

COMPASS projects using IDI data

Barry Milne & COMPASS team



**COMPASS
RESEARCH CENTRE**

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

10th Research Colloquium
Statistics NZ, Conference Room
1 August 2016

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.





- ▣ Administrative Cohort Idea (2011)

- ▣ The IDI

- ▣ COMPASS Projects using IDI
 - ▣ Better Start National Science Challenge
 - ▣ Child poverty analyses
 - ▣ Other possibilities

Administrative Cohorts



NZ has a number of administrative datasets containing data on transactions with the health system:

- [Laboratory Claims Collection \(Labs\)](#)
- [Maternity and Newborn Collection \(MNIS\)](#)
- [Mortality Collection \(MORT\)](#)
- [National Booking Reporting System \(NBRS\)](#)
- [National Booking Reporting System Data Warehouse \(NBRS DW\)](#)
- [National Immunisation Register \(NIR\)](#)
- [National Minimum Dataset \(Hospital Events\) \(NMDS\)](#)
- [National Non-admitted Patient Collection \(NNPAC\)](#)
- [New Zealand Cancer Registry \(NZCR\)](#)
- [Pharmaceutical Collection \(Pharms\)](#)
- [Primary Health Organisation Enrolment Collection \(PHO\)](#)
- [Programme for the Integration of Mental Health Data \(PRIMHD\)](#)

Administrative Cohorts

- ❑ Compiled as a series of “events”, stored more-or-less in the order they happen
- ❑ Creates “event-level” files which are continually growing in length as new events are added
- ❑ Can analyse events for certain sets of individuals if the individuals’ NHI numbers are known (NHI field on all collections)
- ❑ This may involve ‘reshaping’ the data so that there’s one row per individual (‘wide’ file), as opposed to one row per event (‘long’ file)

- If administrative datasets were combined and reshaped into ‘wide’ files, this would comprise a huge (nearly national) cohort, which could be followed over time
- This could be subsetted into any number of cohorts, depending on the needs of the researcher
 - Birth cohorts (complete!)
 - Cohorts of a certain age (or ages)
 - Cohorts sharing certain demographic characteristics.
 - Cohorts sharing certain medical conditions (e.g., cancer cohort).
 - Other cohorts.

Example: Cohort of all births in NZ public hospitals, NMDS, Jan 1-7, 2002 (N=900)



COMPASS
RESEARCH CENTRE

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

Obs	dob	dod	SEX	BTH_WT2	evstdate1	Diag1	LOS1	_2002	_2003	_2004
1	x	.	F	3560	x	Z380	1	0	0	0
2	x	.	M	3220	x	P073	3	0	0	2
3	x	.	M	3765	x	Z380	0	1	0	0
4	x	.	F	2940	x	Z380	0	0	0	0
5	x	.	M	3530	x	Z380	1	1	0	0
6	x	.	M	3910	x	Z380	0	0	0	0
7	x	.	M	3320	x	P599	6	1	0	0
8	x	.	M	3355	x	Z380	5	0	0	0
9	x	.	M	2665	x	Z380	1	0	0	0
10	x	.	M	2930	x	Z380	5	0	0	0

