

# Ten Years of the Integrated Data Infrastructure (IDI): What have we learned and what are the unresolved issues?



COMPASS  
RESEARCH CENTRE

FACULTY OF ARTS  
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

## Barry Milne

COMPASS Seminar Series

3 May 2022

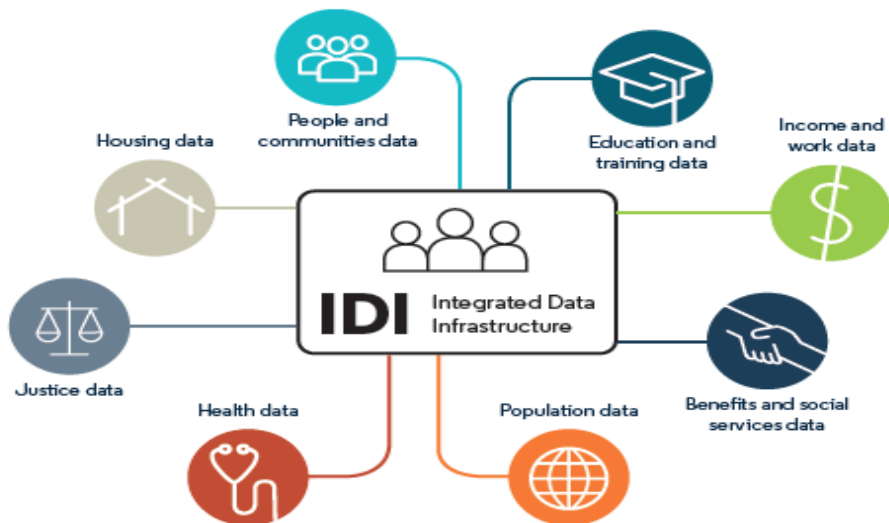
# 1. What is the IDI?



An integrated database containing confidentialised microdata about people and households

- ❑ Linked at the person-level (for most data tables)
  - ❑ Linkable, through tax data, to ‘Longitudinal business database (LBD)’
- ❑ ‘Ever resident’ population
- ❑ Longitudinal
- ❑ Intergenerational

Stats NZ's Integrated Data Infrastructure (IDI) is a large research database containing de-identified microdata about people and households.



The IDI contains person-centred microdata from a range of government agencies, Stats NZ surveys including the 2013 Census, and non-government organisations. For more information about data in the IDI, see [www.stats.govt.nz/integrated-data/integrated-data-infrastructure](http://www.stats.govt.nz/integrated-data/integrated-data-infrastructure)

The Longitudinal Business Database (LBD) complements the IDI with microdata about businesses. For more information about data in the LBD, see [www.stats.govt.nz/integrated-data/longitudinal-business-database](http://www.stats.govt.nz/integrated-data/longitudinal-business-database)

## Benefits and social services data

- ACC injury claims – from 1994
- Benefits – from 1990
- Children's Action Plan – from 1996
- Child, Youth, and Family – from 1991
- Early Start Project – from 2016
- Family Start – from 2008
- Student loans and allowances – from 1992
- Working for Families – from 2003
- Youth services – from 2004

## Education and training data

- Early childhood education participation – from 2008
- Industry training – from 2001
- Primary education – from 2007
- Programme for the International Assessment of Adult Competencies – from 2014
- Secondary education – from 2004
- Targeted training – from 2001
- Tertiary education – from 1994

## Health data

- B4 School Checks – from 2011
- Cancer registrations – from 1995
- Chronic conditions – from 2007
- General medical services claims – from 2002
- Health tracker – 2006–14
- Immunisation – from 2006
- interRAI – from 2014
- Laboratory claims – from 2003
- Maternity – from 2003
- Mortality – from 1988
- National Booking Reporting System – from 2003
- National Needs Assessment and Service Coordination Information System (SOCRATES)
- National non-admitted patient collection – from 2007
- NES enrolments – from 2019
- Pharmaceuticals – from 2005
- PHO enrolments – 2003–2019
- Population cohort demographics and addresses – from 2004
- PRIMHD – from 2008
- Privately funded hospital discharges – from 2001
- Publicly funded hospital discharges – from 1988

## Housing data

- Social housing – from 2000
- Tenancy – from 2000

## Income and work data

- Household economic survey – from 2006
- Household labour force survey – from 2006
- NZ income survey – from 2006
- Survey of family, income, and employment – 2002–10
- Tax and income – from 1999

## Justice data

- Court charges – from 1992
- New Zealand crime and victims survey – from 2018
- NIA links – from 2009
- Recorded crime: offenders – from 2009
- Recorded crime: victims – from 2014
- Sentencing and remand – from 1998

