New Zealand

The Integrated Data Infrastructure (IDI):
New Zealand's Bold Data Experiment

COMPASS Seminar Series
3 June 2016

Barry Milne

COMPASS Research Centre

University of Auckland

COMPASS RESEARCH CENTRE

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

Disclaimer

Access to the data presented was managed by Statistics New Zealand under strict micro-data access protocols and in accordance with the security and confidentiality provisions of the Statistic Act 1975. The findings are not Official Statistics. The opinions, findings, recommendations, and conclusions expressed are those of the researcher, not Statistics NZ.

The University of Auckland

Outline

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- Integrated Data Infrastructure (IDI)
  - What is it? Where did it come from?
  - Māori data issues
- Use of the IDI
  - Who is using it and how?
- Privacy and risk mitigation
  - Privacy protocols
  - Managing threats to privacy

Integrated Data Infrastructure (IDI)

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Integrated Data Infrastructure

Find out about the Integrated Data Infrastructure (IDI) and how it is used. This page includes related links for researchers who want to use the IDI.

About the Integrated Data Infrastructure

The Integrated Data Infrastructure (IDI) combines information from a range of organisations (such as health and education data) to provide the insights government needs to improve social and economic outcomes for New Zealanders.

With all personal information removed, integrated data gives a safe view across government so agencies can deliver better services to the public and ensure investment is made where it's needed most. Integrated data is particularly useful to help address complex social issues such as crime and vulnerable children.

Structure

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Education
- Early Childhood Education Providers
- Primary Education – From 2007
- Secondary Education – From 2004
- Tertiary Education – From 1994

Student Loans and Allowances
- From 1992

Benefits and services
- Benefit Dynamics Data – From 1990
- Child benefit, and 2nd and 3rd tier benefit payments – From 1993
- Youth Services – From 2006

Personal tax
- EMS – From 1999
- IR3 – From 1997
- IR20 and IR4S – From 1995

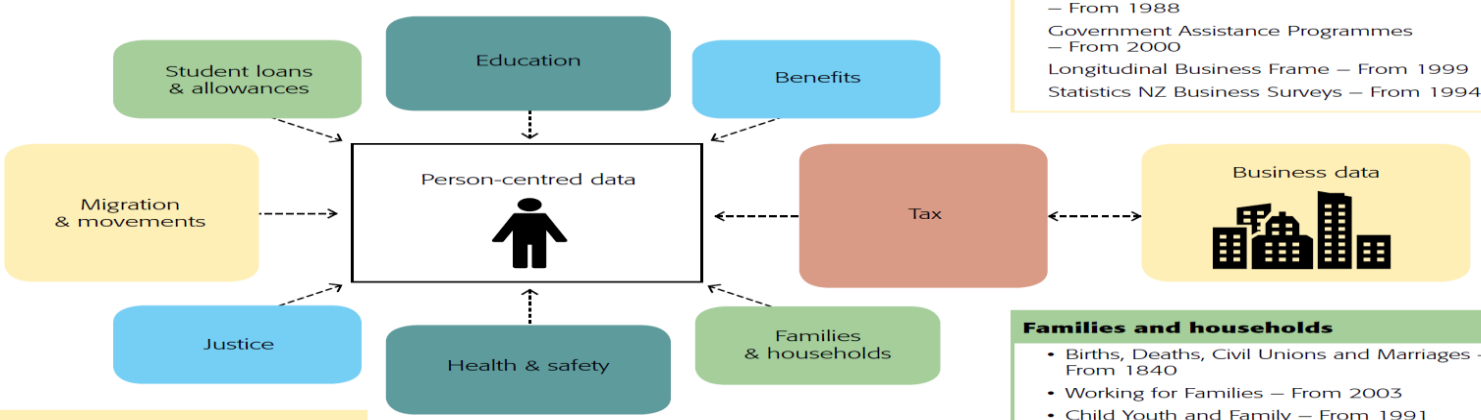
Business Centred Business tax
- IR10 – From 1999
- IR4 – From 1999
- GST – From 1992

Customs – Overseas merchandise trade – From 1988

Government Assistance Programmes – From 2000

Longitudinal Business Frame – From 1999

Statistics NZ Business Surveys – From 1994



Migration and movement
- Border movements – From 1997
- Visa applications – From 1997
- Departure and Arrival cards – From 1997
- Longitudinal Immigration Survey of NZ – 2005-2009
- Migrant Survey – From 2012

