
The Graduate Outcome Project: Using Data from the Integrated Data Infrastructure Project

MALCOLM REES

Massey University, New Zealand

Submitted to the *Journal of Institutional Research* on 14 November 2013, accepted for publication 14 March 2014.

Abstract

This paper reports on progress to date with a project underway in New Zealand involving the extraction of data from multiple government agencies that is then combined into one comprehensive longitudinal integrated dataset and made available to trial participants in a way never previously thought possible. The dataset includes school leaver achievement data, enrolments and completions data from the tertiary education sector, with earnings data in the years after graduation from the Inland Revenue Department, government assistance through benefit dependency data from the Ministry of Social Development, and border-crossings data from Customs. This dataset allows us to track the destinations and earnings of graduates over a number of years after they have lost contact with the institution. This is authoritative population level data on some of the variables we all measure in other ways that is presented in one simple table and tracked over the years.

In the first instance, analysis of the data has focused on young graduates as this is a priority group in New Zealand. Through Massey University's participation, however, and because of our interest in data for our slightly different student demographic, we have been able to extend the analysis and utility to include data for all age groups and thus make use of a very comprehensive set of information. This article describes the data governance through Statistics New Zealand, analysis via the Ministry of Education, and the potential utility of the data for one participating tertiary provider.

Keywords: Graduates, integrated, data, infrastructure, employment.

The Graduate Outcomes Project (GOP) was developed by the Ministry of Education (MoE), in New Zealand in response to a government directive primarily aimed at improving educational outcomes for young graduates. The data used for the project was extracted using Statistics New Zealand's Integrated Data Infrastructure (IDI). This article introduces the GOP, but firstly describes the IDI and how the Graduate Outcomes data is derived. This is just one example of many projects in New Zealand utilising the IDI.

For many years government agencies in New Zealand have undertaken research independently of each other on behalf of the government and various other stakeholders. Until recently, however, it has not been possible to fully integrate these separate agency

datasets. The IDI project is a solution to this problem and integrated data from multiple government agencies can now be extracted in a way that ensures confidentiality, security, reliability and accuracy for any analysis or reporting.

Policy framework

The policy framework that underpins the IDI is based on three Acts of government. The first is the *Privacy Act, 1993* and its associated Code of Practice, the second is the *Statistics Act, 1975* and the third is the *Tax Administration Act, 1994*. These underpin the collection and storage of data and how it can be used for statistical or research purposes. They describe what data can be stored, how long it can be stored and how it can be used for research or statistical purposes.

In order to provide an assurance to members of the public that the data is safe, a series of strict protocols are applied to the collection, retention, integration and distribution of the information gathered (Statistics New Zealand, 2012). A privacy impact assessment was produced in 2013 when the IDI was extended. This document sets out four clear principles of the data governance:

1. The public benefits outweigh the privacy concerns about the use of the data and risks to the integrity of the official Statistics System, the original source data collections, and/or other government activities.
2. Integrated data will only be used for statistical or research purposes.
3. Data integration will be conducted in an open and transparent manner.
4. Data will not be integrated when an explicit commitment has been made to respondents that prevents such action (Statistics New Zealand, 2013a, p. 4).

Statistics New Zealand was assigned the task of managing the integrated dataset and this work was further supported through a previous Cabinet decision in 1997 that required: “Where databases are integrated across agencies from information collected for unrelated purposes, Statistics New Zealand should be the custodian of these datasets in order to ensure public confidence in the protection of individual records’ (Cabinet minutes, 1997, M31/4).

Datasets that make up the IDI

The datasets included in the IDI as at August 2013 can be seen in Figure 1. These were as follows:

- Accident Compensation Corporation: injury data
- Department of Corrections: sentencing data
- Inland Revenue: person and business tax data, student loans and allowances data
- Ministry of Business, Innovation and Employment – migration and movements data
- Ministry of Education: secondary school achievement data, tertiary education data
- Ministry of Justice: charges data
- Ministry of Social Development: benefit data, student loans and allowances data
- New Zealand Customs Service: departure and arrival cards data
- Statistics NZ: Household Labour Force Survey data
- Statistics NZ: New Zealand Income Survey data

- Statistics NZ: Survey of Family Income and Employment data
- Statistics NZ: Longitudinal Immigration Survey of New Zealand data
- Statistics NZ: Longitudinal Business Database data.