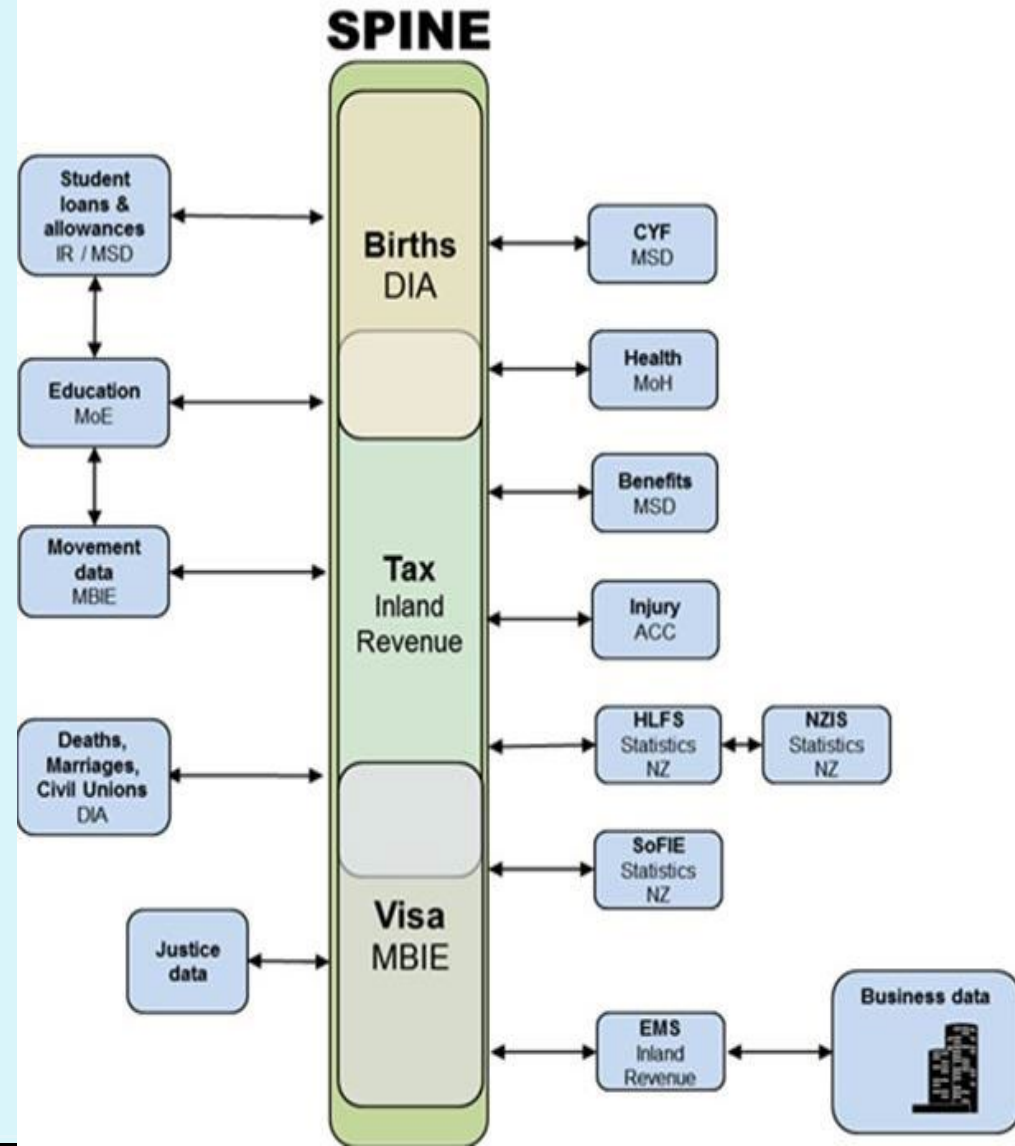
## People and communities data

- Auckland City Mission – from 1996
- Disability survey – 2013
- Driver licence and motor vehicle registers
- General social survey – 2008–2018
- Longitudinal immigration survey of NZ – 2005–09
- Migrant survey – from 2012
- Te Kupenga – 2013

## Population data

- Census – 2013, 2018
- Births, deaths, and marriages – from 1840
- Border movements – from 1997
- Civil unions – from 2005
- Departure and arrival cards – from 1997
- Visa applications – from 1997