Health and safety
- Population Cohort Demographics – From 2004
- NHI and PHO Addresses – From 2004 and 2005, respectively
- PHO Enrolments – From 2003
- Chronic Conditions – Complete from 2007
- Cancer Registrations – From 1995
- Mortality – From 1988
- Pharmaceuticals – From 2005
- Publicly Funded Hospitals Discharges – From 1988
- NNPAC – National Non-admitted Patient Collection – From 2007
- Lab and GMS Claims – From 2003 and 2002, respectively
- NIR – From 2006
- B4School checks – From 2011
- PRIMHD – From 2008

Families and households
- Births, Deaths, Civil Unions and Marriages – From 1840
- Working for Families – From 2003
- Child Youth and Family – From 1991
- Tenancy – From 2000
- Household Economic Survey – From 2006
- Household Labour Force and NZ Income Surveys – From 2006
- Survey of Family Income and Employment – 2002-2010
- Social Housing Data – From 1980

Justice
- Recorded crime: victims – From 2014
- Recorded crime: offenders – From 2009
- Court charges – From 1992
- Sentencing and remand – From 1998

Census
- 2013 Census

ACC
- Claims and medical codes 1994-2015

New Zealand

The University of Auckland

History of NZ Integrated Data

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- Integrated Data
  - Data collected for one purpose from one agency merged at unit level with data collected for another purpose from [potentially] another agency

- 1997 Cabinet Directive
  - “where datasets are integrated across agencies from information collected for unrelated purposes, Statistics NZ should be custodian of these datasets in order to ensure public confidence in the protection of individual records”

History of NZ Integrated Data

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- SNZ Integration projects prior to IDI
  - Student Loans and Allowances integrated Dataset
    - IRD, MSD (StudyLink), MoE
  - Linked Employer-Employee Data (LEED)
    - IRD, SNZ Business Data Frame
  - LEED-MSD Benefit Dynamics Data
  - Employment Outcomes of Tertiary Education Study
    - LEED-MoE
  - LEED-Household Labour Force Survey

SNZ
Data Integration Policy

The Data Integration Policy (DIP) defines Statistics NZ's policy on integrating personal data. It applies to all integration of personal data that we undertake for statistical or related research purposes, including for the IDI.

The DIP includes a set of principles, which are based on the United Nations Economic Commission for Europe *Principles and guidelines on confidentiality aspects of data integration undertaken for statistical or related research purposes,* as follows:

Principle 1: The public benefits of integration outweigh both privacy concerns about the use of data and risks to the integrity of the Official Statistics System (OSS), the original source data collections, and/or other government activities.

Principle 2: Integrated data will only be used for statistical or research purposes.

Principle 3: Data integration will be conducted in an open and transparent manner.

Principle 4: Data will not be integrated when an explicit commitment has been made to respondents that prevents such action.

SNZ
Data Integration Policy

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Principle 4: Data will not be integrated when an explicit commitment has been made to respondents that prevents such action.

a) What commitments, both actual and implied, have been given to respondents about how their personal information will be used?

Respondents to our surveys have been informed that their information may be used for statistical purposes, but we ensure that no individual can be identified in data or results that are released.

The administrative data collections in the IDI[5] have no explicit commitments that prevent data from being integrated. Data use for statistics and research is an exemption to privacy legislation and codes.

Māori Data & Data Integration / IDI

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- **Indigenous Data Sovereignty**
  - Data is subject to the laws of the nation from which it is collected (incl Tribal nations)
  - 2007 UN Declaration on the Rights of Indigenous Peoples – Article 15:
    - "Indigenous peoples ... right to the dignity and diversity of their cultures, traditions, histories and aspirations … shall be appropriately reflected in education and public information"
- **Maori Data Sovereignty** recognizes that Maori data should be subject to Maori governance.
- Maori Data Sovereignty supports **iwi sovereignty** & the realisation of iwi aspirations.

Māori Data & Data Integration / IDI

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- Data is a highly valuable strategic asset for Maori development
- Data on Māori ubiquitous in IDI
  - Ethnicity (Census; Health); ancestry/iwi (Census, MoE)
- In the age of Big-Data, Maori want
  - access to data to support decision-making
  - to be involved when big-data is used to make decisions about them
    - Partnership model: work with government agencies
    - Ownership, Control, Access, Possession (OCAP®) http://fnigc.ca/

Use of the IDI

Structure

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Education
- Early Childhood Education Providers
- Primary Education – From 2007
- Secondary Education – From 2004
- Tertiary Education – From 1994

Benefits and services
- Benefit Dynamics Data – From 1990
- Child benefit, and 2nd and 3rd tier benefit payments – From 1993
- Youth Services – From 2006