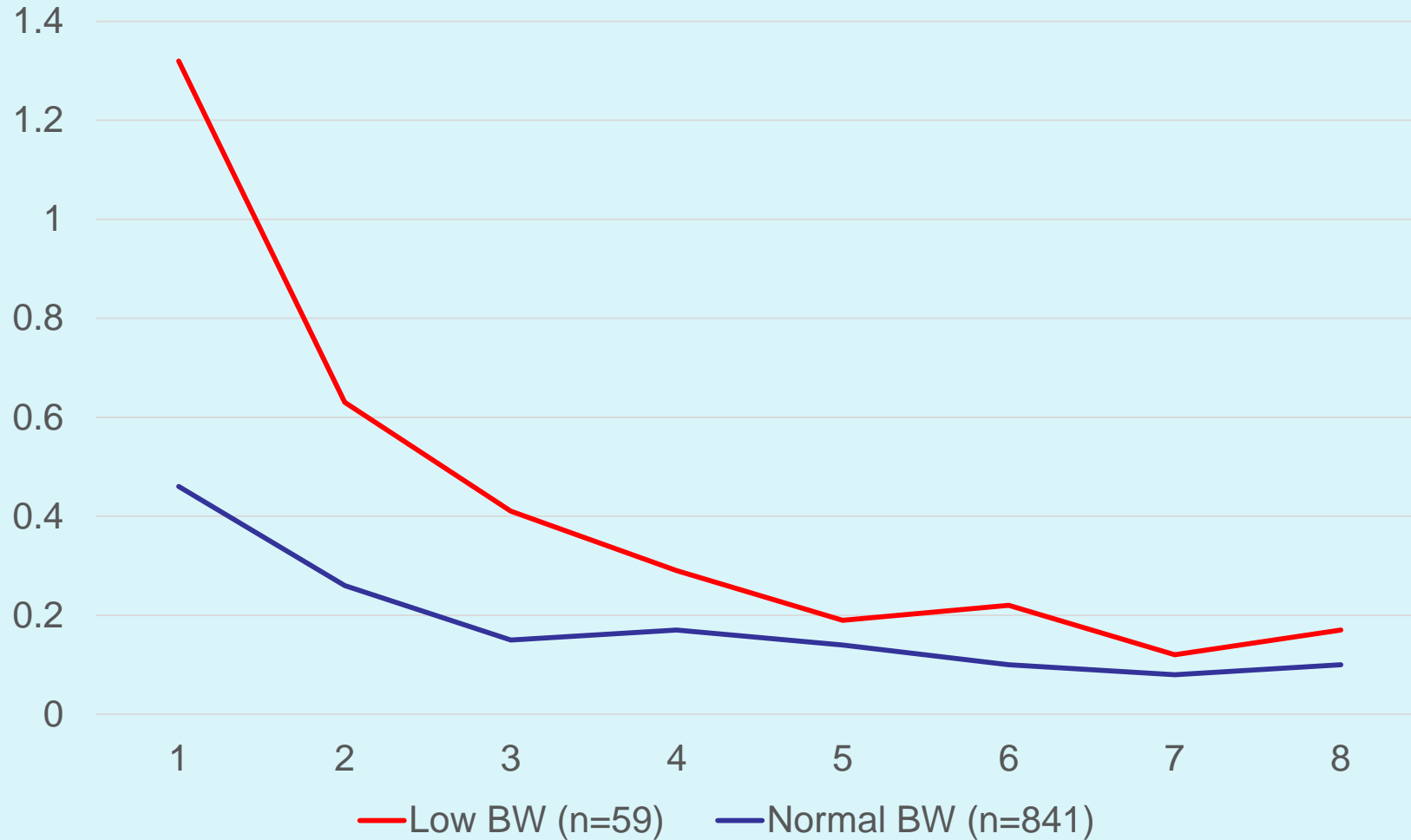
Example: Cohort of all births in NZ public hospitals, NMDS, Jan 1-7, 2002 (N=900)



Basic descriptives		
Sex (% male)		51.1
Ethnicity (%)	NZ European	58.3
	NZ Māori	21.6
	Pasific	10.1
	Asian	5.7
	Other	4.3
NZ Dep quintiles (%)	1	14.5
	2	16.0
	3	17.6
	4	23.4
	5	28.5
Birthweight in grams (mean [sd])		3380 [619]

Example: Cohort of all births in NZ public hospitals, NMDS, Jan 1-7, 2002 (N=900)

Hospitalisations, ages 1-8, by birthweight



Example: Cohort of all births in NZ public hospitals, NMDS, Jan 1-7, 2002 (N=900)