# 1. What is the IDI?



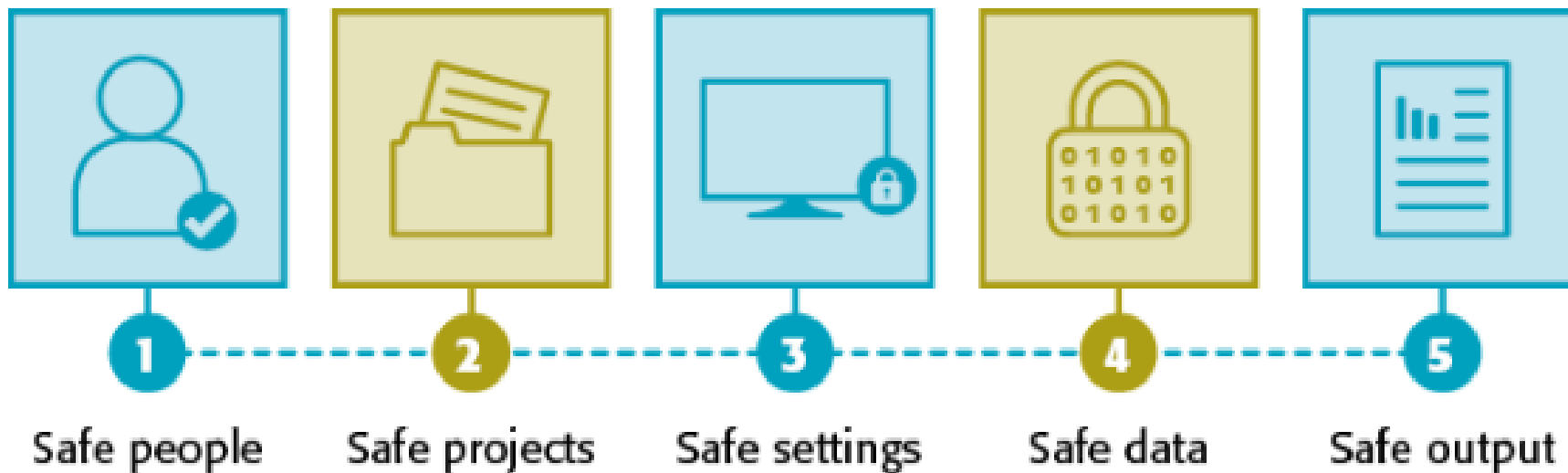
# History of the IDI

- 2011 IDI prototype
  - Labour market focused to allow linking of individual, household, and business-level data, with the following components
    - iLEED (integrated Longitudinal Employment and Education Data)
    - Migration and movements data
    - Longitudinal Business Database (LBD) data
  - Privacy impact assessment (Feb 2012)
- 2013 Cabinet agree to expand IDI – Better Public Services
  - Seven additional privacy impact assessments (June 2013 – July 2016) as additional datasets (ACC, Education, Justice, Census, health) were added
  - Overarching privacy impact assessments July 2017
- 2020-2021 Covid-19 data (wage subsidies, Immunisation Register)
- New data additions undergo a ‘Data Ethics and Privacy Assessment’

## The five safes

We use a 'five safes' framework to ensure that we provide access to data only if all of these five conditions are met.

### The 'five safes' framework for the IDI



- ❑ Application deadlines every 6 weeks
  - ❑ Describe project (public good research)
  - ❑ List researchers
    - All need to be confidentiality trained
  - ❑ List data requested
    - Only the tables you are approved to use will be ‘unlocked’
  - ❑ Ngā Tikanga Paihere framework. Need to demonstrate:
    - Experience working with the population(s) of interest
    - Populations of interest support the research, and have been (and will continue to be) consulted about the research
    - The value of the research to the populations of interest
    - How potential risks (if any) to the populations of interest will be managed/mitigated

# 2. What have we learned?

## i. The IDI is a popular research tool

Approved Microdata projects

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Projects using integrated data	7	14	22	36	62	64	96	97	78	81	20
Projects not using integrated data	13	12	18	28	8	11	11	9	4	4	0
Total approved or in progress projects	20	26	40	64	70	75	107	106	82	85	20

Table from Michael Challands, Stats NZ



# 2. What have we learned?

## ii. Publications increasing year on year

Title	Creator	Item Type	Year
> Boosting productivity in New Zealand by unleashing digitalisa...	Yashiro et al.	Report	2022
> Investigating the transmission risk of infectious disease outbre...	Turnbull et al.	Journal Arti...	2022
Patterns across debt and debtors to government: Connections...	Social Wellb...	Book	2022
Implementing pay equity for social services	Social Wellb...	Book	2022
Analysis: COVID-19 vaccine uptake for disabled people	Social Wellb...	Book	2022
> Sweat Equity: Student Scholarships in Aotearoa New Zealand's...	Soar et al.	Journal Arti...	2022
Effectiveness of hospital transfer payments under a prospectiv...	Schumacher	Journal Arti...	2022
Income and extratropical cyclones in New Zealand	Roy et al.	Journal Arti...	2022
Longitudinal Associations of Mental Disorders With Dementia:...	Richmond-...	Journal Arti...	2022
Indigenous people in Aotearoa New Zealand are overrepresen...	Rapana et al.	Journal Arti...	2022
A statistical person register in New Zealand: Progress and chal...	Page et al.	Journal Arti...	2022
Residential mobility and potentially avoidable hospitalisations ...	Nathan et al.	Journal Arti...	2022
Inter-relationships between geographical scale, socio-econom...	Mills et al.	Journal Arti...	2022
> Melatonin dispensing and polypharmacy rates for New Zealan...	McLay et al.	Journal Arti...	2022
> Firm Productivity Growth and the Knowledge of New Workers	Kirker and S...	Report	2022
> The impact of the 2018 Families Package Accommodation Sup...	Hyslop and ...	Report	2022
> Essays on Disaster Risk Management and Policy in New Zealand	Hoang	Thesis	2022
> Quality of life, quality of business, and destinations of recent g...	Grimes et al.	Journal Arti...	2022
> Social determinants and inequitable maternal and perinatal ou...	Dawson et al.	Journal Arti...	2022
The Devil is in the Details: Identifying Unbiased Link between ...	Dasgupta et...	Journal Arti...	2022
> Skills, Economic Crises and the Labour Market	Dasgupta a...	Report	2022
Further reductions in the prevalence of obesity in 4-year-old ...	Daniels et al.	Journal Arti...	2022
Economic insecurity during the COVID-19 pandemic: insights f...	Clyne and S...	Journal Arti...	2022
> Relative contribution of trends in myocardial infarction event r...	Camacho et...	Journal Arti...	2022
New Zealand Socio-economic Index 2018.	Boven et al.	Book	2022
> A longitudinal linkage study of occupation and ischaemic hear...	Barnes et al.	Journal Arti...	2022
> How wages respond to the job-finding and job-to-job transiti...	Ball et al.	Book	2022
> Who benefits from firm success? Heterogeneous rent-sharing ...	Allan and M...	Report	2022
> Benchmarking New Zealand's frontier firms	Zheng et al.	Book	2021