Personal tax
- EMS – From 1999
- IR3 – From 1997
- IR20 and IR4S – From 1995

Student Loans and Allowances
- From 1992

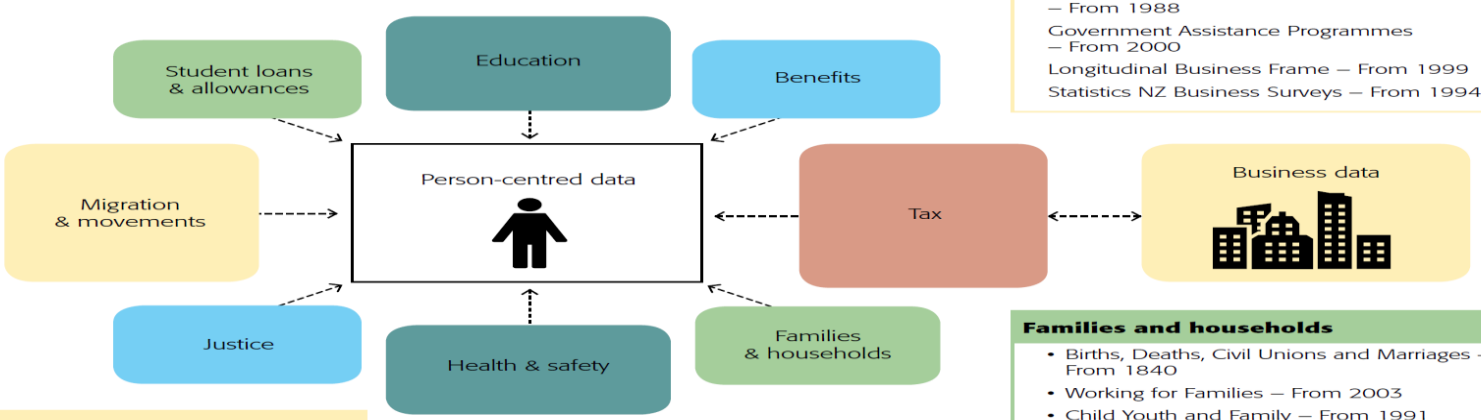
Business Centred Business tax
- IR10 – From 1999
- IR4 – From 1999
- GST – From 1992

Customs – Overseas merchandise trade – From 1988

Government Assistance Programmes – From 2000

Longitudinal Business Frame – From 1999

Statistics NZ Business Surveys – From 1994



Migration and movement
- Border movements – From 1997
- Visa applications – From 1997
- Departure and Arrival cards – From 1997
- Longitudinal Immigration Survey of NZ – 2005-2009
- Migrant Survey – From 2012

Health and safety
- Population Cohort Demographics – From 2004
- NHI and PHO Addresses – From 2004 and 2005, respectively
- PHO Enrolments – From 2003
- Chronic Conditions – Complete from 2007
- Cancer Registrations – From 1995
- Mortality – From 1988
- Pharmaceuticals – From 2005
- Publicly Funded Hospitals Discharges – From 1988
- NNPAC – National Non-admitted Patient Collection – From 2007
- Lab and GMS Claims – From 2003 and 2002, respectively
- NIR – From 2006
- B4School checks – From 2011
- PRIMHD – From 2008

Families and households
- Births, Deaths, Civil Unions and Marriages – From 1840
- Working for Families – From 2003
- Child Youth and Family – From 1991
- Tenancy – From 2000
- Household Economic Survey – From 2006
- Household Labour Force and NZ Income Surveys – From 2006
- Survey of Family Income and Employment – 2002-2010
- Social Housing Data – From 1980

Justice
- Recorded crime: victims – From 2014
- Recorded crime: offenders – From 2009
- Court charges – From 1992
- Sentencing and remand – From 1998

Census
- 2013 Census

ACC
- Claims and medical codes 1994-2015

New Zealand

The University of Auckland

Who is using the IDI?

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- 96 IDI projects listed on StatsNZ website
- 70 led by government departments
  - MBIE: 24; Treasury: 12; MinEdu: 9; MSD: 5
- 20 led by Universities
  - UOA: 9; AUT: 4; Otago: 3
- 6 led by other groups
  - Motu:4

http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/researchers-using-idi.aspx

Who is using the IDI? Government

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- Assessing the impact of KiwiSaver on the saving behaviour and retirement outcomes of New Zealanders
- How successful is New Zealand in retaining qualifications?
- Measuring housing affordability in New Zealand as a Tier One statistic
- The factors associated with student loan borrowers not being able to repay their loans, or to only repay very slowly
- The impacts of cancer, chronic disease, and acute health events on future employment, earnings, and benefit receipt among the working-age population
- The influence of education on outcomes
- The New Zealand rental sector: Who rents what, where, and from whom?

