Childhood Development: Population Research Panel Discussion

Data Linkage Research Conversation Series 2013

Professor John Lynch
Professor Sven Silburn
Dr Sally Brinkman
Dr Steve Guthridge
Mr Sam Luddy

Moderated by:
Professor Annette Braunack-Mayer
Hosted by:
The University of Adelaide

Professor John Lynch

How can we give every child the best start in life?

Better information through data linkage

John Lynch

Professor of Public Health, University of Adelaide

NHMRC Australia Fellow

Investigators:

University of Adelaide:	Cathy Chittleborough, Lisa Smithers, Angela Gialamas, Daniel Scalzi Jesia Berry, Murthy Mittinty, Alyssa Sawyer, Loc Do, Michael Davies, Michael Sawyer
Flinders University:	Clare Bradley, James Harrison (Injury)
Telethon Inst CHR:	Sally Brinkman, Tess Gregory, Angela Kinnell (AEDI and NAPLAN)
WCHN:	Kerrie Bowering , Pauline McEntee (Child and Family Health Services - FHV, hospitalizations),
DECD:	David Englehart, Sam Luddy, Sandy Burton, Trish Strachan (AEDI and NAPLAN)
SA Health:	David Banham (Potentially Avoidable Admissions to hospital), Wendy Scheil (perinatal)

Australian Government
National Health and Medical Research Council
NHMRC

THE UNIVERSITY OF ADELAIDE
AUSTRALIA

Australian Government

Productivity Commission

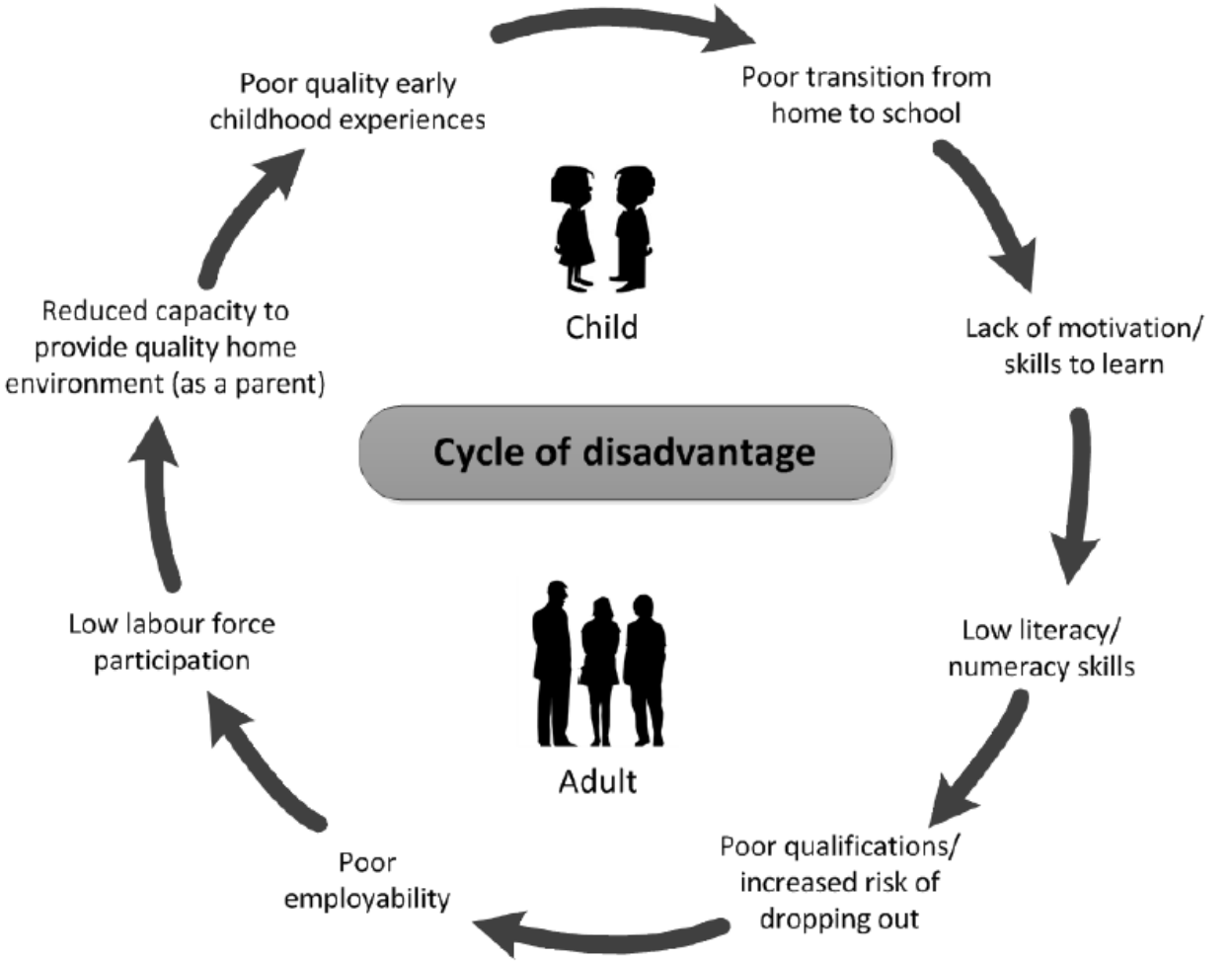
Deep and Persistent Disadvantage in Australia