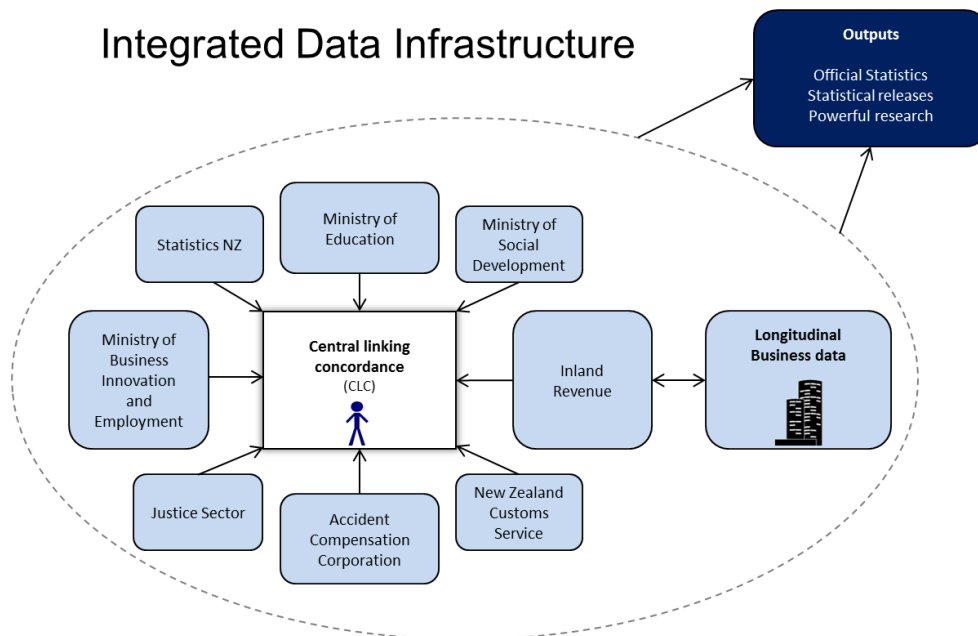


Figure 1. Datasets included in the Integrated Data infrastructure as at August 2013.

Limitations of the IDI data

There is a series of detailed business rules around how the linking of the data takes place. High rates of data linking are achieved through careful matching of unique identifiers within the data. There are limitations and possible errors with all data integration projects through erroneous linking; for example, where two records are linked when they should not have been or where two records should have been linked but were not. A unique identifier is applied to the data at individual record level by Statistics New Zealand and all identifying fields are removed or encrypted. The integrated data is then held by Statistics New Zealand and access is granted only for statistical or research purposes.

One of the key features of the IDI for a university is that the various components of the IDI are collected at the population level and at the level of the individual graduate. There is neither the sampling nor response-bias error that would be inherent in analysis if we had done it ourselves through surveying or other means.

Who can access data?

As the protocols describe, only approved researchers have access to the IDI data. Access to the data within the IDI is at the Government Statistician's discretion and is governed by the *Statistics Act*.

The Government Statistician may approve access to microdata if they are satisfied that:

- the data is needed to complete statistical research for the public good
- the researcher has the necessary research, knowledge, and skills to carry out the work
- the information will be used only for the purposes of the approved research
- the security and confidentiality of the microdata are protected.

Once researchers are approved they must sign a declaration of secrecy and follow strict rules to ensure information about individual people, households or businesses is not published or disseminated. Additionally, Inland Revenue restrict access to their specific data within the IDI and this is currently restricted to government employees working on Statistics NZ premises. Even the participation of Massey University required consent from us to release the data to the MoE before any analysis could take place. The MoE then undertook the analysis on our behalf. An additional consent is required before any institutional-level reporting takes place.

There are numerous examples of research that have used the IDI dataset. These are all available for download on the Stats NZ web site (Statistics New Zealand b). Projects to date relating to the tertiary education sector include:

- Papadopoulos, T. (2012). *“Who left, who returned and who was still away”*, *Migration patterns of 2003 graduates, 2004–2010*. Ministry of Business Innovation and Employment (MBIE)
- Mahoney, P., Park, Z., & Smyth, R. (2013). *Moving on Up: What young people earn after their tertiary education*.
- MoE. “The influence of education on outcomes” work in progress.
- MoE. “Who doesn’t participate in tertiary education” work in progress.

Graduate Outcomes Project

This project came about through the current national government’s election commitment in 2011 to boost skills and employment by increasing the education achievement for 25 to 34-year-olds. In order to increase the transparency of the information, resources have been provided to help inform study and employment decisions. The project had a number of specific parameters such as time-series data that was required as the focus was on young leavers who completed qualifications and the data should be disaggregated to level of study and field of study. The datasets for the project came from four of the sources in the IDI, namely:

- tertiary education data via the MoE
- earnings information from the IRD
- welfare benefit receipts from the MSD
- movements data from MBIE.

Limitations of the dataset include:

- The dataset is silent as to whether employment is part-time or full-time.
- The data does not have occupation code data (but it does have employer’s industry code).

- The analysis is restricted to young graduates. A young graduate is defined by the duration of the qualification. In the case of a 3-year degree, this is 24 years, while for a 5-year bachelor's degree, this is 26 years. Master's graduates are defined as young if they complete under the age of 27 years and doctoral graduates under the age of 29 years.
- The data is blind to what happens to graduates once they go overseas; we can see that someone has left New Zealand but cannot tell if they are in work or what they are earning while they are overseas.

The results of the Graduate Outcomes Project are described in the report called 'Moving On Up: What Young People Earn After Their Tertiary Education' (Mahoney, Park, & Smyth, 2013).