Who is using the IDI? Universities

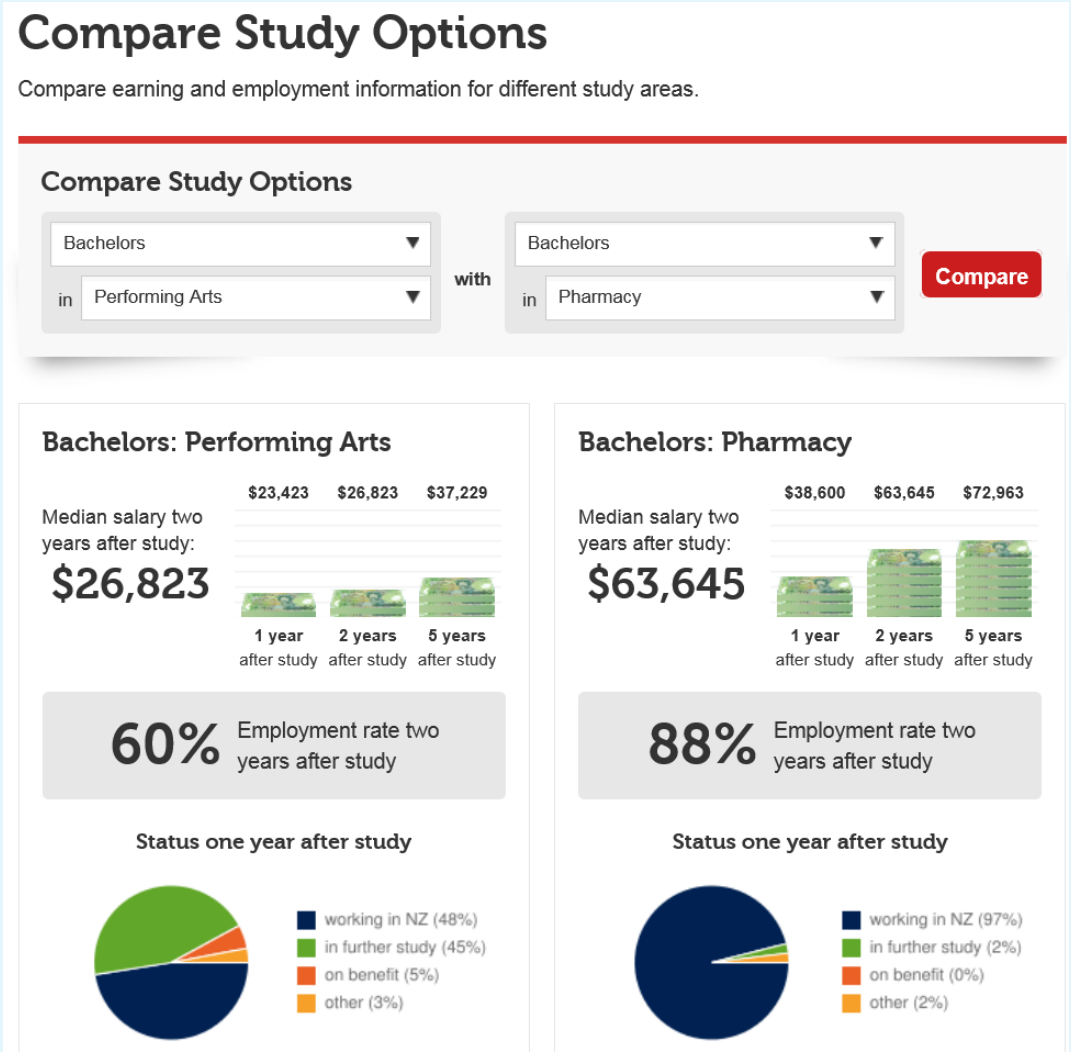
COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- A retrospective, quasi-experimental study to evaluate the effect of the Family Start home visiting programme on children's health, safety, and well-being, their preparedness for school, and maternal well-being
- BODE3 and HIRP-led VHIN research in the IDI
- Childhood poverty: precursors, persistence, prognosis, and policy
- Delivering a new measure of neighbourhood disadvantage for New Zealand
- Impact of the MeNZB™ vaccine on gonorrhoea
- Using birth information to predict reaching key early childhood development indicators: identifying at risk populations
- VIEW-IDI: Vascular risk informatics using epidemiology & the web research programme within the IDI

Outcomes of tertiary study

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

What should I study?