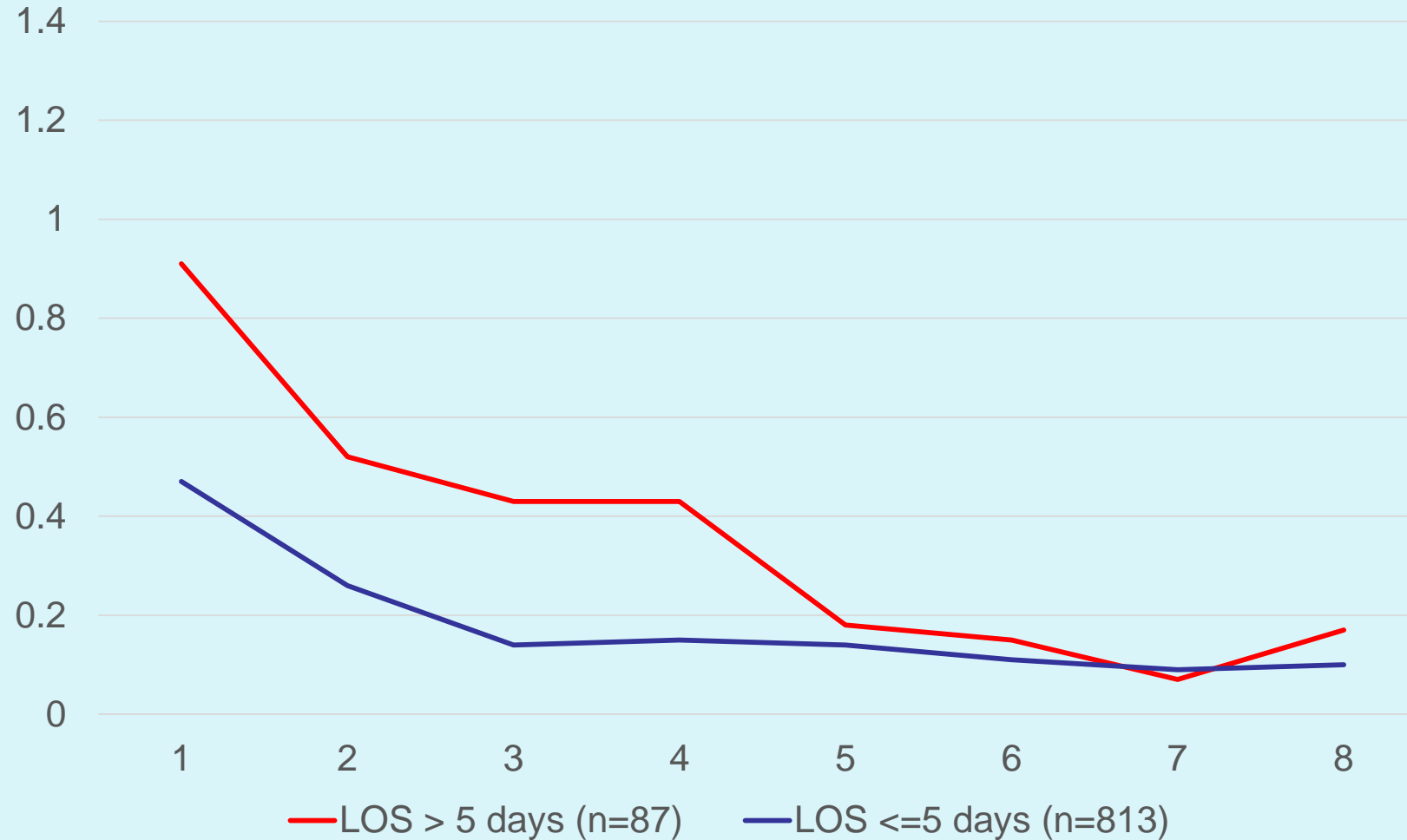


COMPASS
RESEARCH CENTRE

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

Hospitalisations, ages 1-8, by birth event LOS

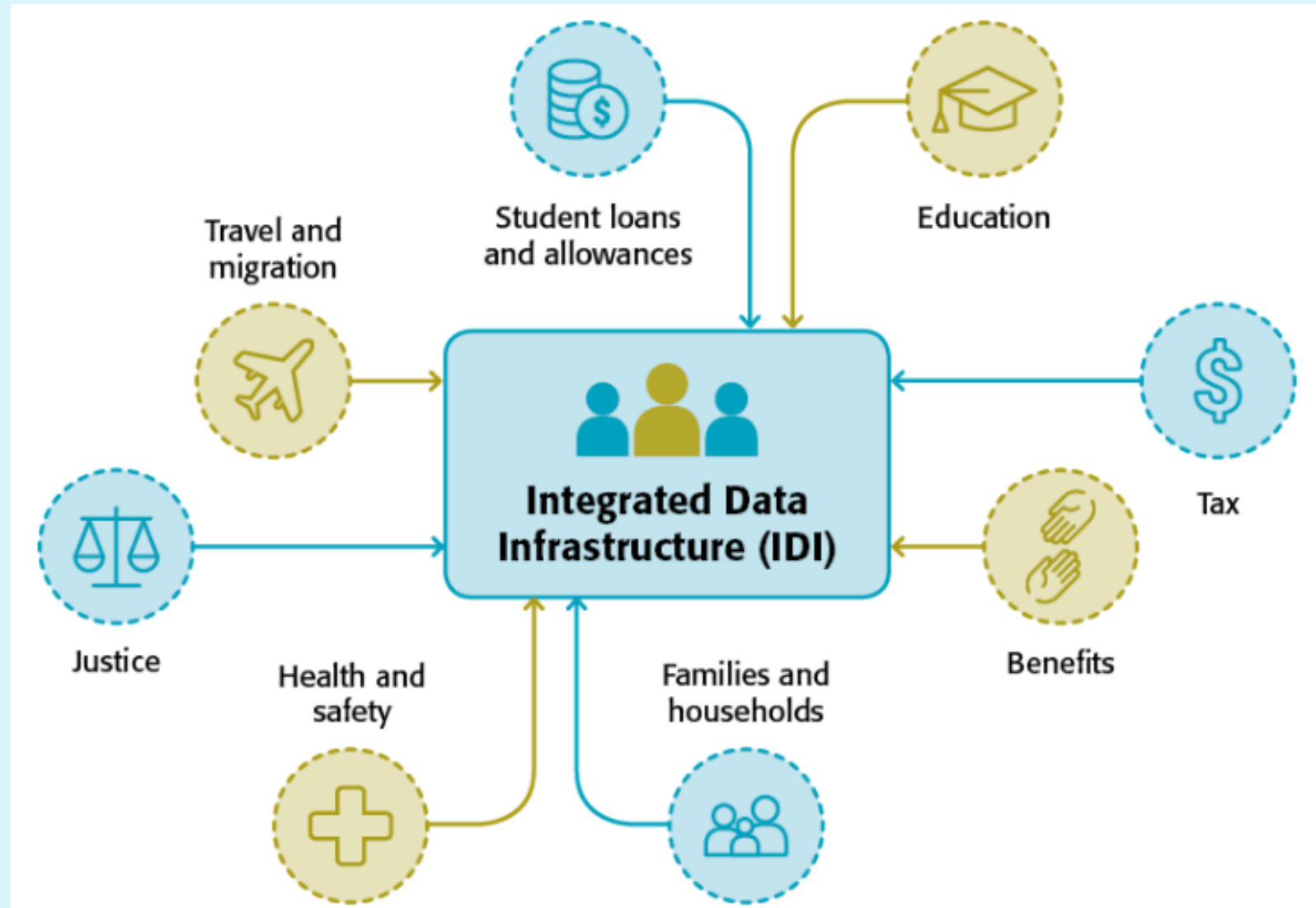


- ❑ There are lots of possibilities for cohort analysis just with hospitalisation data
 - ❑ Let alone all health data...
 - ❑ What if we had ALL NZ admin data sources linked together at the person level...
- ❑ Integrated Data Infrastructure (IDI)!!!
- ❑ IDI can be viewed as a gigantic cohort (or series of cohorts), with enormous breadth
 - ❑ And that's just one use!