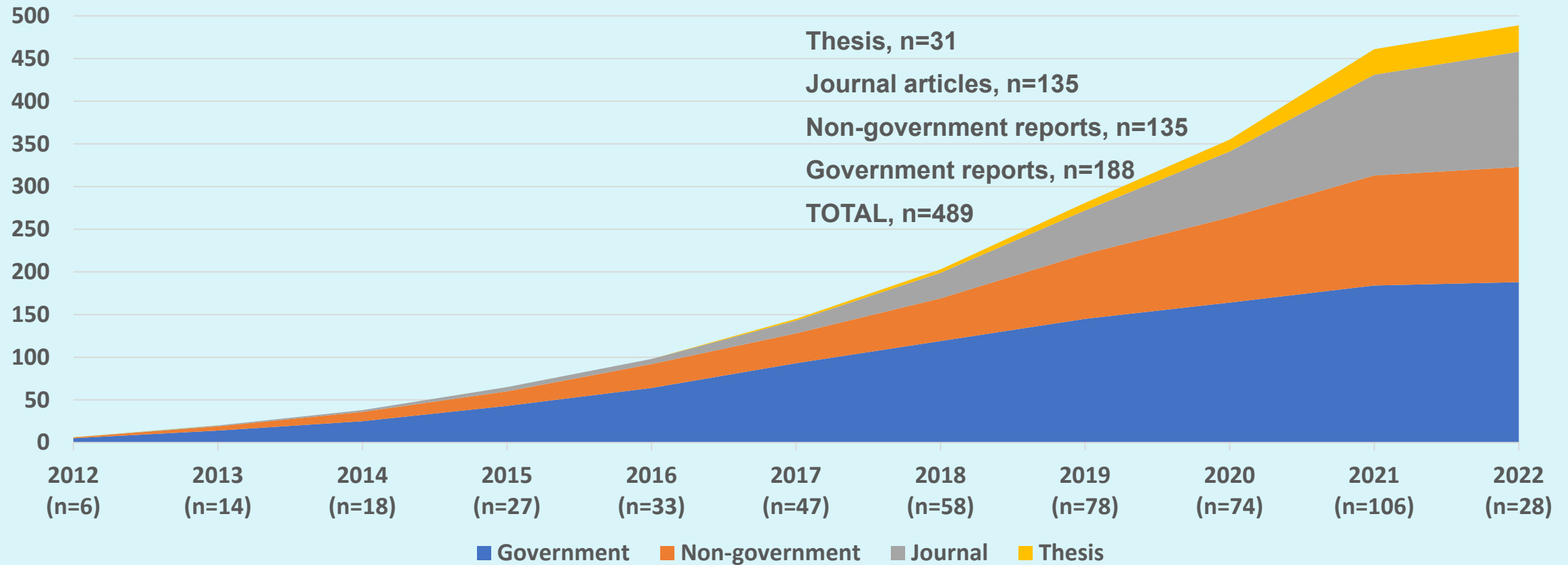
Info	Notes	Tags	Related
Item Type	Journal Article		
Title	Investigating the transmission risk of infectious disease outbreaks through the Aotearoa Co-incidence Network (ACN): a population-based study		
Author	Turnbull, S. M.		
Author	Hobbs, M.		
Author	Gray, L.		
Author	Harvey, E. P.		
Author	Scarrold, W. M. L.		
Author	O'Neale, D. R. J.		
Abstract			
Publication	The Lancet Regional Health-Western Pacific		
Volume	20		
Issue			
Pages	100351		
Date	2022		y
Series			
Series Title			
Series Text			
Journal Abbr			
Language			
DOI			
ISSN			
Short Title	Investigating the transmission risk of infectious disease outbreaks through the Aotearoa Co-incidence Network (ACN)		
URL	https://www.sciencedirect.com/science...		

<https://vhin.co.nz/research/>

# 2. What have we learned?

## ii. Publications increasing year on year

IDI Publications



## 2. What have we learned?

### ii. Publications increasing year on year

- Journal articles (n=135)
  - n=41 in 2021 and n=17 by April 2022 (one a week)
  - Most articles (n=111, 82%) published in international journals
  - Median impact factor of Journals = 2.7 (4.6 in 2020/21)
    - n=23 (17%) publications in Journals with impact factors >5
    - n=7 (5%) publications in Journals with impact factors >10

## 2. What have we learned?

### iii. Few reports taking a ‘social investment approach’

- Only 10/107 reports 2015-17 have a social investment focus
- Characteristics of Children at Greater Risk of Poor Outcomes as Adults. **The Treasury (2016)**
- Using Integrated Administrative Data to Understand Children at Risk of Poor Outcomes as Young Adults. **The Treasury (2015)**
- The Employment and Income Effects of Eight Chronic and Acute Health Conditions. **The Treasury (2016)**
- School to Work: What Matters? Education and Employment of Young People Born in 1991. **Ministry of Education (2016)**
- What Happened to People Who Left the Benefit System During the Year Ended 30 June 2016. **Ministry of Social Development (2015)**
- Using Integrated Administrative Data to Identify Youth Who Are at Risk of Poor Outcomes as Adults. **The Treasury (2015)**
- Economic Outcomes of Youth Not in Education, Employment or Training (NEET). **The Treasury (2016)**
- Using IDI Data to Estimate Fiscal Impacts of Better Social Sector Performance. **The Treasury (2015)**
- Research Using Administrative Data to Support the Work of the Expert Panel on Modernising Child, Youth and Family. **The Treasury (2017)**
- The Impact of Tertiary Study on the Labour Market Outcomes of Low-Qualified School Leavers. **The Treasury (2015)**
- **Caveat: I may have missed/misclassified some; results of some may not have been published**

## 2. What have we learned?

### iv. Publications cover diverse areas

- Labour market (Motu, MBIE, AUT, Victoria)
- Productivity (Motu, MBIE, Productivity Commission)
- Education (MOE, Motu, NZ Initiative)
- Health (Universities)
- Housing (Otago, MBIE, SWA)
- Wellbeing (Motu, Treasury, SWA, Universities)
- Deprivation (Auckland)
- Covid (Te Punaha Matatini)
- Methods (Stats NZ, Others)