http://www.careers.govt.nz/ tools/compare-study-options/

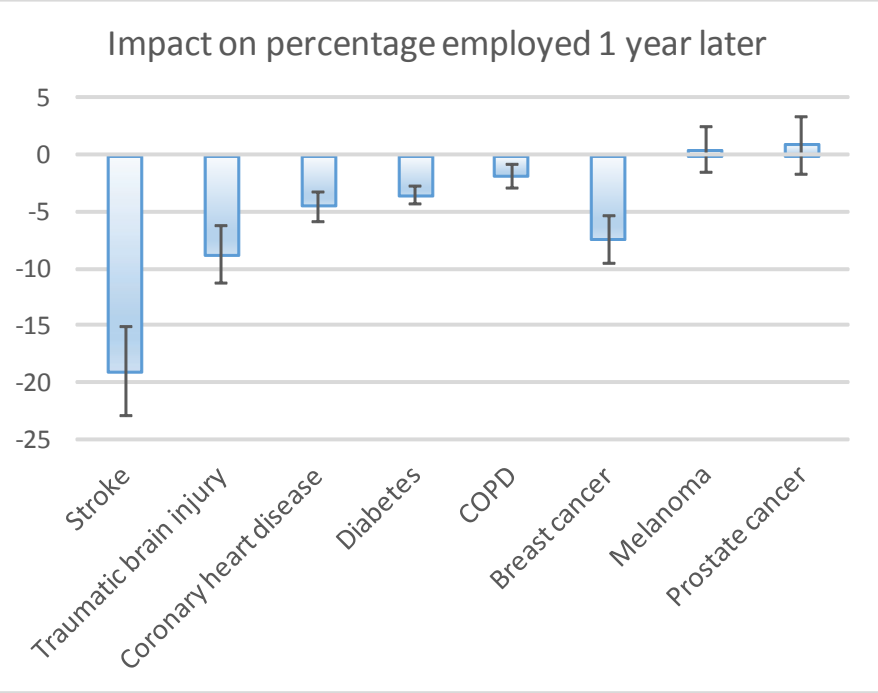
New Zealand

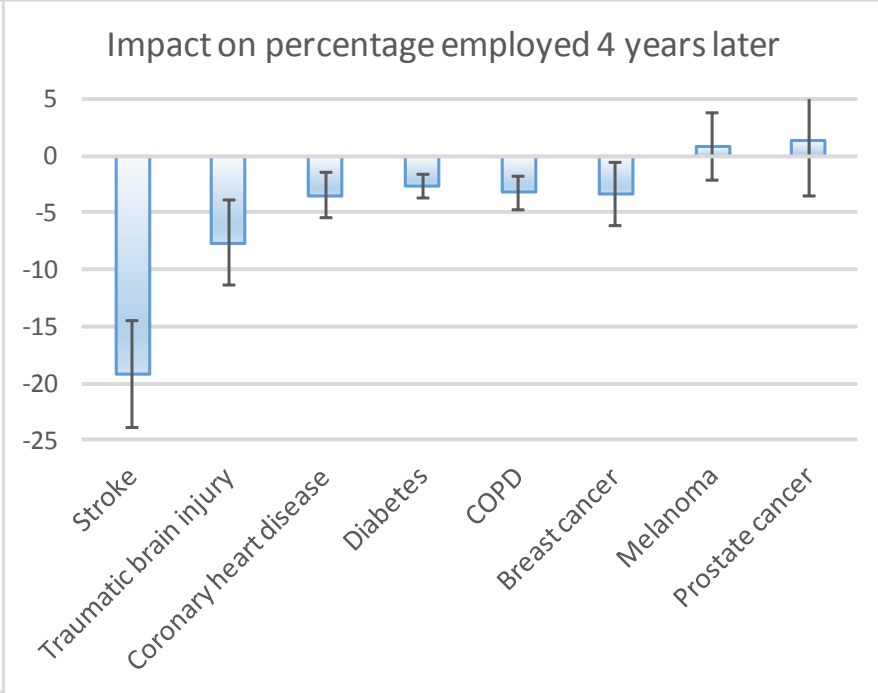
The University of Auckland

Impact of health conditions

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- What impacts do eight selected health conditions have on the **employment rates** of working adults who experience them?