Integrated Data Infrastructure

The Integrated Data Infrastructure (IDI) is a large research database containing microdata about people and households. Data is from a range of government agencies, Statistics NZ surveys including the 2013 Census, and non-government organisations. Researchers use the IDI to answer complex questions to improve outcomes for New Zealanders.





COMPASS IDI PROJECTS

Virtual Health Information Network (VHIN)

- ❑ Network of NZ health researchers, with an interest in use and analysis of (whole) population linked health data
 - Create and sustain an environment that captures value from linking health data and related social and economic data, through world leading health research
 - Tony Blakely (UO), Jereon Douwes (Massey), Simon Ross & Sheree Gibb (MOH), Barry Milne & Andrew Sporle (UOA)
 - Catalyst projects
 - Getting the denominator right (UOA: Jackson, Mehta, Exeter)
 - Cost of CVD in New Zealand (UO: Blakely)
 - Occupational and Pharmaceutical risk factors for Congenital Malformations Host (Massey: Mannelje, Borman, Eng, Douwes)
 - <https://www.facebook.com/groups/1658066281071881/>

1. Better Start National Science Challenge

- ❑ The Better Start Challenge focuses on three critical areas of childhood development that have been linked to life course outcomes.
- ❑ These are:
 - ❑ Obesity
 - ❑ Literacy
 - ❑ Mental Health

- ❑ Will also investigate Autism Spectrum Disorder (ASD)

Big Data's Contribution - COMPASS

- ❑ Aim of the Big Data group (incl COMPASS) is to provide Challenge level support to each of the research streams.

- ❑ Synergies in centralising this function – particularly around the IDI
 - ❑ Economies of scale and scope (data access processes and protocols; analytical techniques)

- ❑ Barry Taylor, Rick Audas (UO)
- ❑ Barry Milne (COMPASS, UOA)

What data will we use?



COMPASS
RESEARCH CENTRE

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

- ❑ Birth Record data
- ❑ B4 School Check
- ❑ NMDS
- ❑ Census and other socio-economic data

- ❑ Record created at birth.
- ❑ Includes data on birthweight and gestational age.
- ❑ Also links to mother's characteristics and father's characteristics.