Productivity Commission Staff Working Paper

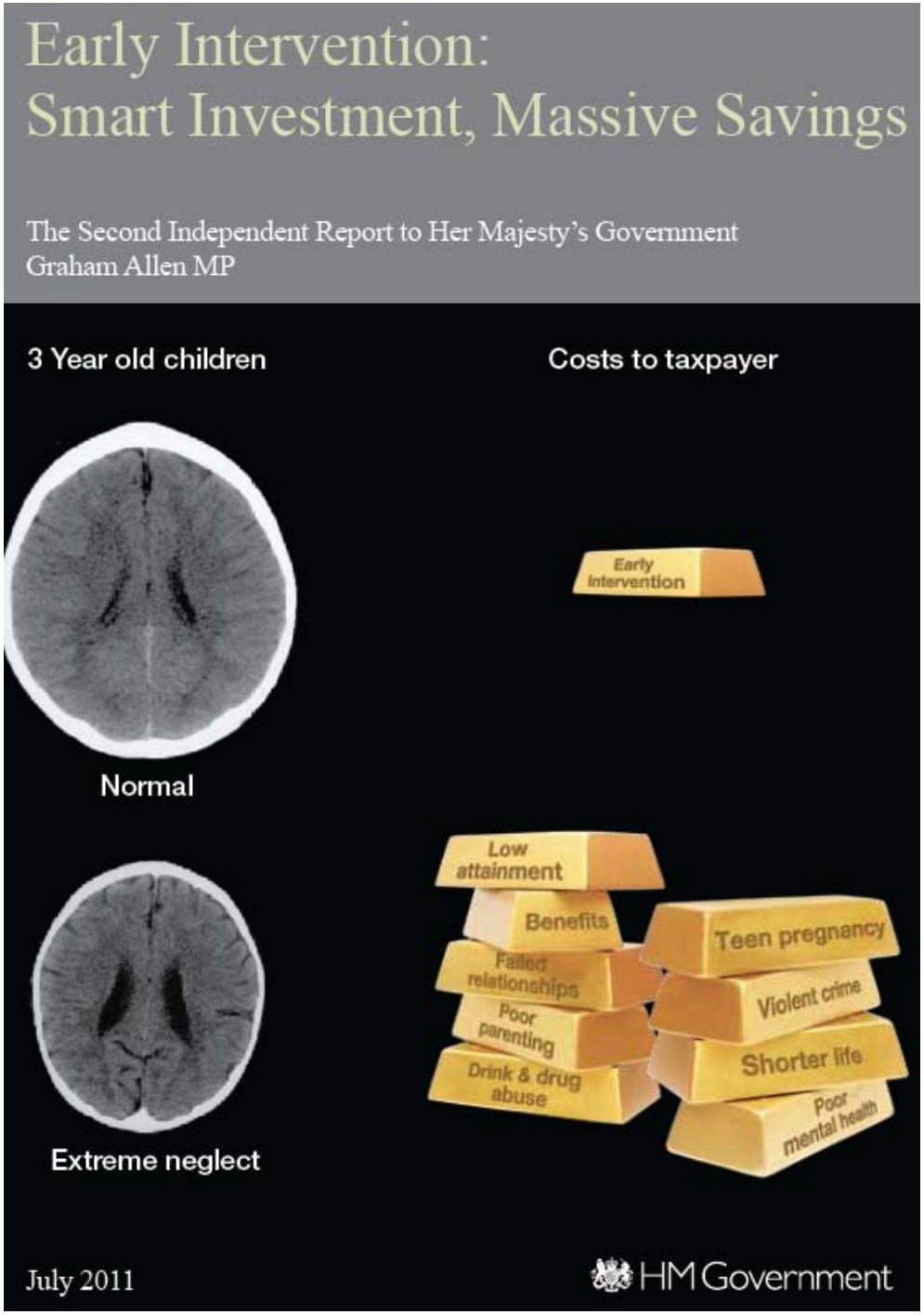
July 2013

THE UNIVERSITY OF ADELAIDE AUSTRALIA

Figure 4.3 **The cycle of disadvantage can start early in life**



Source: The Smith Family (2010, p. 5).



"My first Report detailed the immense penalties to society and to the individual of failing to provide a strong foundation of social and emotional capabilities early in life.

This second Report focuses more on addressing the vast financial and economic costs."

Letter to the Prime Minister, David Cameron, July 2011

THE UNIVERSITY OF ADELAIDE AUSTRALIA

PERSPECTIVE

Skill Formation and the Economics of Investing in Disadvantaged Children

James J. Heckman

This paper summarizes evidence on the effects of early environments on child, adolescent, and adult achievement. Life cycle skill formation is a dynamic process in which early inputs strongly affect the productivity of later inputs.

Four core concepts important to devising sound social policy toward early childhood have emerged from decades of independent research in economics, neuroscience, and developmental psychology (*1*). First, the architecture of the brain and the process of skill formation are influenced by an interaction between genetics and individual experience. Second, the mastery of skills that are essential for economic success and the development of their underlying neural pathways follow hierarchical rules. Later attainments build on foundations that are laid down earlier. Third, cognitive, linguistic, social, and emotional competencies are interdependent; all are shaped powerfully by the experiences of the developing child; and all contribute to success in the society at large. Fourth, although adaptation continues throughout life, human abilities are formed in a predictable sequence of sensitive periods, during which the development of specific neural circuits and the behaviors they mediate are most plastic and therefore optimally receptive to environmental influences.