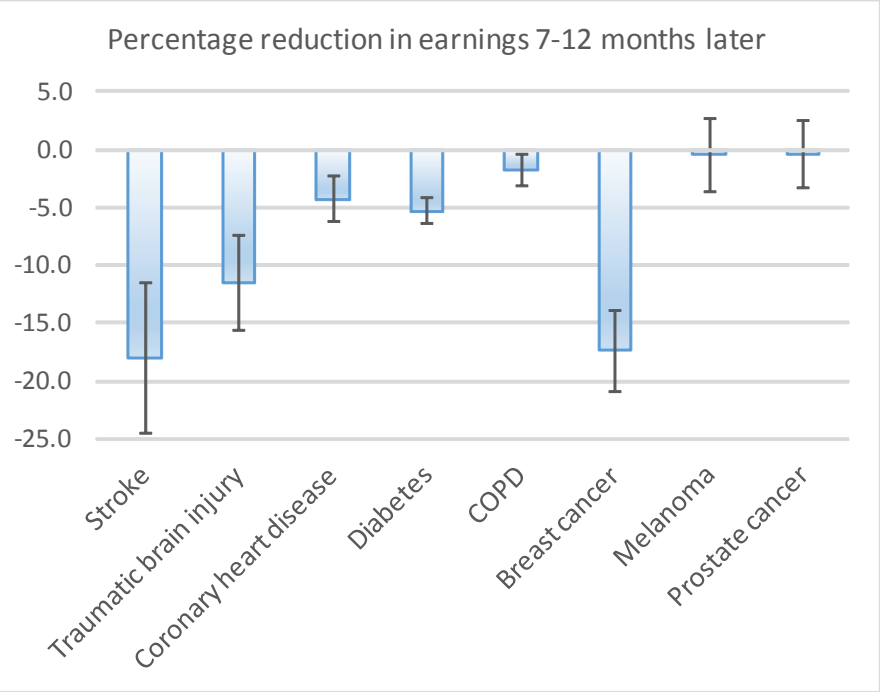


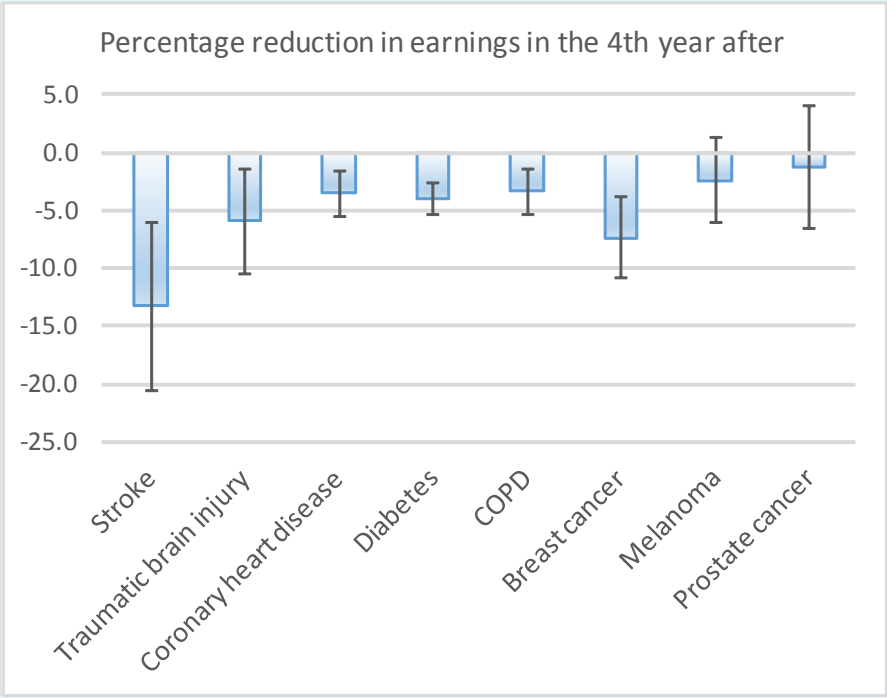
Research undertaken by Silvia Dixon, Treasury

Impact of health conditions

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- What impacts do eight selected health conditions have on the **earnings** of working adults who experience them?





Research undertaken by Silvia Dixon, Treasury

Childhood poverty

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Childhood poverty: precursors, persistence, prognosis, and policy

The project seeks to evaluate whether it is possible to use the Integrated Data Infrastructure (IDI) data to define childhood poverty. This involves defining households, determining households in poverty, and determining the children within those households.

This is part of a wider project to undertake a series of investigations looking at these questions:

1. Can children in poverty – and their families – be assessed using routine data available through the IDI?; and, assuming the answer to, (1) is yes;
2. Who becomes trapped in childhood poverty and why?
3. What is the duration of poverty experience for children, and how does the timing and persistence impact later health, education, employment, social, and justice outcomes?
4. What is the effect of lifting children from poverty?

Dr Barry Milne
University of Auckland

New Zealand
The University of Auckland

Better Start National Science Challenge

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Using birth information to predict reaching key early childhood development indicators: identifying at risk populations

This application will inform the Better Start National Science Challenge on risk factors associated with poor childhood outcomes (obesity, mental health, and poor literacy) with the purpose of improving targeted interventions.

This research will investigate the extent to which children at risk of these negative outcomes can be identified early using variables such as birth weight, parent's relationship status and age, and ethnicity. These risk factors will, after being linked to the before school checks table, provide evidence to the Challenge that these variables identify children at risk of poor developmental outcomes.

Dr Rick Audas and Dr Barry Milne
University of Otago

Privacy

Privacy protocols

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

It's not about 'you' as an individual

Your personal information is never seen by researchers. This is done in two ways. First – the information is anonymised. Personal identifying information like names, addresses, and days of dates of birth are removed. Second – all research findings are confidentialised. This means that information is grouped in a way which makes it impossible to identify individual people.