Some of the key findings from that analysis include:

- Median earnings for young bachelor's graduates are 53% higher than the national median five years after graduation¹.
- Employment rates increase with the level of qualification gained (56% of young bachelor's graduates were in employment one year after graduation and a further 38% were in further study). In contrast to this, only 37% of sub-degree or certificate graduates were in employment and 48% were in further study.
- Very few people who complete a qualification are on a benefit in the first five years after study (2% for bachelor's graduates).
- Young graduates who complete medical qualification have the highest median salary five years after graduation (\$110,000).
- Dental and pharmacy graduates are the next highest earners (\$76,100 & \$75,100 respectively).
- Bachelor's degree in Creative Arts have the lowest earnings and have a relatively high rate of benefit receipt.
- Qualifications associated with high rates of further study include:
 - Natural and physical sciences (58% in further study after 1 year)
 - Society and culture
 - Health
 - Agriculture
 - Environmental studies.

Why Massey University got involved in the trial?

Massey University became aware of the IDI project through the Moving On Up Project Report. However, because 40% of our enrolled students are over 29 years of age, and because many are studying by distance, the report did not apply to a significant proportion of our graduates. To some extent our participation has been a test of the possibilities of extracting data at the institutional level, but with a focus on all age groups. For the purposes of analysis the data has been categorised into four categories as follows; *Young, Young–34 years, 35–44 years, 45+ years*. This analysis was provided in July 2013.

¹ The data on earnings both in the Moving On Up report and in this article relate to earnings for the tax year ending 31 March 2009–2010. Data have been converted to 2011 NZ dollars using the Labour Cost Index.

Results

What the Massey University earnings data by qualification shows is that there is a clear differential in earning capacity by qualification at year one for young graduates; however, the difference diminishes between master's and doctorate graduates by year five (see Figure 2). The biggest increase in earnings for young Massey graduates is at the master's level with an increase of \$17,700 after five years, followed by bachelor's level with an increase of \$12,700 and finally doctorates with an increase of only \$7,900 after five years.

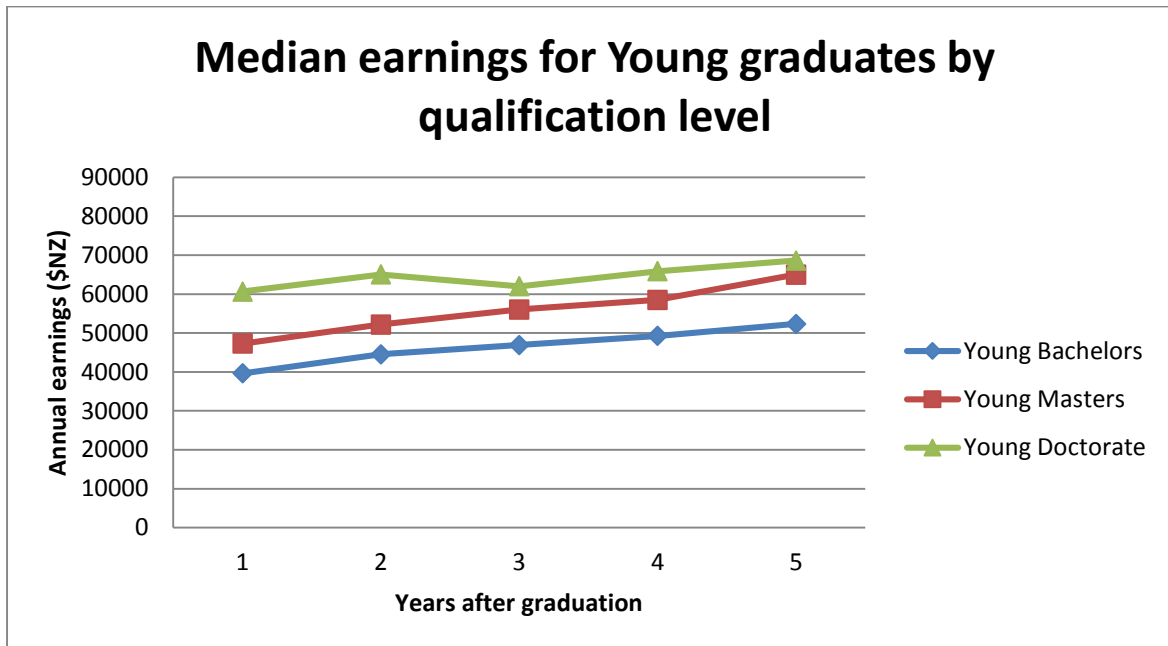


Figure 2. Young graduates' median earnings by qualification level over the five years since graduation.

Figure 3 shows a change to the pattern of increases in earnings with bachelor's graduates showing a consistent increase annually for a total increase at the 5-year mark of \$12,000 (similar to the young bachelor's graduates but master's graduates in this age category show the biggest total increase in earnings over that 5-year period of just over \$21,000 although with a dip in earnings at year four and a dramatic increase in year five). Doctoral graduates show a big increase overall of \$15,000 with most of that occurring by year four and a levelling off at year five.