A landmark study concluded that "virtually every aspect of early human development, from the brain's evolving circuitry to the child's capacity for empathy, is affected by the environments and experiences that are encountered in a cumulative fashion, beginning in the prenatal period and extending throughout the early childhood years" (*2*). This principle stems from two characteristics that are intrinsic to the nature of learning: (i) early learning confers value on acquired skills, which leads to self-reinforcing motivation to learn more, and (ii) early mastery of a range of cognitive, social, and emotional competencies makes learning at later ages more efficient and therefore easier and more likely to continue.

ing practices and lack of positive cognitive and noncognitive stimulation. A child who falls behind may never catch up. The track records for criminal rehabilitation, adult literacy, and public job training programs for disadvantaged young adults are remarkably poor (*3*). Disadvantaged early en

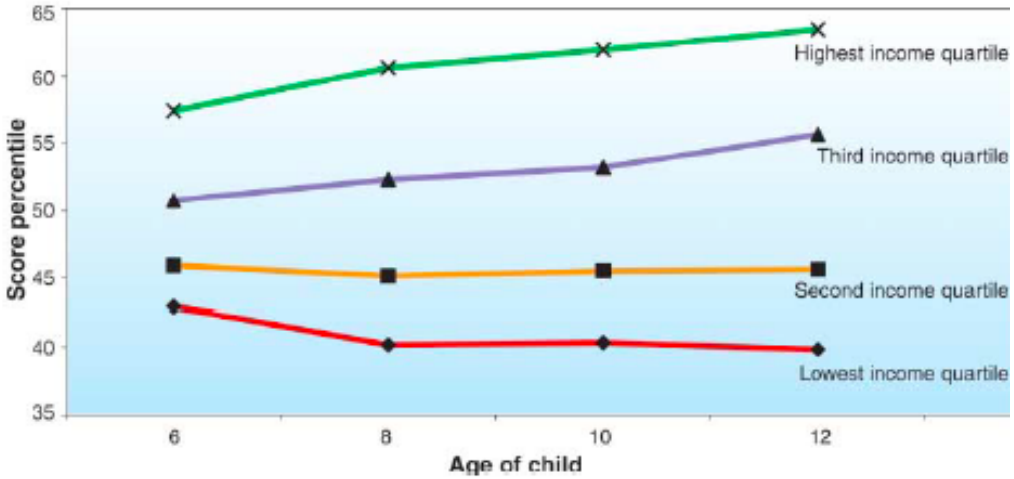


Fig. 1. Average percentile rank on Peabody Individual Achievement Test–Math score by age and income quartile. Income quartiles are computed from average family income between the ages of 6 and 10. Adapted from (*3*) with permission from MIT Press.

Department of Economics, University of Chicago, Chicago, IL 60637, USA. Department of Economics, University College Dublin, Dublin 4, Ireland. E-mail: jjh@uchicago.edu

THE UNIVERSITY OF ADELAIDE AUSTRALIA

DECD Strategic Plan 2012-2016

"The system will identify objectives, measures and indicators that are integrated across early childhood settings, schools, regions and the State. These will focus on what is making a difference for children and young people.

We will publish performance data and report on our achievements. We will research different approaches, consider available evidence, and promote the most effective practices.

Current investments will be evaluated and we will reinvest where appropriate."

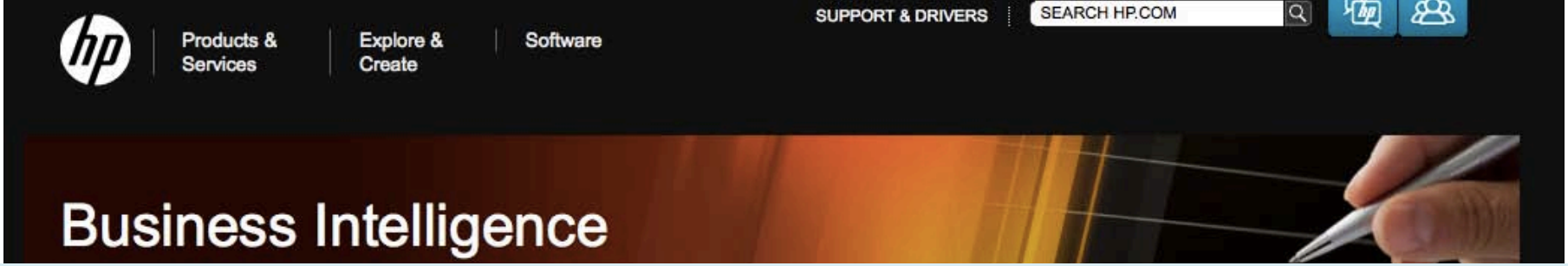
DECD expenditures 2010-2011: ~ $2.5 billion

THE UNIVERSITY OF ADELAIDE AUSTRALIA

What information systems will help secure these goals?

and

Where might data linkage fit?



Transforming data into valuable business assets

Business Intelligence solutions—through connected intelligence—is your answer to the world of disconnected information.

Connected intelligence ties data that was once scattered across your business ecosystem together in a way that allows knowledge and intelligence to flow effortlessly across your entire business ecosystem.

The solutions enable your enterprise to drive greater value from information by aligning your information strategy with your business objectives and respond to emerging opportunities