B4 School Check

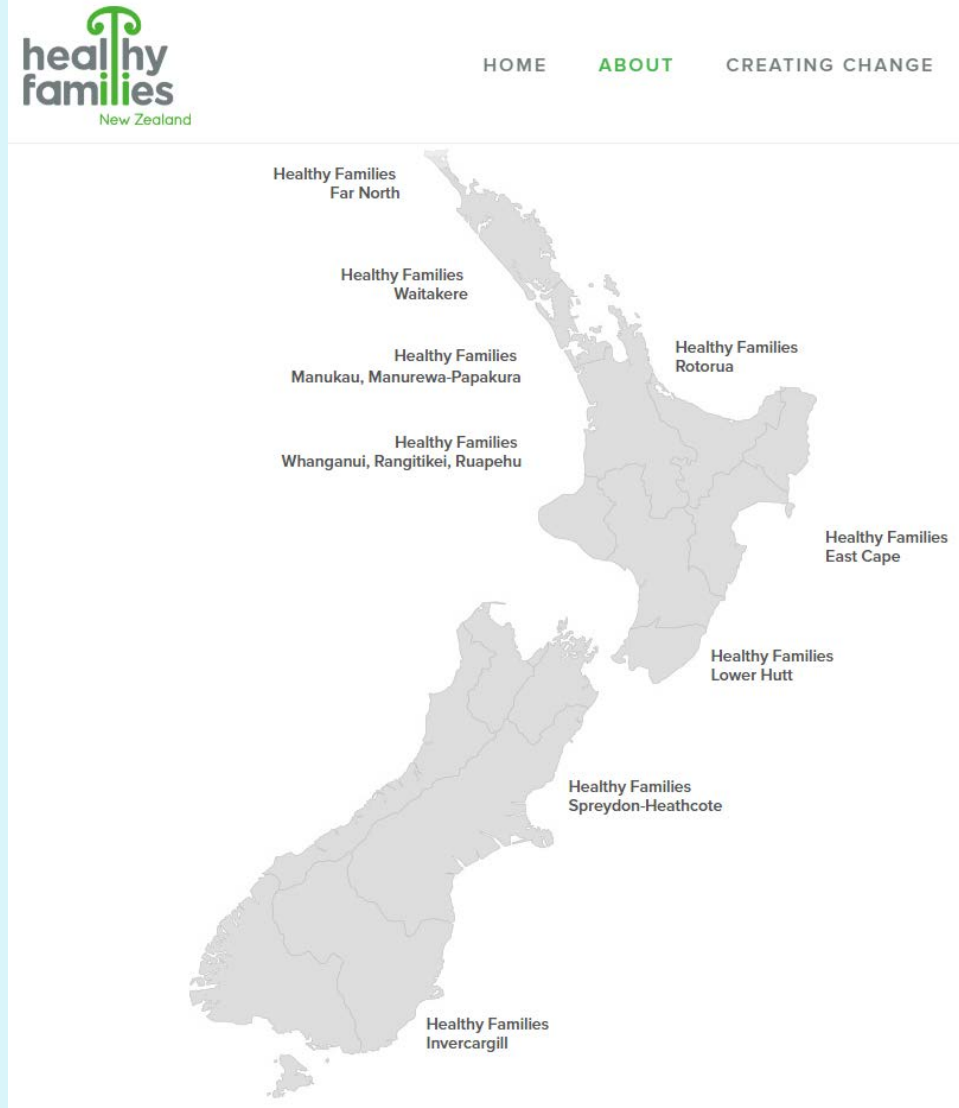
- ❑ Checks conducted before children enter school.
- ❑ Covers a range of key health and developmental indicators.
- ❑ Critically for us, BMI is captured, as are indicators of behavioural/self control issues.
- ❑ Has been collected since 2009 and now there is 90%+ coverage.

- ❑ Hospital records.
- ❑ Mostly useful for health trajectories. Can estimate predictors and counts of admissions.
- ❑ May also use ED data
- ❑ Sadly no primary care data

- ❑ ACC – allows for the identification (and count) of injury claims.
- ❑ School data (NCEA achievement, but also discipline and interventions)
- ❑ Surveys – the possibility of linked survey/admin data
- ❑ Pharmaceutical – may help us identify the incidence and prevalence of some conditions
- ❑ PRIMHD – Integrated MH data.
- ❑ Socrates – Disabilities, incl. data on ASD

- ❑ Only two systematic measures of weight and height over the lifespan (birth and B4 School).
- ❑ Would be nice if there was another at age ten and again at 18 But there isn't.
- ❑ Broad theme will be to look at the long-term sequelae of EC obesity.

- ❑ First task: track obesity over time across region
 - ❑ which areas are showing good/poor trends, controlling for population differences (SES, ethnicity)?
 - ❑ What are different areas doing differently?



- Locally driven health promotion, tailored to the community and delivered where people live, learn, work and play, can be successful in addressing the underlying causes of preventable diseases.
- The Healthy Families NZ teams are working with locally-based organisations to identify, design and implement changes to help people make healthier choices and live healthier lives.