Keeping information safe

Statistics NZ operates within a 'five safes' framework to ensure that access to microdata is only provided if all of the following conditions can be met:

- safe people – researchers can be trusted to use data appropriately and follow procedures
- safe projects – the project has a statistical purpose and is in the public interest
- safe settings – security arrangements prevent unauthorised access to the data
- safe data – the data itself inherently limits the risk of disclosure
- safe output – the statistical results produced do not contain any disclosive results.

New Zealand
The University of Auckland

Privacy protocols

Six privacy impact assessments undertaken (to date)

- Privacy impact assessment for the Integrated Data Infrastructure (February 2012)
  Identifies privacy risks associated with the Integrated Data Infrastructure prototype and the IDI delivered in December 2012, and outlines the processes for managing them.

The IDI was extended in 2013 to include justice sector and health and safety data.

- Integrated Data Infrastructure extension: Privacy impact assessment (2nd ed) (October 2014)
  Identifies privacy risks associated with the Integrated Data Infrastructure extension delivered in August 2013, and outlines the processes for managing them.
- Integrated Data Infrastructure extension: Privacy impact assessment (3rd ed) (October 2015)
  Identifies privacy risks associated with the Integrated Data Infrastructure extension delivered in July 2015, and outlines the processes for managing them.
- Integrated Data Infrastructure extension: Privacy impact assessment (4th ed) (January 2016)
  Identifies privacy risks associated with the Integrated Data Infrastructure extension delivered in October 2015, and outlines the processes for managing them.
- Integrated Data Infrastructure extension: Privacy impact assessment (5th ed) (March 2016)
  Identifies privacy risks associated with the Integrated Data Infrastructure extension delivered in February 2016, and outlines the processes for managing them.
- Integrated Data Infrastructure extension: Privacy impact assessment (6th ed) (April 2016)
  Identifies privacy risks associated with the Integrated Data Infrastructure extension delivered in April 2016, and outlines the processes for managing them.

Access protocols

Access principles

Research proposals to access microdata will be assessed using the following principles:

- Access to microdata must be for statistical purposes and/or bona fide research purposes.
- Access to microdata must be consistent with the Statistics Act and other relevant legislation.
- Access to microdata is at the discretion of the Government Statistician.
- Access to microdata must protect respondents' confidentiality.
- Access to microdata must not adversely affect Statistics New Zealand's relationship with respondents.
- Decisions on requests for access to microdata will be provided through transparent processes.

Identified privacy risks

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

Risk	Likelihood of occurring	Consequence if it did occur	Risk rating
Individuals being re-identified in the data	Unlikely	Major	Medium
Unfavourable public perception of the data integration	Possible	Moderate	Medium
Maintaining data security	Unlikely	Severe	High
Data used for non-approved purposes	Rare	Major	Low

Privacy Risk Mitigation

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

1. Re-identification
   - Removal of identifiers, 'Safe Data'

Variables that must be removed from research datasets include:

- first names
- middle names
- last names
- titles
- business names
- day of date of birth
- day of date of death
- day of date of disposal
- day of date last seen
- address information.

New Zealand
The University of Auckland

Privacy Risk Mitigation

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

1. Re-identification

- Bone-fide researchers, 'Safe People'

Researchers using the IDI data will only have access to the specific datasets required for their research questions. Processes are in place to assess a potential researcher's integrity and experience. Once researchers are approved they must complete training in applying confidentiality methods and sign both a Declaration of Secrecy and a Researcher Undertaking before they can access the data.

The undertaking includes an agreement to:

- not attempt to identify particular persons or organisations
- not attempt to match the information with any other unit record level data source or list of persons or organisations
- not disclose, either directly or indirectly, information protected by the Statistics Act 1975 with any individual not approved by Statistics NZ
- apply confidentiality measures to all output to ensure that no individual person or organisation can be identified

New Zealand
The University of Auckland

Privacy Risk Mitigation

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

1. Re-identification
   - Confidentialised Output: 'Safe Output'
   - Counts
     - Small cell counts suppressed (<6): Can never release attribute information about a single individual, or even a small number of individuals
     - All cells 'random rounded to base 3'
   - Other
     - Income means to nearest $100 only
     - Minimum and maximums not released unless >=6 have those values

New Zealand
The University of Auckland

Privacy Risk Mitigation

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

1. Re-identification
   - DataLab Environment: 'Safe Settings'
     - Cannot transfer or email files outside of datalab environment (e.g., to another computer, usb drive)
     - No printing, photography
     - Non-approved people not allowed in datalab environment
   - Research is for statistical purposes, and is a public good research project: 'Safe projects'
     - All projects approved by Government Statistician (independent of Government by law)
     - Projects encouraged to seek ethical approval