THE UNIVERSITY OF ADELAIDE AUSTRALIA

SA and NT Early Childhood Development Project

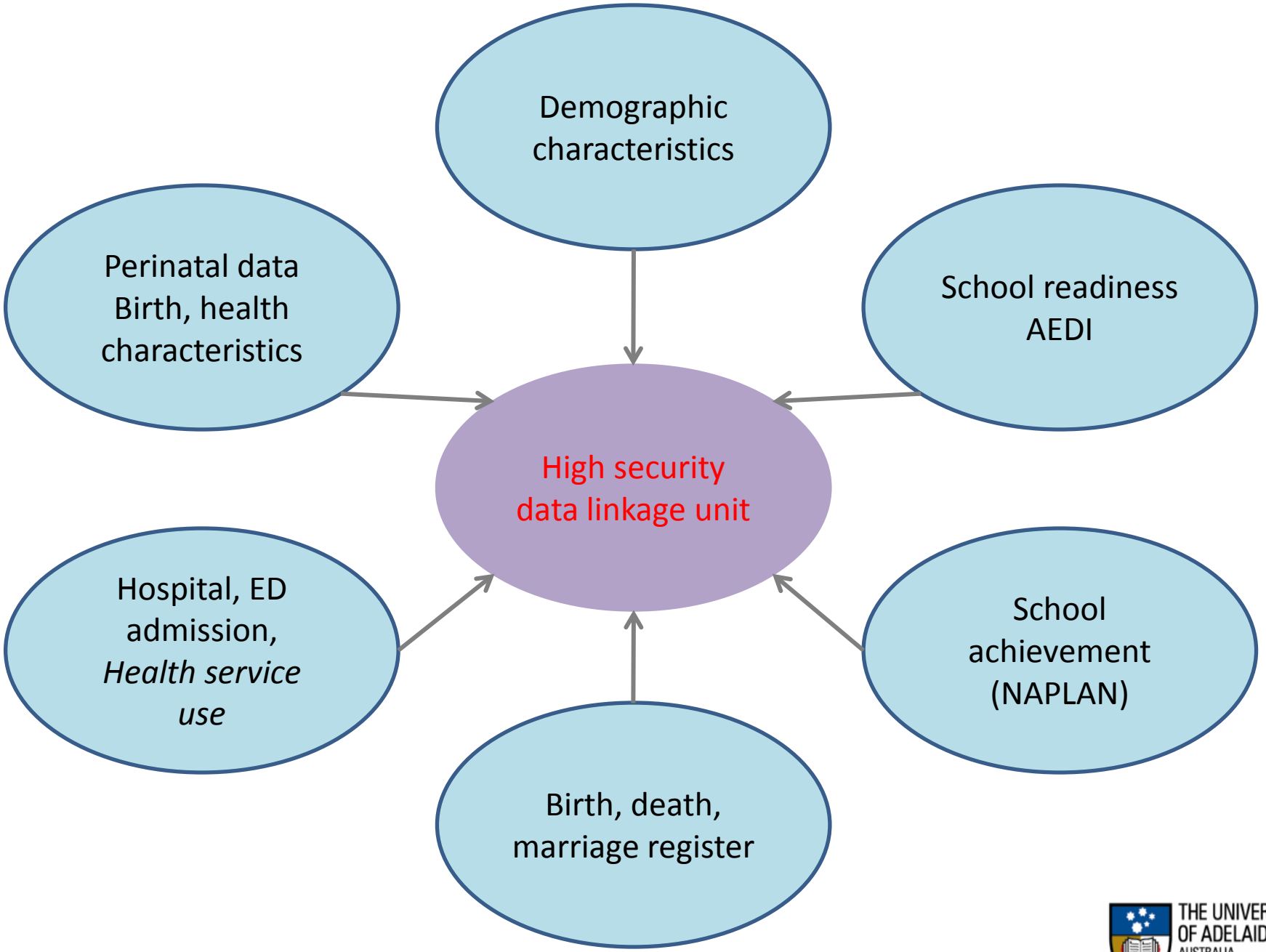
The effects of early life conditions and experiences on child development and learning: a whole of population study

Contains information on all children born in SA from 1999-2011

from birth through to age 12 (~ 200,000 children)

SA·NT DataLink

THE UNIVERSITY OF ADELAIDE AUSTRALIA



* Data comes from 13 different datasets

THE UNIVERSITY OF ADELAIDE AUSTRALIA

Data Sets and Years Available

Born	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Births													
Perinatal													
Deaths													
WCHN	e-crib												
							e-chims from Sept 2005						
ISAAC													
ED													
Dental													
AEDI													
School Census							not available until 2005						
Running Records Yr 1-2 school									missing	not reliable until 2007			
NAPLAN										only state based LAN before 2008			
ESL								not reliable until 2006					
Behaviour													
Attendance													

Current Data Set: Born 1999-2005

Extended Data Set: Born 2006-20011

Age	Yr School	Born	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1			2000	2001	2002	2003	2004	2005	2006						
2			2001	2002	2003	2004	2005	2006	2007						
3			2002	2003	2004	2005	2006	2007	2008						
4		4 Year Health Check	2003	2004	2005	2006	2007	2008	2009						
5	Reception	AEDI	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
6	1	RRs	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
7	2	RRs	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
8	3	NAPLAN	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
9	4		2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
10	5	NAPLAN	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
11	6		2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
12	7	NAPLAN	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
13	8		2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
14	9	NAPLAN	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
15	10														
16	11														
17	12														

THE UNIVERSITY OF ADELAIDE AUSTRALIA

Research Questions

1. Investigate the consequences of pregnancy complications, poor birth outcomes, and social, economic and demographic disadvantage at birth on child health (PPAs), school readiness at age 5, and school achievement (literacy, numeracy) up to age 12.

2. Develop a data-driven risk prediction model that can be used to facilitate more accurate identification of families who will benefit from the South Australia Family Home Visiting (SA-FHV) program.

3. Obtain a better understanding of the social and health origins of poor literacy and numeracy in school.

4. Describe the epidemiology of hospital emergency presentations and PPAs for children aged 0 to 8 years. Improve understanding about the relationship between key perinatal and sociodemographic factors, and hospital emergency presentations and PPAs.