# Diversity of IDI investigations: COVID

- ❑ 'Aotearoa Coincidence Network'
- ❑ Transmission risk estimated for NZ population using data on 'connections' between individuals (employed at the same workplaces or attend the same schools)
- ❑ Health vulnerability estimated for each SA2 by number elderly or with long term conditions

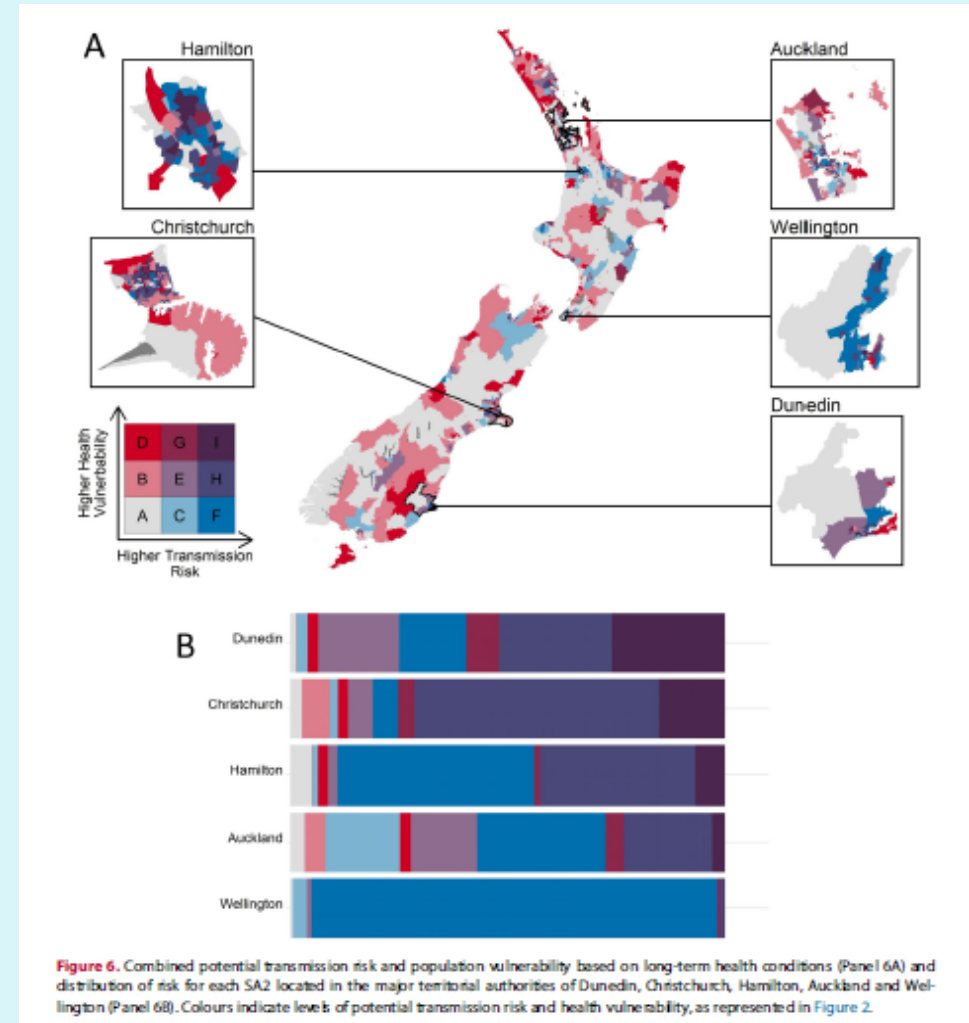


Figure 6. Combined potential transmission risk and population vulnerability based on long-term health conditions (Panel 6A) and distribution of risk for each SA2 located in the major territorial authorities of Dunedin, Christchurch, Hamilton, Auckland and Wellington (Panel 6B). Colours indicate levels of potential transmission risk and health vulnerability, as represented in Figure 2.

# Diversity of IDI investigations: Vegetation diversity and asthma

Born 1998 (Birth records)