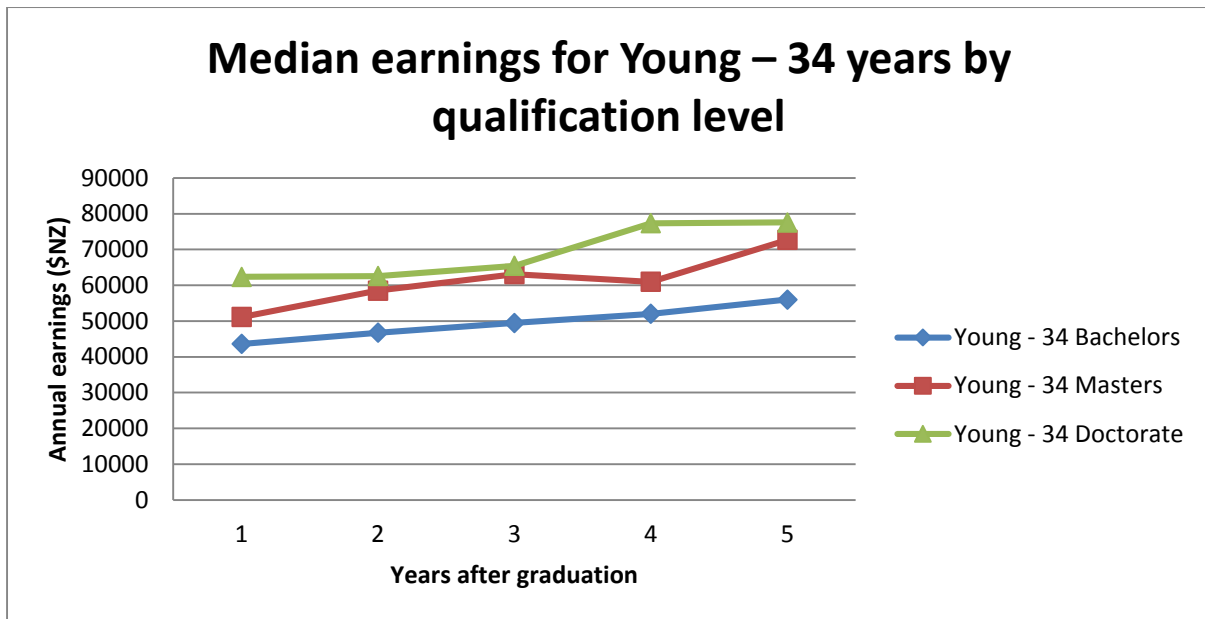


Figure 3. Median earnings for Young–34 years graduates for the five years since graduation.

Figure 4 shows that for the 35–44-year-old graduates there is a much higher starting income in year one of over \$70,000 for both master’s and doctorates compared with bachelor’s graduates. The difference in income after five years shows increases of \$13,000 for bachelor’s, \$11,000 for master’s and only \$7,000 for doctorates. By year three master’s and doctorates are earning approximately the same income.

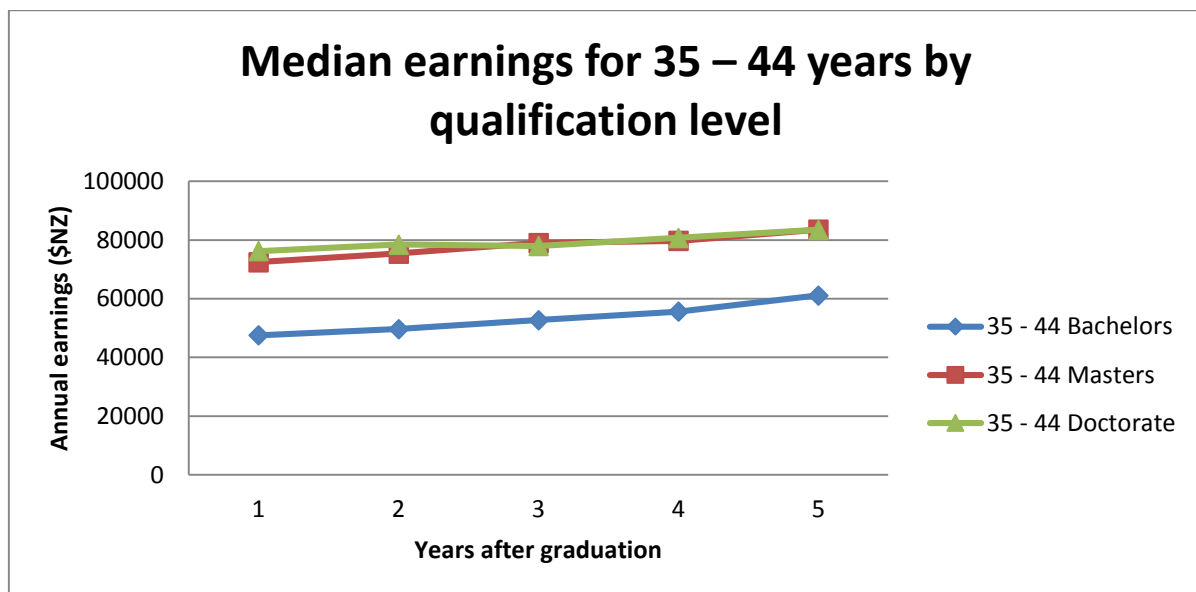


Figure 4. Median earnings for 35–44-year-old graduates by qualification level for the five years since graduation.