THE UNIVERSITY OF ADELAIDE AUSTRALIA

Data Sets and Years Available

Born	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Births													
Perinatal													
Deaths													
WCHN	e-crib												
							e-chims from Sept 2005						
ISAAC													
ED													
Dental													
AEDI													
School Census							not available until 2005						
Running Records Yr 1-2 school									missing	not reliable until 2007			
NAPLAN										only state based LAN before 2008			
ESL								not reliable until 2006					
Behaviour													
Attendance													

Current Data Set: Born 1999-2005

Extended Data Set: Born 2006-20011

Age	Yr School	Born	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1			2000	2001	2002	2003	2004	2005	2006						
2			2001	2002	2003	2004	2005	2006	2007						
3			2002	2003	2004	2005	2006	2007	2008						
4		4 Year Health Check	2003	2004	2005	2006	2007	2008	2009						
5	Reception	AEDI	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
6	1	RRs	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
7	2	RRs	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
8	3	NAPLAN	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
9	4		2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
10	5	NAPLAN	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
11	6		2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
12	7	NAPLAN	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
13	8		2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
14	9	NAPLAN	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
15	10														
16	11														
17	12														

THE UNIVERSITY OF ADELAIDE AUSTRALIA

Dr Sally Brinkman

Proof of Concept 3

- This Proof of Concept project aims to use linked health, AEDI and education data to explore differences within and across jurisdictions in patterns of child development.

- This will support the investigation of the extent to which different jurisdictional systems/policies result in differences in children's development.

- The aim of the Proof of Concept project has both infrastructure and epidemiological aims.

Change in the proportion of children developmentally vulnerable by State/Territory from 2009 to2012

	2009		2012		Comparative result
	No. of children	Developmentally vulnerable on one or more domain/s	No. of children	Developmentally vulnerable on one or more domain/s	
New South Wales	82,710	21.3	88,921	19.9	↑
Victoria	57,277	20.3	63,584	19.5	↑
Queensland	52,603	29.6	57,994	26.2	↑
Western Australia	26,052	24.7	30,631	23.0	↑
South Australia	15,009	22.8	17,355	23.7	↓
Tasmania	5,699	21.8	6,086	21.5	↑
Northern Territory	2,865	38.7	3,117	35.5	↑
Australian Capital Territory	4,180	22.2	4,594	22.0	↑

FRASER MUSTARD CENTRE

A collaboration between

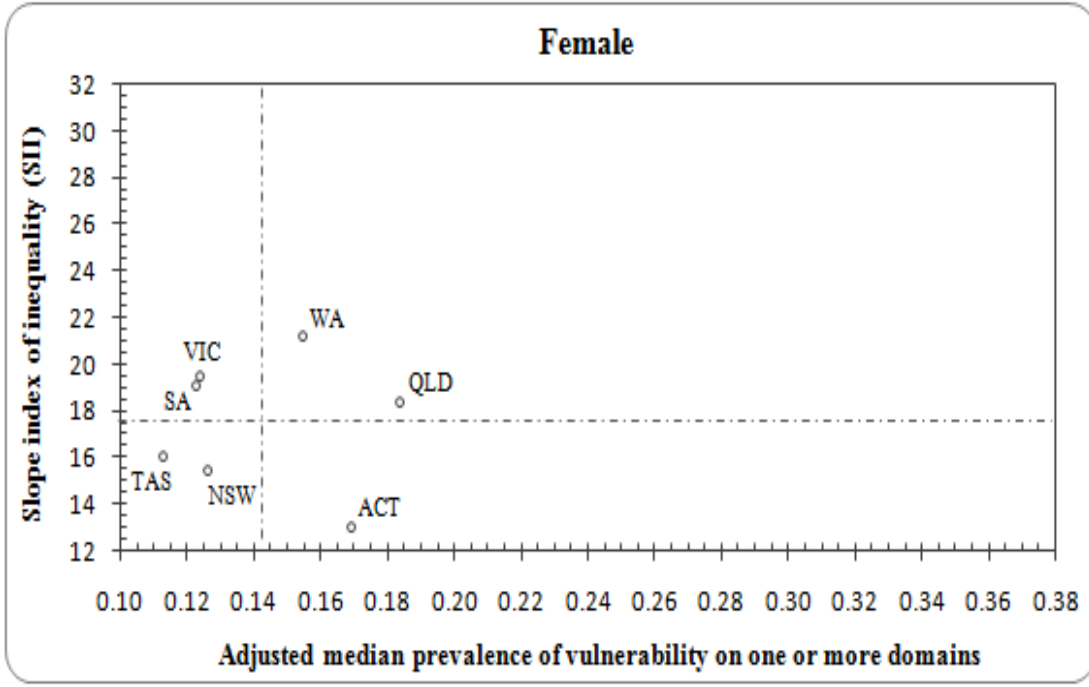
Government of South Australia
Department for Education and Child Development