Privacy Risk Mitigation

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

1. Re-identification
   - All SNZ safeguards necessary in my view

- Re-identification Disasters
  - Latanya Sweeney and release of anonymised health care data in Massachusetts
  - Tennis gambling scandal
  - Care.Data initiative in UK

Privacy Risk Mitigation

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

2. Unfavourable public perception
   - Research Report, 'Public attitudes to data integration' commissioned by SNZ, undertaken by Opus found:
     - Most New Zealanders appear to have a relatively positive perception of data integration by public sector agencies
     - Attitudes to data integration do not appear to be strongly associated with particular types of social groups
     - Acceptability of data integration appears to be largely influenced by the individual's own personal experiences
     - There also appears to be a more general value-based concern around appropriate use

OPUS

New Zealand
The University of Auckland

Privacy Risk Mitigation

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

3. Maintaining Data security

Statistics NZ has well-established policies, procedures, and systems in place to ensure adequate measures of physical and electronic security, including:

- physical security systems controlling entry to premises and sections of premises to authorised people only
- visitors to Statistics NZ are subject to strict registration and supervision procedures; systems ensure their activities are confined to legitimate business
- access to data is in accordance with Statistics NZ's Security Framework
- access to Statistics NZ's IT systems requires a valid userid and password. A Statistics NZ security office actively audits and reviews security processes and addresses new and emerging threats.

Additional security arrangements for the infrastructure.

- All data collections and associated electronic workspaces will be secured (access will only be authorised for project personnel who need to access data for specific tasks, and to selected IT administrators who are required to maintain the IT system).
- Datasets will not be available to third parties.
- Regular audits of individuals able to access the dataset will be made.

Privacy Risk Mitigation

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

4. Data used for non-approved purposes
   - E.g., not for a public good research project, but instead to identify an individual
   - Five safes as per 1.

Summary

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- The IDI is a large, whole population dataset made up of constituent data from govt agencies, linked at the person level
- Use to date has mainly been led by govt agencies, but use by academic researcher is increasing
- Robust systems are in place to protect the privacy of individuals and public confidence in the initiative
  - But one screw up could ruin it for everyone….

QUESTIONS?

Re-identification risk

COMPASS RESEARCH CENTRE
FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND
Whare Wānanga o Tāmaki Makaurau

- Massachusetts Group Insurance Commission released anonymised health data on state employees (1997)
  - To enable research to improve healthcare
  - Mass Governor assured public the privacy was protected by removal of identifiers
- MIT CompSci grad student, Latanya Sweeney
  - Accessed the health data
  - Accessed electoral roll ($20) for Cambridge, Mass (where Governor lived), incl name, address, ZIP code, birthdate and sex of every voter
  - Six in Cambridge shared Gov's birthdate; only three were men and only one lived in his ZIP code
  - She mailed all of the Gov's health records to him…

Re-identification risk

As part of the investigation, John Templon, an investigative data reporter for BuzzFeed News, spent more than a year analyzing 26,000 professional men's matches and found 15 players who lost matches with unusual betting patterns "startlingly often." (Match-fixing is also believed to occur in professional women's tennis, but the BuzzFeed-BBC investigation focused only on men's tennis, so we are in this article, too.) BuzzFeed and BBC didn't name these players, citing a lack of evidence of wrongdoing and possible alternative explanations for underperformance, including injury. But BuzzFeed did release an anonymized version of the data it used on GitHub, including a file containing betting odds and the year for 129,271 matches.

Quickly, people wrote on Twitter and on GitHub that the data could be de-anonymized, thereby identifying the 15 players Templon mentioned. Ian Dorward, a London-based tennis bettor who used to set and adjust tennis betting lines for a bookmaker, emailed me the list of what he believed to be the 15 names. After Chris Bol, a data analyst based in Utrecht, the Netherlands, published the same names, Dorward went public with his findings, which criticized BuzzFeed for making the data relatively easy to crack.[1]

http://fivethirtyeight.com/features/why-betting-data-alone-cant-identify-match-fixers-in-tennis/ 16 Jan 2016

Re-identification risk

- Care.Data Initiative (UK, 2014)
  - Detailed NHS data made available to researchers
- Privacy campaigners noted
  - If details of an individual's treatment were in the public domain (e.g., Ed Milliband's nose operation), that individual identifiable in the dataset, and all treatments within NHS able to be revealed for that individual
- In the case of IDI, this would be mean all health details, tax details, benefit details, justice details, education details…
  - The dangers of n=1…
  - And the importance of SNZ's safes…