Figure 5 shows for the 45 year + age group, higher starting incomes for all groups and an increase for bachelor’s graduates of \$9,000; however, a levelling off for master’s earnings with only a \$4,000 increase and no increase in income at all for doctoral graduates (there was insufficient data for year five).

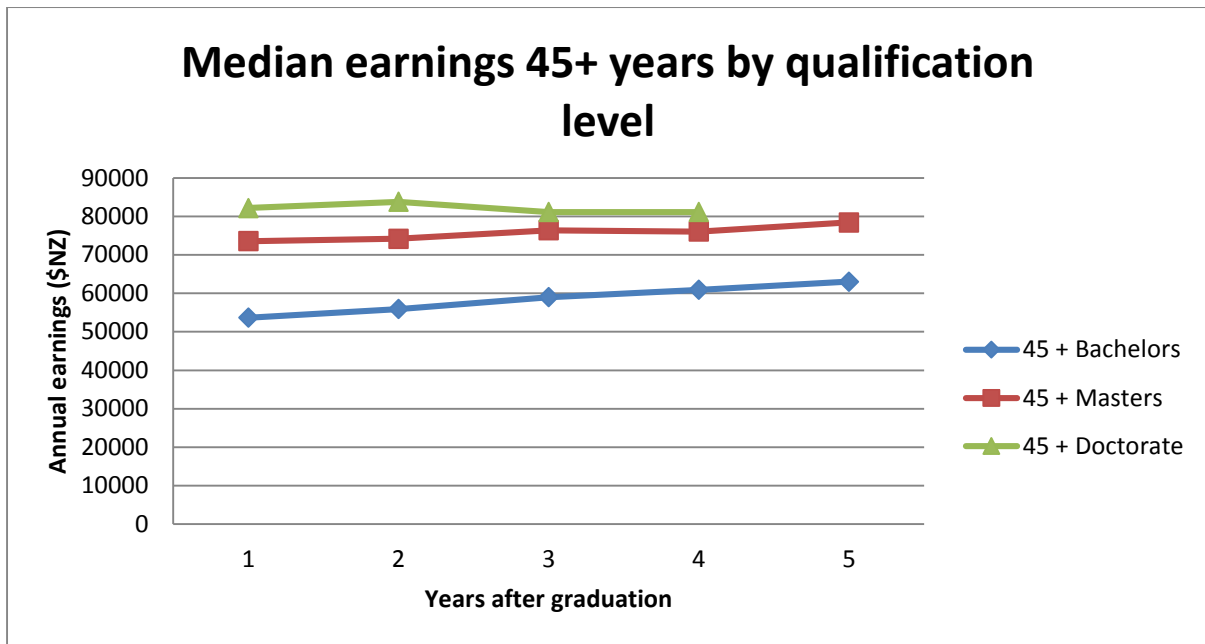


Figure 5. Median earnings for 45 years + graduates by qualification level for the five years since graduation.

The differences in earning capacity over the five years post-graduation are shown in Table 1.

Table 1

The Difference in Earnings Between Year 1 and Year 5 After Graduation for Each Qualification Type

Age category	Bachelor’s degree \$	Master’s degree \$	Doctorate degree \$
Young	12,700	17,724	7965
Young–34 years	12,300	21,531	15,194
35–44 years	13,585	11,078	7,270
45 + years	9,300	4,872	-1,013

Looking at the differences by qualification types by age band we see that for bachelor’s qualifications that there is a consistent increase in earnings regardless of age group (see Figure 6).