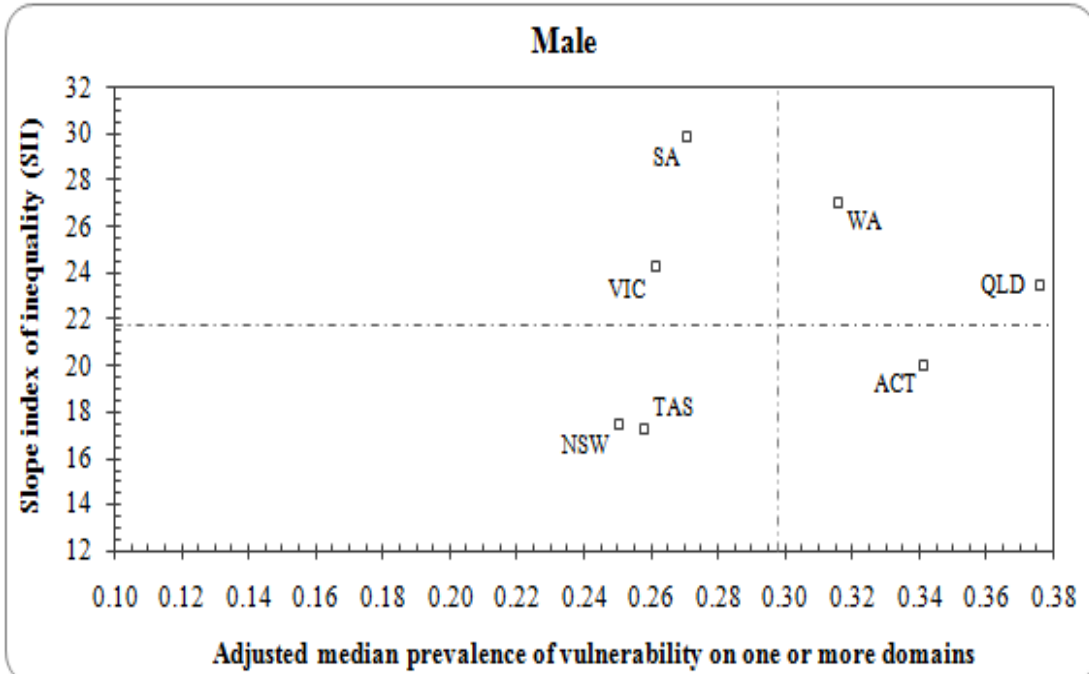
Telethon Institute for Child Health Research

Inequality in Child Development

Top right corner = high inequality with high vulnerability.

Bottom left corner = low inequality and low vulnerability





Source: Brinkman et al. BMJ Open (2012)

Proof of Concept 3

- Evaluation should not just be limited to programs, policies should also be evaluated

- National linked data provides opportunities for natural experiment designs to evaluate different policies:
  - minimum of 15 hours universal access to preschool
  - CHN delivered through NonGovs versus State Govt and implications for universality of access and service quality/fidelity

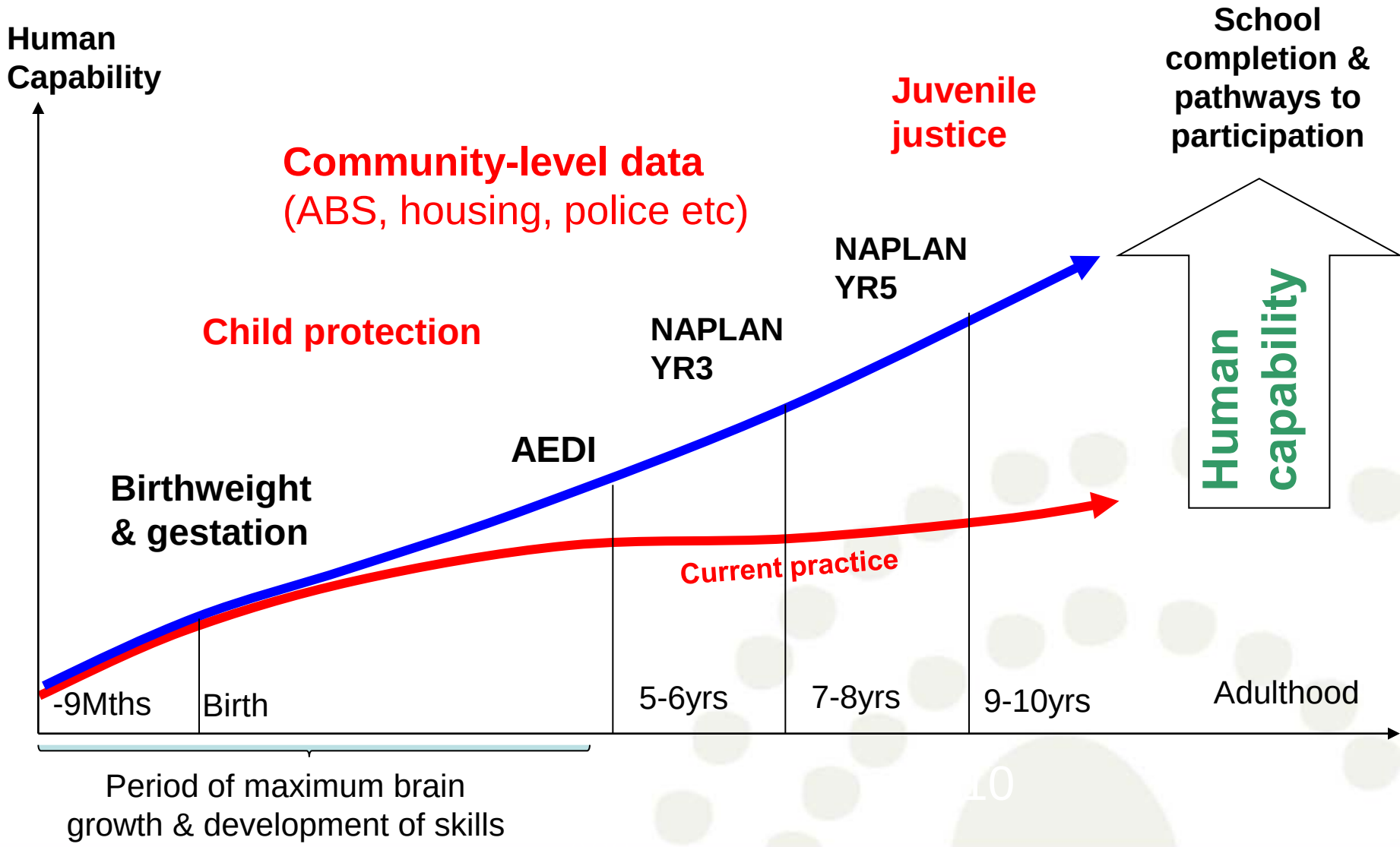
Professor Sven Silburn

Early life determinants of school education outcomes in the NT: a data linkage study