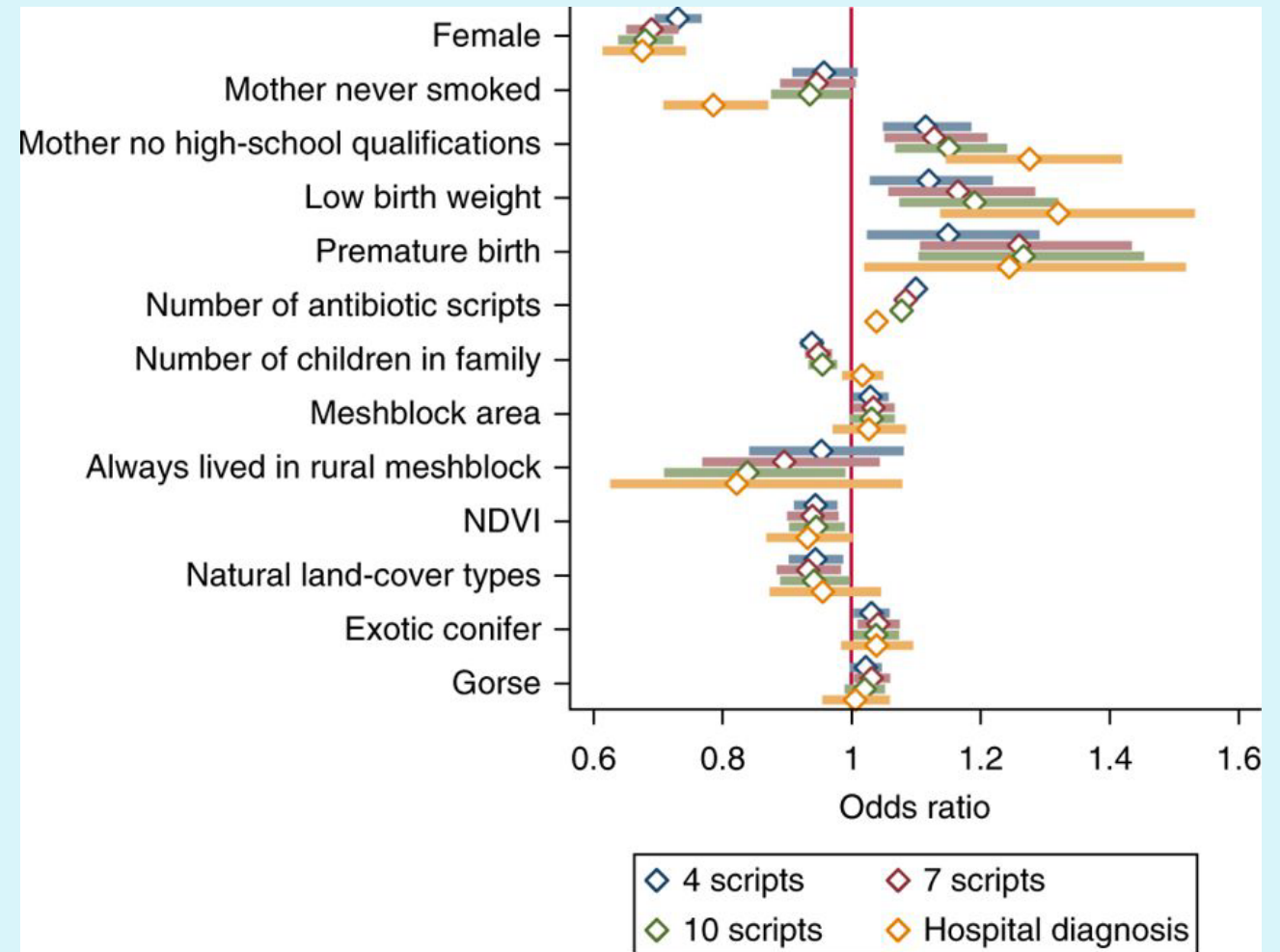
Area of residence 1998-2016  
(Address table)



Area-level greenness index  
(NDVI: Normalized difference  
vegetation index) 1998-2016  
(satellite imagery)



Asthma 2005-2016  
(Hospitalization and  
pharmaceutical data)



Donovan et al. (2018). Vegetation diversity protects against childhood asthma: results from a large New Zealand birth cohort. *Nat Plants*;4(6):358-364.

# Diversity of IDI investigations: Outcomes of very preterm birth



- Most children born very preterm (23-24w) able to be resuscitated and 66% survived to age 10
- Most able to participate in and perform well at school
- Gradient of school achievement across gestational age, right up to early term (37-38w)

**BACKGROUND AND OBJECTIVES:** As outcomes for extremely premature infants improve, up-to-date, large-scale studies are needed to provide accurate, contemporary information for clinicians, families, and policy makers. We used nationwide New Zealand data to explore the impact of gestational age on health and educational outcomes through to adolescence.

**METHODS:** We performed a retrospective cohort study of all births in New Zealand appearing in 2 independent national data sets at 23 weeks' gestation or more. We report on 2 separate cohorts: cohort 1, born January 1, 2005 to December 31, 2015 (613 521 individuals), used to study survival and midterm health and educational outcomes; and cohort 2, born January 1, 1998 to December 31, 2000, and surviving to age 15 years (146 169 individuals), used to study high school educational outcomes. Outcomes described by gestational age include survival, hospitalization rates, national well-being assessment outcomes at age 4 years, rates of special education support needs in primary school, and national high school examination results.

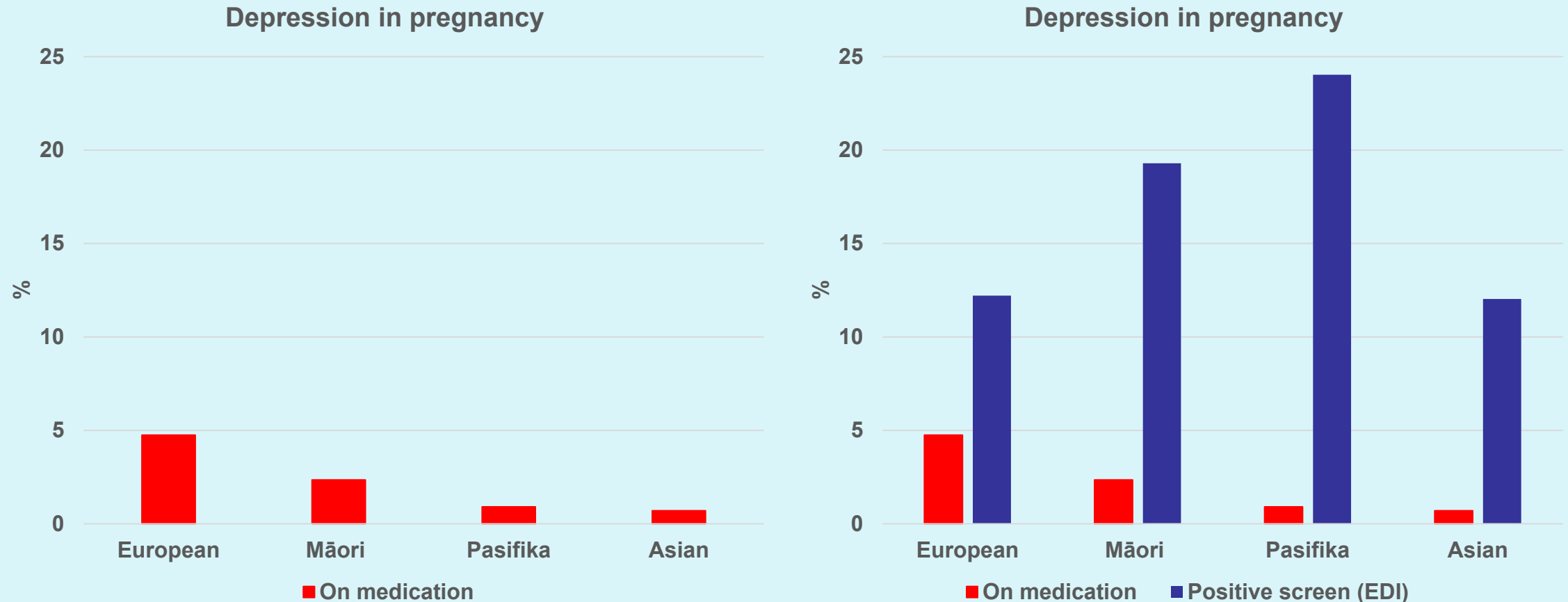
**RESULTS:** Ten-year survival increased with gestational age from 66% at 23 to 24 weeks to >99% at term. All outcomes measured were strongly related to gestational age. However, most extremely preterm children did not require special educational support and were able to sit for their national high school examinations.

