measure, was developed within the IDI using data on offending from Police for the period 2014-2019.

- *How were individual students identified and matched across datasets for attendance and offending history?*

The analyses were based on data in the IDI. The IDI is a large research database curated and hosted by Stats NZ. It contains matched, de-identified data on individuals and households in New Zealand collected by government agencies, Stats NZ surveys and non-governmental organisations. Stats NZ links an individual’s records across multiple datasets (such as school attendance, PHO enrolment, etc) before removing all identifiable information such as names, addresses, date of birth and agency unique identifiers (e.g. IRD or NHI numbers). Researchers are not able to identify any individuals in the IDI, only a unique identifier which

changes during every refresh (there are three refreshes in a year). Using the unique identifier, we can link across datasets. Researchers are only allowed to extract data at an aggregated level. All outputs were checked by Stats NZ to ensure anonymity was protected during the extraction process.

- *What measures were in place to ensure students' privacy in this data linkage?*

Data in the IDI are maintained and operated by Stats NZ. Stats NZ has a strong reputation as a data expert and trusted custodian of public information and is bound by the Statistics Act 1975 and the Privacy Act 1993 to protect the identities of people in the data it holds. Further information can be found here: <https://www.stats.govt.nz/integrated-data/how-we-keep-integrated-data-safe/>

To meet these legislative requirements, Stats NZ has a comprehensive guide and applied 'Five Safes' and 'Nga Tikanga Paihere' frameworks to manage safe access to the rich source of information about New Zealand people, households and businesses. Stats NZ provides access to integrated data if all the 'five safes' conditions are met: safe people; safe projects; safe settings; safe data; and safe output.

Safe people	Users of the IDI are vetted and trained in the safe use of data.
Safe projects	Research projects must be in the public interest. Research is restricted to the analysis of groups not individuals.
Safe settings	Data can only be accessed in secure settings.
Safe data	Personal identifiers are removed or encrypted by Stats NZ. Researchers get access <i>only</i> to the data relating to their research.
Safe output	Researchers must confidentialise output before data can be taken out of the Data Lab to ensure individuals cannot be identified.

Source: Stats NZ

- *Under what authority did the ERO access this data?*

ERO had commissioned the project to the Social Investment Agency (SIA) and SIA was approved by Stats NZ to conduct the research. SIA has experienced researchers that are familiar with the IDI and use of the IDI is a core function of the Agency. All researchers in the IDI are vetted and trained to use the IDI data safely. All IDI projects are scrutinised to ensure they are in the public interest before the research is approved and data access is granted.

2. *"Four times as likely to live in social housing," with 12 percent of chronically absent students living in social housing, compared to 3 percent of all students. Could you explain how this data link was determined? Specifically:*

- *What data was used to connect attendance information to students' social housing status?*

This data was provided to the IDI by MoE and Housing NZ (HNZ).

- *Which agencies were involved in sharing this information?*

Agencies who provided the data in the IDI were not involved in this analysis. As mentioned, ERO commissioned the project to SIA to conduct the project and SIA is responsible for this research. The findings are not official statistics of those agencies.

- *What specific data about social housing was provided to ERO?*

No specific data were given to ERO, however we provided anonymised descriptive statistics and findings from regression analysis.

3. *“At age 23, young adults who were chronically absent cost \$4,000 more than other young people,” with particular costs noted in corrections, hospital admissions, and receiving benefits. Could you specify:*
- *What data was used to link student attendance information with hospital admissions, corrections, and benefit receipt data?*
 - *What specific data points were shared to determine the cost differences?*
 - *Which agencies were involved, and what data was exchanged?*

The information used to derive the costs in this analysis was:

Agency	Description of IDI source data	Inflation adjustment
Ministry of Social Development	Dollar value of main benefit received	Adjusted to 2023 dollars using CPI
Ministry of Health/Te Whatu Ora	Casemix cost weighted usage from public hospitals	Using WIES 23 cost values
Pharmac	Subsidies attached to pharmaceutical prescriptions filled	Adjusted to 2023 dollars using CPI
Corrections	Lengths of sentences served combined with management costs per day	Using latest daily management cost
Whaikaha	Costs of disability support services recorded in the SOCRATES database	Adjusted to 2023 dollars using CPI

As above, these agencies were not directly involved in this analysis and the SIA relied on data they had supplied to the IDI. The SIA is responsible for this research and the findings are not official statistics of those agencies.

- *How was privacy maintained in the data-sharing process for these individuals?*

As above, privacy was maintained via Stats NZ safeguards surrounding use of the IDI, including the obligations placed on researchers granted access to the IDI.

- *Additionally, were students or their caregivers informed and asked to provide consent before any of this data was shared across agencies for the purposes of this report?*

The data that informs this report is routinely shared by agencies with Stats NZ to add to the IDI processes, rather than a data sharing exercise specific to this research project.

Parents/caregivers are informed on the school enrolment form that information collected at school is loaded into the Integrated Data Infrastructure (IDI) to be used for research purposes.

If you are not satisfied with this response, you have a right to seek an investigation or review by the Ombudsman. Information about how to make a complaint is available at www.ombudsman.parliament.nz or by calling 0800 802 602.

Yours sincerely

A handwritten signature in blue ink, appearing to read 'K Anderson', with a stylized flourish at the end.

Kirsty Anderson
Manager, Office of the Chief Executive