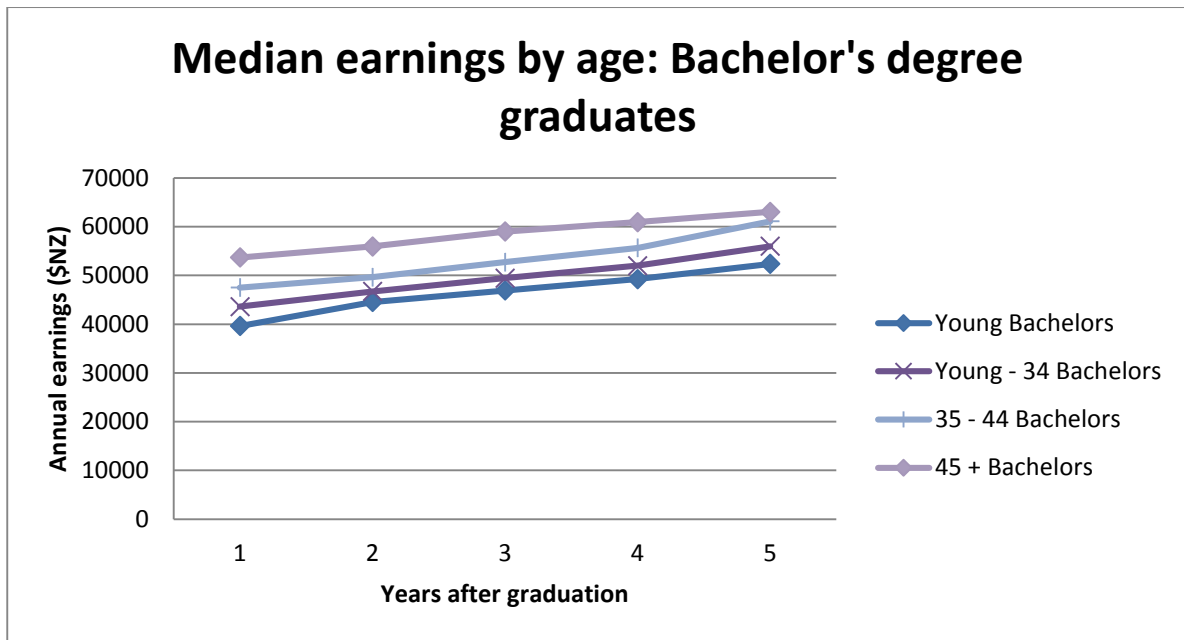


Figure 6. Median earnings for bachelor degree graduates for the five years since graduation.

Master’s graduates still have an earnings differential for the younger graduates; however, the differential is not so clear for the older age groups where the earnings reduces for 35–44 and again for the 45+ age group (see Figure 7).

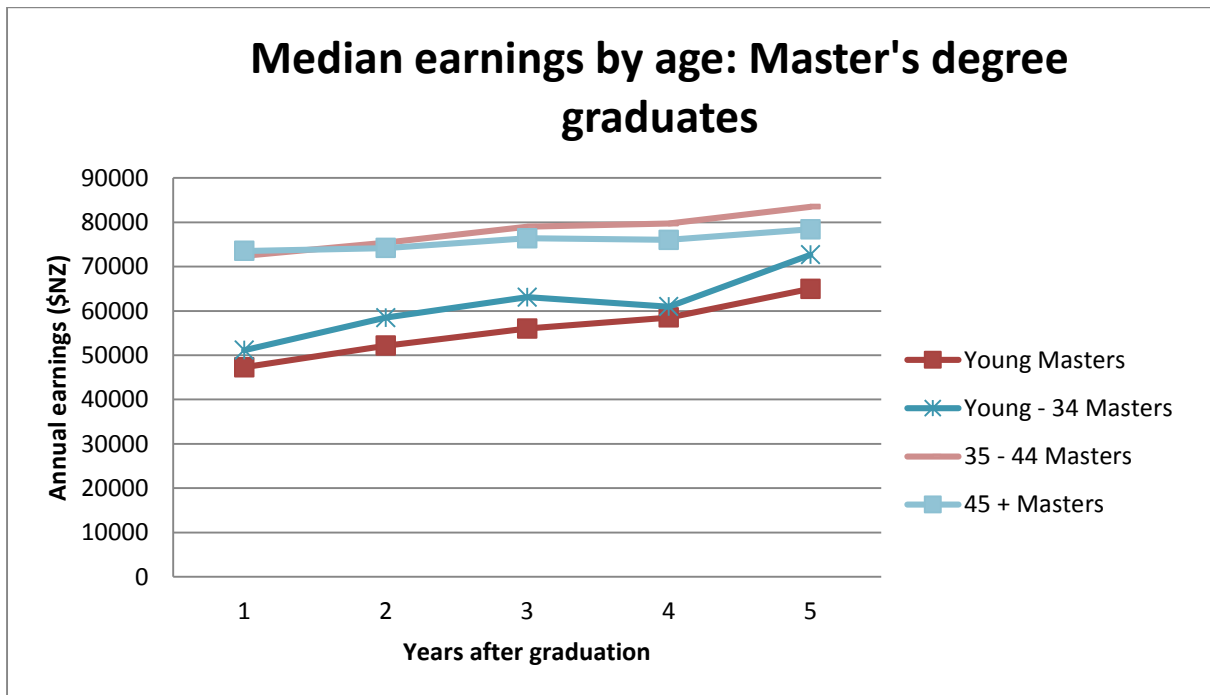


Figure 7. Median earnings for master’s degree graduates over the five years since graduation.

Doctoral graduates show quite different earnings trend by age than for the other qualifications, with only a modest increase in salary over time for young graduates and there is actually a decline in salary for the 45+ age group over four years (insufficient data was available for the 5-year analysis) (see Figure 8).