**CONCLUSIONS:** Within a publicly funded health system, high-quality survival is achievable for most infants born at periviable gestations. Outcomes show improvement with gestational ages to term. Outcomes at early-term gestation are poorer than for children born at full term.



# 2. What have we learned?

## vi. Comparing IDI results to ‘collected’ data can be illuminating



Modified from Svardal et al., 2021; <https://pubmed.ncbi.nlm.nih.gov/34260316/>

New Zealand

The University of Auckland

# 2. What have we learned?

## Strengths

- Whole population
  - Transmission risk, Turnbull et al., 2022
- Able to study small and difficult to reach populations
  - pre-term babies, Berry et al., 2018; refugee children, Charania et al., 2018; 2020
- All life stages able to be studied
- Cross-domain exposures and outcomes
  - Berry et al., 2018
  - Donovan et al., 2018

## Weaknesses

- Primarily service use data
  - Will not include problems/issues for which service has not been sought
  - May be biases in who receives services
  - Interpret with care, Svardal et al., 2021
- Primarily deficit focussed
- Limited follow-up for many datasets
- Variable data quality and documentation
- Difficult to determine family and household structures longitudinally
- Missed and incorrect links

# 3. What are the issues?

## i. Social Licence

“permission to make decisions about management and use of the public’s data”

- Kalkman et al. 2019; Paprica et al. 2019: Public acceptance for secondary data use when there is transparency about how the data are being used, and there is trust in the institutions
- Do people know how their data are being used?
  - Unlikely... Nielsen survey found that 40% of adults didn't know enough about what StatsNZ does (generally) to know whether or not to trust them

# 3. What are the issues?

## ii. Ethics

- Ethical governance of IDI projects is ‘uneven’, and depends on
  - The researcher’s understanding of their obligations and the researcher’s employer/sector
    - University researchers:  
No-one forcing them, but ethical governance systems in place
    - Government researchers:  
Might have systems in place and might be required to use them
    - Other researchers:  
?? May not have ethical governance systems in place ??

# 3. What are the issues?

## ii. Ethics

- Informed consent often lacking
  - Individuals may have signed a form indicating their data may be used for ‘research or scientific purposes’, but...
    - Often under duress (need to use a service)
    - Unclear what public understand by ‘research or scientific purposes’
  - Lessof (2009): the important question regarding use of administrative data without informed consent is “whether an individual’s health, interests or confidentiality could be affected negatively”
    - Use of IDI data unlikely to affect health, and confidentiality is guarded (‘Five Safes’) ... but... ethical assessment by diverse panel would help assess whether individuals’ interests could be affected negatively

# 3. What are the issues?

## iii. Māori data sovereignty

“Māori data should be subject to Māori governance”

- ‘Big data divide’ – those represented in the data often not those using and benefitting from data use
- Data structured to respond to government priorities, NOT to benefit Indigenous nations (Rowe et al., 2021)
  - Ongoing community-level consent crucial
- Ngā Tikanga Paihere framework goes some way to address sovereignty concerns

# 3. What are the issues?

## iii. Māori data sovereignty

“Māori data should be subject to Māori governance”

- Other frameworks go further
  - For datasets where indigenous populations are over-represented, Manitoba Multigenerational Cohort requires approval from the First Nations Health and Social Secretariat of Manitoba and the Manitoba Metis Community Research and Ethics Protocol, in addition to an ethical assessment (Hamad et al. 2021)
- Is the current governance arrangement of the IDI the best way to recognize Māori data sovereignty??