- ❑ Not much in IDI.
- ❑ Will examine school achievement data but may need to look outside IDI for population assessments of literacy

- ❑ First indicators at B4 School check.
- ❑ Hopefully can pick up cases through PRIMHD, NMDS, ED and possibly Pharms and Lab data.
- ❑ Will tend to under-report less serious cases which are very often undiagnosed and untreated, but possible biases in those seeking treatment.
 - ❑ Lends itself to methods work, comparing survey data to admin records

2. Child poverty Analyses - Potential Questions



- ❑ Who becomes trapped in childhood poverty?
- ❑ What are the health effects and health costs of childhood poverty?
- ❑ Do different types of deprivation have different effects on outcomes?
- ❑ What is the duration of poverty experience for children, and how does the timing and persistence impact later outcomes?
- ❑ What is the effect of lifting children from poverty?
- ❑ To what extent do different factors mediate the association between child poverty and health, and what are the likely health benefits and cost benefits of intervening on these factors?

Data sources - Census

- ❑ Census classifies (and assigns IDs for):
 - Individuals
 - Within Families
 - Within Households
 - Within Dwellings

- ❑ This structure allows
 - ❑ Construction of family/household indicators, such as household poverty (income poverty measure only)

Household poverty

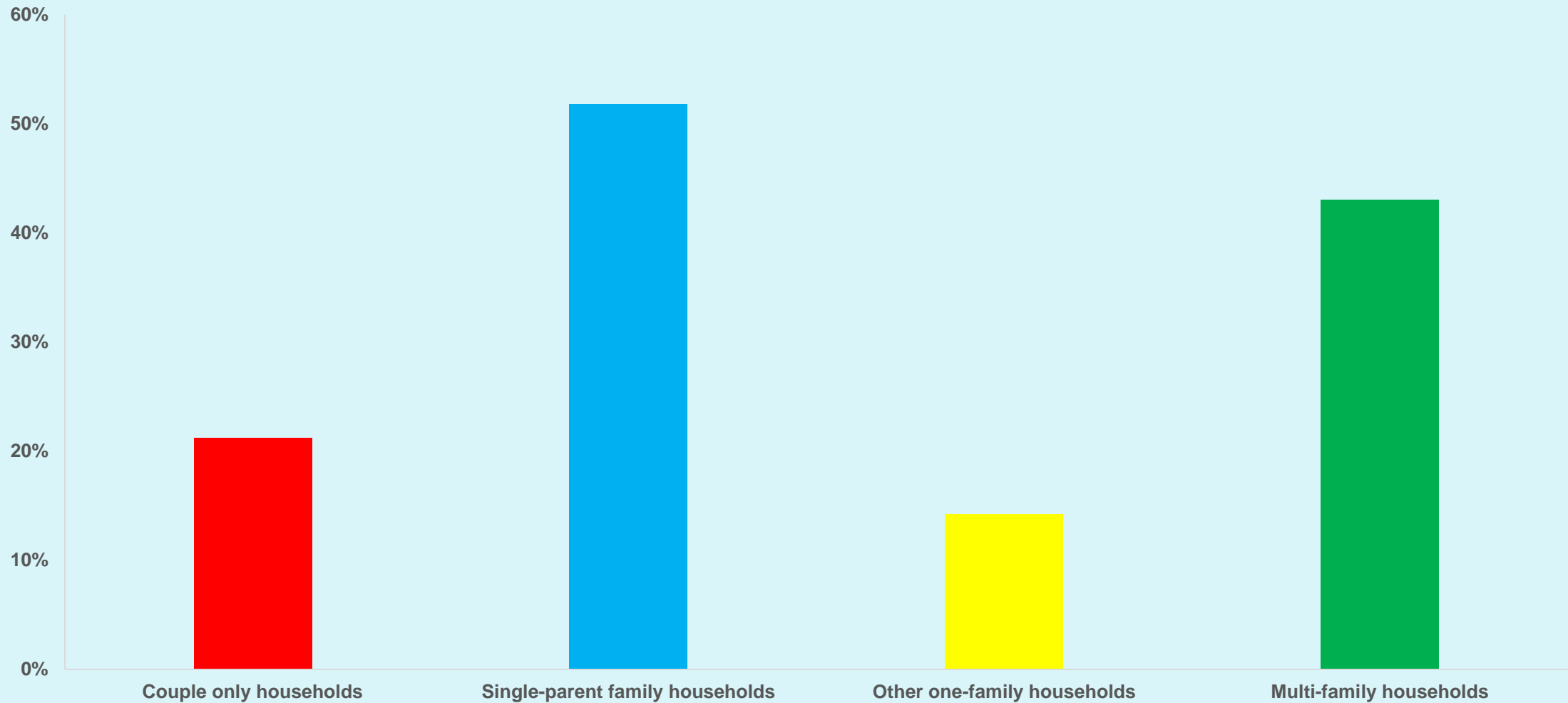


COMPASS
RESEARCH CENTRE

FACULTY OF ARTS
THE UNIVERSITY OF AUCKLAND

Whare Wānanga o Tāmaki Makaurau

Households in poverty (%)



Data sources - Surveys

- ❑ Household Economic Survey
 - ❑ Very detailed characterisation of both income and deprivation poverty

- ❑ Survey of Families, Income & Employment (SOFIE)
 - ❑ Longitudinal data on both income and deprivation poverty (so can track movements in and out of poverty)

- ❑ Children in surveys CAN be linked to other IDI datasets

Data sources - Outcomes & Mediators

▣ Outcomes

- ▣ Health: Hospitalisations, pharmaceuticals, B4School Check
- ▣ Education: NCEA attainment