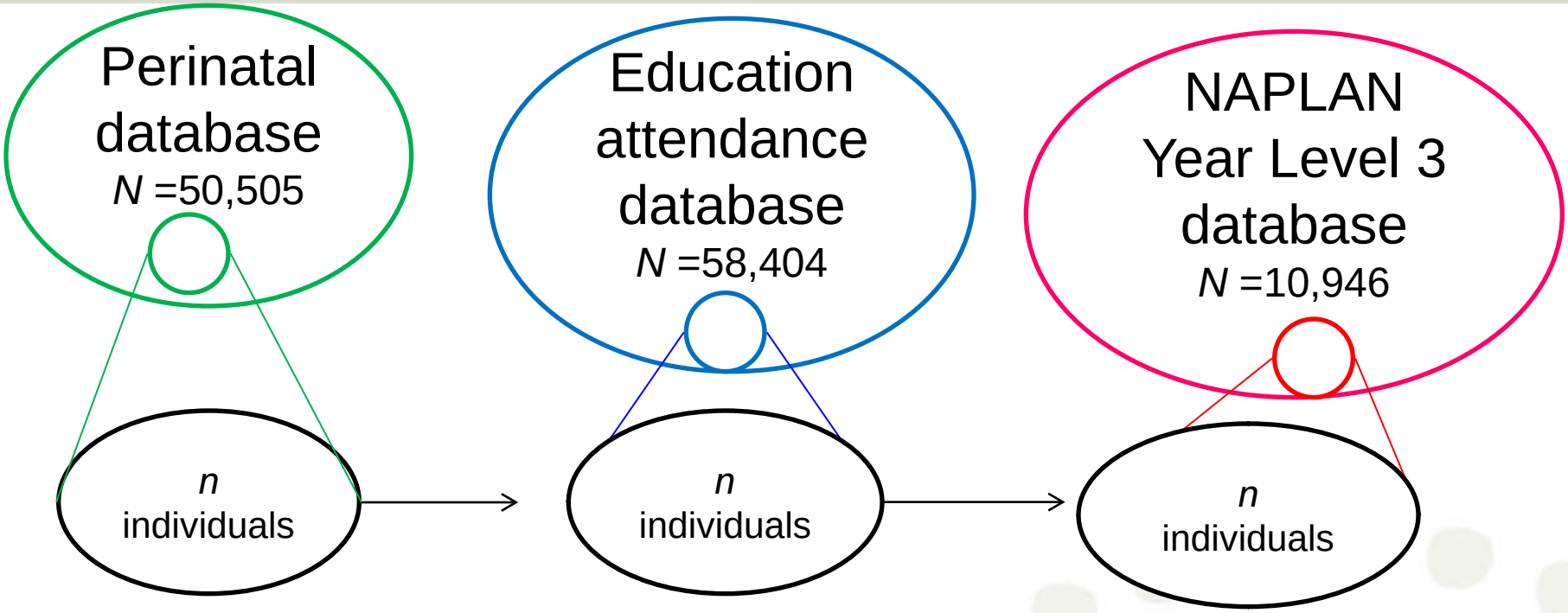
Sven Silburn, John McKenzie Eric Grist,

Centre for Child Development & Education,
Menzies School of Health Research,
Charles Darwin University, Darwin, Australia.

Life-span human development



NT Linked dataset: Initial exploratory analysis



Selected Variables

- *ethnicity*
- *birth weight*
- *gestational age*
- *age of mother*
- *alcohol*
- *smoking*