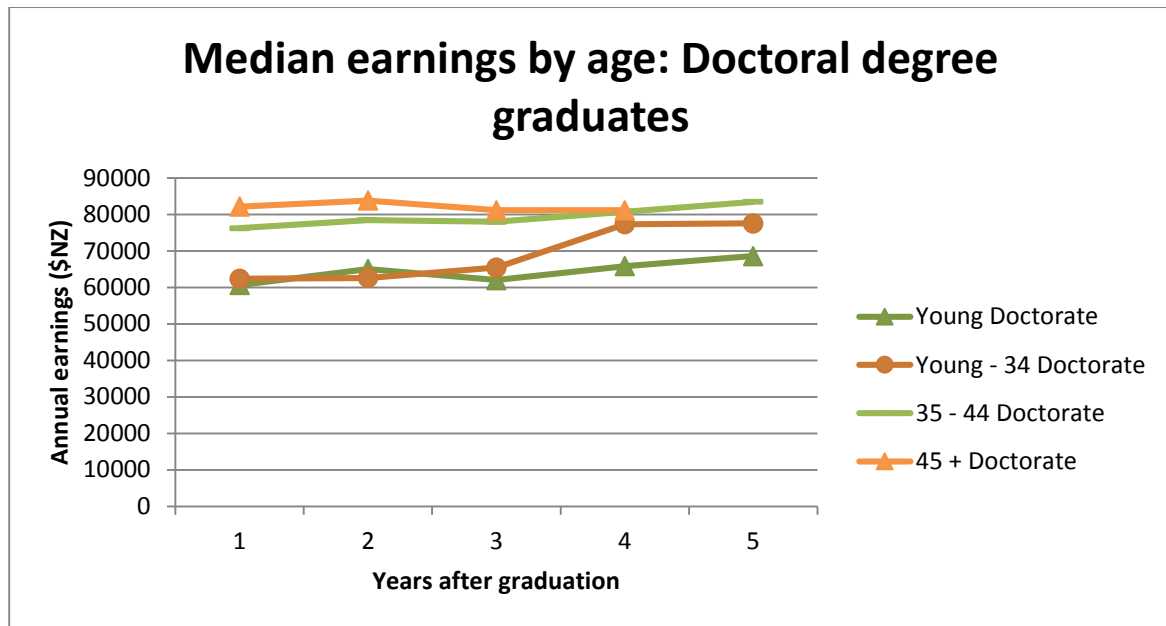


Figure 8. Median earnings for doctoral degree graduates by age group for the five years after graduation.

Using the IDI data, comparisons can be made by broad or narrow field of study (see Figure 9). This is just one example that compares the sciences broad fields of study. Engineering, Information science, and Architecture and Building take a similar earnings trajectory with an average median earnings of \$58,817 five years after graduation, whereas, Agriculture and Natural Sciences average \$51,794 over the same period.

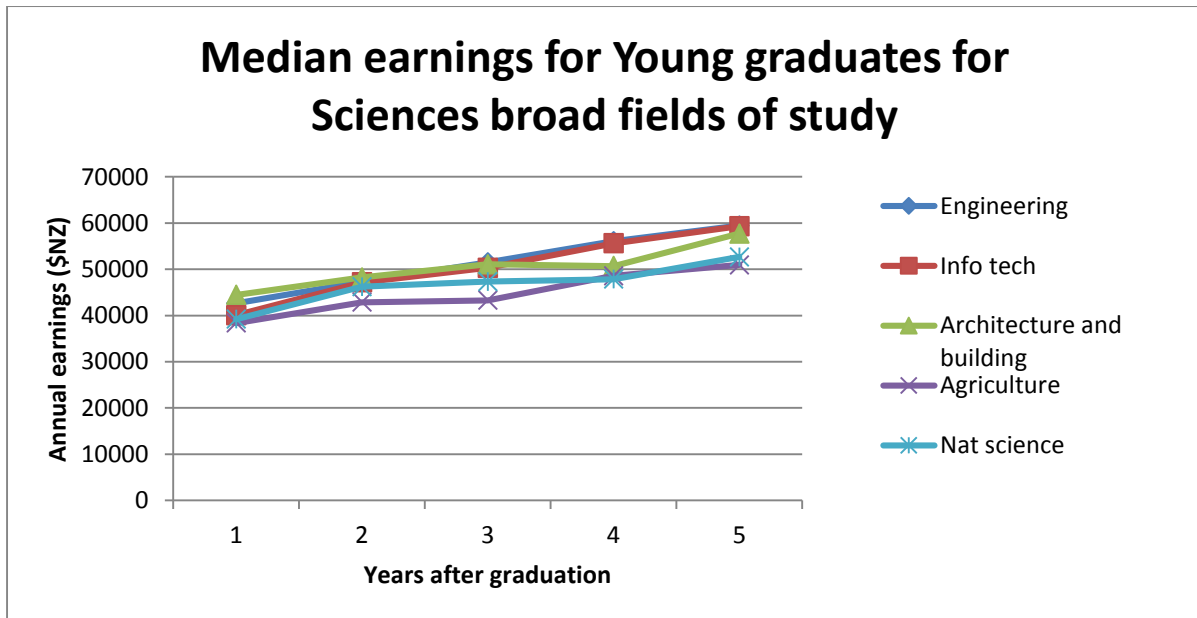


Figure 9. Median earnings for Young sciences graduates by broad field of study.

Migration

Another of the variables extracted in this analysis has been the movements data provided by MBIE. The migration of people from New Zealand is not an unexpected or new

phenomenon, although until the IDI data became available, very detailed analysis of emigration by educational qualification has been difficult to obtain. Table 2 shows the percentage of Massey graduates overseas five years after graduation for both bachelor's and master's graduates. There is insufficient data regarding doctoral graduates to be able to report this with any certainty.