- ▣ Justice: Convictions
- ▣ MSD: benefit receipt
- ▣ IRD: Earnings

▣ Mediators

- ▣ Surveys: Crowding, Rent affordability, parental substance use (SOFIE poss better)
- ▣ Other: B4School check, ECE

3. Other possibilities: Better measures of income in surveys

- ❑ Comparison of income recorded by IRD and self-reported income in the Census
 - ❑ 1) Assess discrepancy between recorded (IRD) and self-reported income for those who do report their income on the census
 - Does this vary across the income distribution? By social group?
 - ❑ 2) For those who did not report income in the census
 - Test several imputation models for income
 - Compare imputed income with observed income(IRD)



4. Other possibilities: Intergenerational Analyses



Psychiatric family history and schizophrenia risk in Denmark: which mental disorders are relevant?

P. B. Mortensen*, M. G. Pedersen and C. B. Pedersen

Criminal conviction among offspring with parental history of mental disorder

K. Dean^{1*}, P. B. Mortensen², H. Stevens², R. M. Murray¹, E. Walsh¹ and E. Agerbo²

Higher Risk of Offspring Schizophrenia Following Antenatal Maternal Exposure to Severe Adverse Life Events

Ali S. Khashan, MSc; Kathryn M. Abel, MRCP, MRCPsych, PhD; Roseanne McNamee, PhD; Marianne G. Pedersen, MSc; Roger T. Webb, PhD; Philip N. Baker, DM, FRCOG; Louise C. Kenny, PhD, MRCOG; Preben Bo Mortensen, MD, DMSc

LETTERS TO THE EDITOR

The importance of father's age to schizophrenia risk

- ❑ Analyses straightforward with population registries (e.g., Denmark, Sweden)

- ❑ ... and possible in New Zealand with linked administrative and Census & DIA data sources
 - ❑ Census & DIA data defines the family connections
 - ❑ Health data collections identify disease states for family members
 - ❑ Other (social, justice, education) data collections can identify other outcomes and experiences

- ❑ COMPASS is very excited by possibilities with IDI data
- ❑ Currently working on Better Start project
- ❑ Other possibilities and ideas
 - ❑ Poverty; Income measurement; Intergenerational analyses