# 3. What are the issues?

## iv. Inequality of access

- High bar to access IDI (Five Safes), so many communities who may benefit from IDI data not able to access it ('big data divide'...)
  - Covid lockdowns highlighted inequality (unable to access 'Safe Setting')
  - Most uses of the IDI are simple ones (counts: "how many people are 'X'?", Ellis, 2017).
  - So... is a technical solution possible that allows broader access to data, relaxing 'Safe People' & 'Safe Setting' (e.g., an extension of NZ.Stat)??
- Paprica et al., 2019: The more the public knows about the data the greater the trust and support (social licence!) for the data-use project
  - What better way to learn about the data than to use it and see its benefits! In so doing...
  - Communities' views and aspirations better reflected in the research about them



- ❑ As an academic researcher, the IDI is a fantastic research tool
  - ❑ The IDI allows research that is unable to be done any other way
- ❑ As a New Zealander, I want the public to know more about the IDI
  - ❑ to discuss its benefits and risks
  - ❑ to have a say in the research that is being undertaken
  - ❑ and to (perhaps...) participate in the research



# QUESTIONS?

# References (1/2)

- Ball C, Crichton S, Templeton R, Tumen S, Ota R, MacCormick C. 2016. Characteristics of Children at Greater Risk of Poor Outcomes as Adults. Wellington, New Zealand: The Treasury
- Berry MJ, Foster T, Rowe K, Robertson O, Robson B, Pierse N. 2018. Gestational Age, Health, and Educational Outcomes in Adolescents. *Pediatrics*. 142(5):e20181016
- Charania NA, Paynter J, Lee AC, Watson DG, Turner NM. 2018. Exploring immunisation inequities among migrant and refugee children in New Zealand. *Human Vaccines & Immunotherapeutics*. 14(12):3026–33
- Charania NA, Paynter J, Lee AC, Watson DG, Turner NM. 2020. Vaccine-Preventable Disease-Associated Hospitalisations Among Migrant and Non-migrant Children in New Zealand. *J Immigrant Minority Health*. 22(2):223–31
- Crichton S, Templeton R, Tumen S. 2015. Using Integrated Administrative Data to Understand Children at Risk of Poor Outcomes as Young Adults. Wellington, New Zealand: The Treasury
- Dixon S. 2015. The Employment and Income Effects of Eight Chronic and Acute Health Conditions. Wellington, New Zealand: The Treasury
- Donovan GH, Gatzolis D, Longley I, Douwes J. 2018. Vegetation diversity protects against childhood asthma: results from a large New Zealand birth cohort. *Nature Plants*. 4(6):358–64
- Earle D. 2016. School to Work: What Matters? Education and Employment of Young People Born in 1991 (Pdf). Wellington, New Zealand: Ministry of Education (MoE)
- Ellis P (2017). Understanding current practice and systems for analysis of integrated data in the Stats NZ Data Lab. Statistics New Zealand, <https://vhin.co.nz/wp-content/uploads/2018/02/use-of-IDI.pdf>
- Hamad AF, Walld R, Lix LM, Urquia ML, Roos LL, Wall-Wieler E (2021). Data Resource Profile: The Manitoba Multigenerational Cohort. *International Journal of Epidemiology*. dyab195, doi: 10.1093/ije/dyab195.
- Judd E, Gibbs J. 2016. What Happened to People Who Left the Benefit System During the Year Ended 30 June 2016. Wellington, New Zealand: Ministry of Social Development
- Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G (2019). Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *J. Med. Ethics*. medethics-2019-105651

# References (2/2)

- Lessof C (2009). Ethical Issues in Longitudinal Surveys. In *Methodology of Longitudinal Surveys*, ed. P Lynn, pp. 35–54. Chichester, UK: John Wiley & Sons, Ltd
- McLeod K, Templeton R, Ball C, Tumen S, Crichton S. 2015. Using Integrated Administrative Data to Identify Youth Who Are at Risk of Poor Outcomes as Adults. Wellington, NZ: The Treasury
- Nielson (2020). Stats NZ's Social Licence for Data Stewardship. Wellington, New Zealand: Statistics New Zealand
- Paprica PA, de Melo MN, Schull MJ. (2019). Social licence and the general public's attitudes toward research based on linked administrative health data: A qualitative study. *CMAJ Open*, 7(1), E40–E46. <https://doi.org/10.9778/cmajo.20180099>
- Rowe RK, Carroll SR, Healy C, Rodriguez-Lonebear D, Walker JD (2021). The SEEDS of Indigenous population health data linkage. *Int J Popul Data Sci*. 6(1):1417, doi: 10.23889/ijpds.v6i1.1417
- Samoilenko A, Carter K. 2015. Economic Outcomes of Youth Not in Education, Employment or Training (NEET). Wellington, New Zealand: The Treasury
- Stats NZ (2020). Ngā Tikanga Paihere: a framework guiding ethical and culturally appropriate data use. Retrieved from [www.data.govt.nz](http://www.data.govt.nz).
- Svardal CA, Waldie K, Milne B, Morton SM, D'Souza S. 2021. Prevalence of antidepressant use and unmedicated depression in pregnant New Zealand women. *Aust N Z J Psychiatry*. 000486742110256
- Templeton R, Crichton S, Tumen S, Rea D, Ota R, Small D. 2016. Research Using Administrative Data to Support the Work of the Expert Panel on Modernising Child, Youth and Family. Wellington, New Zealand: The Treasury
- Templeton R. 2016. Using IDI Data to Estimate Fiscal Impacts of Better Social Sector Performance. Wellington, New Zealand: The Treasury
- Tumen S, Crichton S, Dixon S. 2015. The Impact of Tertiary Study on the Labour Market Outcomes of Low-Qualified School Leavers. Wellington, NZ: The Treasury
- Turnbull SM, Hobbs M, Gray L, Harvey EP, Scarrold WML, O'Neale DRJ. 2022. Investigating the transmission risk of infectious disease outbreaks through the Aotearoa Co-incidence Network (ACN): a population-based study. *The Lancet Regional Health-Western Pacific*. 20:100351