Table 2

Percentage of Graduates Overseas Five Years After Graduation by Age Category

Age category	Bachelor's degree	Master's degree	Doctoral degree
	%	%	%
Young	27	36	Insufficient
Young–34 years	19	31	Insufficient
35–44 years	8	20	Insufficient
45+years	7	8	Insufficient

There is a wealth of information in the movements data that warrants further analysis and reporting. This analysis is ongoing. Some of the key points regarding migration include:

- Both bachelor's and master's graduate migration five years after graduation decreases as the age category increases.
- Graduates in postgraduate banking, finance and related narrow fields of study begin migrating overseas in year two and are still overseas in high numbers (50%) five years after graduation.
- Young bachelor's graduates with the lowest migration overseas (less than 15%, five years after graduation) include Political Science, Agriculture, Earth Science, Accountancy, Education, Building, Communication and Media, and Behavioural Science.

Benefit receipts

There are very few fields of study at bachelor's level or above where there is any significant evidence of benefit receipts, meaning that in most cases graduates at bachelor's level and above are able to obtain employment one year after graduation. The exception is the Creative Arts broad field of study, which has a consistent 5% benefit receipt annually; however, even this is not a very high overall percentage.

While it has been interesting to analyse the benefit dependency metric, the very low level of benefit receipt across all fields of study would suggest that there is little value investigating this further. The very low level means that there is little or no unemployment for those graduating with bachelor's qualifications and above. However, because the data is blind to occupation and to hours of work, we cannot tell if there is under-employment of graduates; that is, we do not know what proportion of our graduates are working in jobs that do not require a degree, nor can we distinguish between people who are working part-time or full-time.

Discussion

This article only provides a very small snapshot of what is possible using the IDI dataset. Much of the focus so far has been on earnings by age category with the extended Massey dataset, going beyond what was initially used for the Graduate Outcomes Project.

Only one university dataset has been reported at this stage; however, a similar national dataset would be useful so that we can make some informed comparisons between our own data and the national metrics. In addition to the analysis by age, the data could also be analysed nationally by a range of other variables such as gender, ethnicity and mode of study. A further analysis of the migration metric is already underway.

One of the key themes emerging from the Massey University data relates to the economic contribution, in particular, for both bachelor's and master's graduates across all age groups. The report from the graduate outcome project states that:

Many economists measure human capital by looking at people's earnings. The reason is that what an employer pays is an indicator of how much value a worker creates – because the employer cannot pay a person more than the value created by the employee. (Moving On Up report, p. 3)

Using earnings as a proxy for economic contribution, our data would suggest that given the earnings trajectory for all the age groups, especially at the bachelor's level, all are making a meaningful economic contribution. We would also predict further that many of those in the older age groups are studying by distance and therefore this mode of study is also making a very valuable contribution to the economy of this country.

What is not so easy to explain is the economic contribution of higher qualifications, except to show that doctoral graduates start with an income premium that remains consistently high; however, their earnings do not increase over time to the same extent as the bachelor's or master's graduates. There could be a number of reasons why this occurs, such as: difference in motivation for undertaking such qualifications in the first place, the high level of migration for many doctoral graduates and thus a bias in the measurement, or the effect that part-time employment may be having on the statistics. Certainly the economic impact of doctoral study warrants further research.

It is very easy to develop a fixation on earnings-related information in this analysis because the earnings data is so reliable and devoid of the limitations we regularly experience using our own survey data, such as the potential for response-bias or sampling error. We do hope, however, that students make career choices based on more than just earnings capacity—through good advice, and through support from academics and parents. In time it is hopeful that learning analytics can be included in the dataset to assist those decisions.

While the IDI is not the complete solution to our data needs, it does move us one step closer to the point where we may not need some of the survey analytics such as our own Graduate Destination Survey (GDS). The point at which we would seriously look at ceasing the GDS would be if occupation information could be included in the integrated dataset.

Author note

The results in this paper are not official statistics, they have been created for research purposes from the Integrated Data Infrastructure (IDI) managed by Statistics New Zealand. The opinions, findings, recommendations and conclusions expressed in this report are those of the author(s) not Statistics NZ. Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business or organisation and the results in this [report, paper] have been confidentialised to protect these groups from identification. Careful consideration has been given to the privacy, security and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the [Privacy impact assessment for the Integrated Data Infrastructure](#) available from www.stats.govt.nz.

The results are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information may be published or disclosed in any other form, or provided to Inland Revenue for administrative or regulatory purposes. Any person who has had access to the unit-record data has certified that they have been shown, have read, and have understood section 81 of the Tax Administration Act 1994, which relates to secrecy. 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.

References

- Mahoney, P., Park, Z., & Smyth, R. (2013). *Moving on up: What young people earn after their tertiary education*. Wellington, NZ: Ministry of Education, New Zealand.
- Papadopoulos, T. (2012). *Who left, who returned and who was still away? Migration patterns of the 2003 graduates 2004–2010*. Wellington, NZ: Ministry of Business, Innovation and Employment.
- Statistics New Zealand. (2012). *Privacy impact assessment for the integrated Data Infrastructure*. Wellington, NZ: Author.
- Statistics New Zealand. (2013a). *Integrated data infrastructure extension: Privacy impact assessment*. Wellington, NZ: Author.
- Statistics New Zealand. (2013b). *How are researchers using the integrated data infrastructure?* Retrieved from http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/researchers-using-idi.aspx.