Selected variables

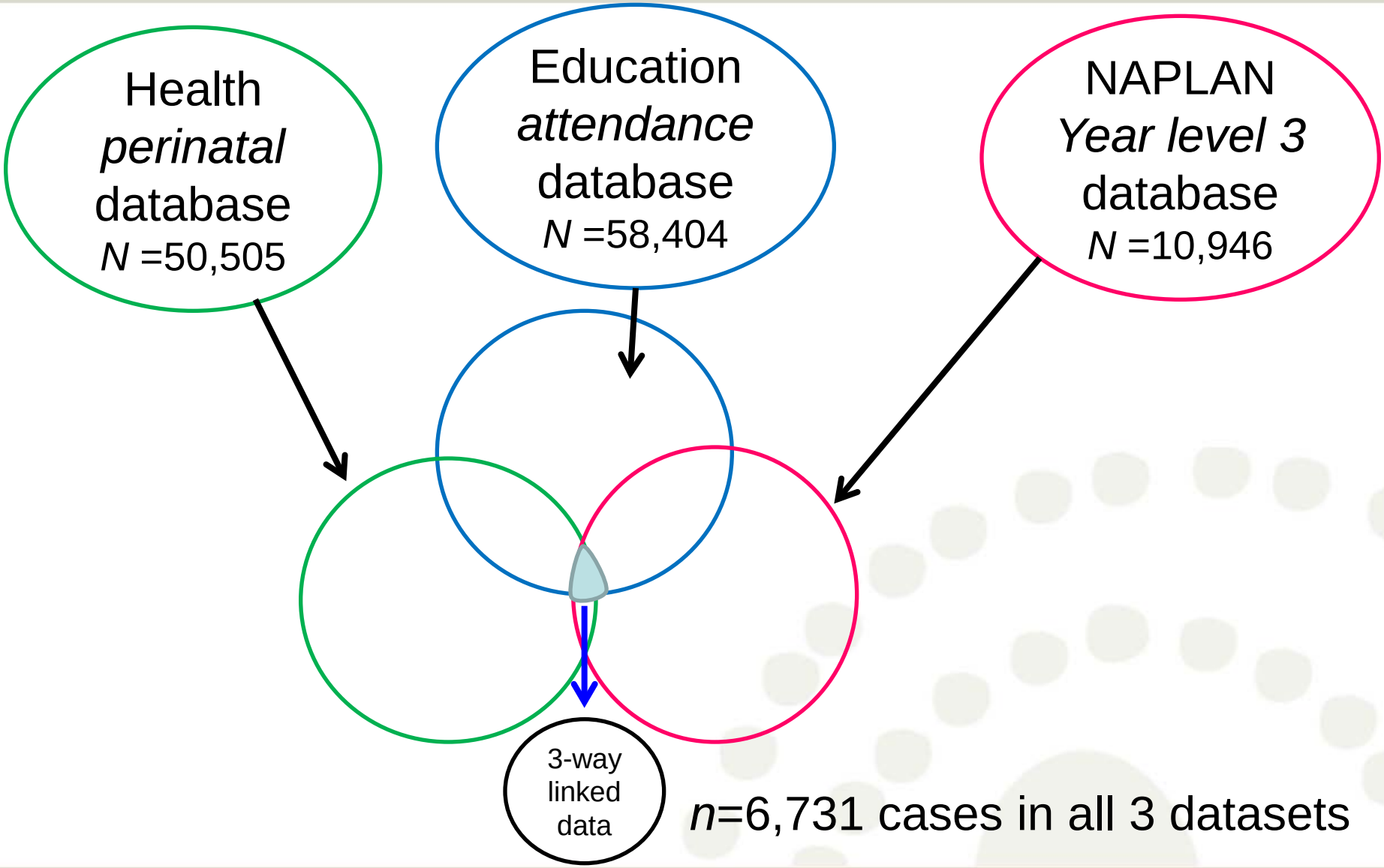
- *school attendance*

Selected variables

- *scaled NAPLAN Year Level 3 (8 years old)*
- *domain description (subject)*

time

Cases that can be followed through time



n=6,731 cases in all 3 datasets

Variables examined

*** = substantial missing or unknown values**

blue = continuous variable

Variable	Perinatal (child)
*aborbaby**	Indigenous status of the baby
birthwt	Birth weight - the first weight of the baby taken after birth (in grams)
gest_age	Estimated gestational age of the baby in completed weeks
sex	gender of the baby at birth
	Antenatal (mother)
abormthr	Indigenous status of the mother
agemum	Age of the mother at birth of child (in years)
*alchvst1**	Alcohol consumed at the time of first antenatal clinic visit
*alch36wk**	Alcohol consumed at week 36 of the pregnancy
*smokvst1**	Smoking reported at the first antenatal clinic visit
*smok36wk**	Smoking reported at week 36 of the pregnancy
	Educational data (child)
perc_exp_ attend	School attendance as a percentage of expected attendance (PEA)
	NAPLAN Year level 3 test (age 8 years, tests in years 2008-2010)
scale_score	scaled test result for children taking NAPLAN Year Level 3 test
naplan_domain	test category (Numeracy, Reading, Writing, Grammar, Spelling)

Correlation matrices (5 continuous)

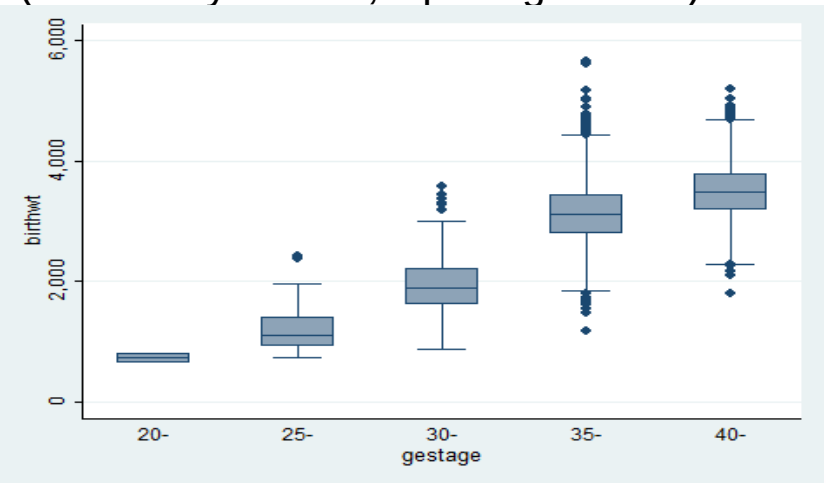
Numeracy, *n*=6587

	agemum	birthwt	gest_age	perc_e~d	scale_~e
agemum	1.0000				
birthwt	0.1183*	1.0000			
	0.0000				
gest_age	0.0081	0.6436*	1.0000		
	0.5087	0.0000			
perc_exp_a~d	0.3386*	0.1706*	0.0979*	1.0000	
	0.0000	0.0000	0.0000		
scale_score	0.3330*	0.1878*	0.1147*	0.6093*	1.0000
	0.0000	0.0000	0.0000	0.0000	

Spelling, *n*=6731

	agemum	birthwt	gest_age	perc_e~d	scale_~e
agemum	1.0000				
birthwt	0.1242*	1.0000			
	0.0000				
gest_age	0.0169	0.6491*	1.0000		
	0.1645	0.0000			
perc_exp_a~d	0.3379*	0.1759*	0.1070*	1.0000	
	0.0000	0.0000	0.0000		
scale_score	0.3229*	0.1532*	0.0976*	0.5469*	1.0000
	0.0000	0.0000	0.0000	0.0000	

Greatest correlation between *birthwt* and *gest_age*
(Numeracy *0.6436,* Spelling *0.6491*)

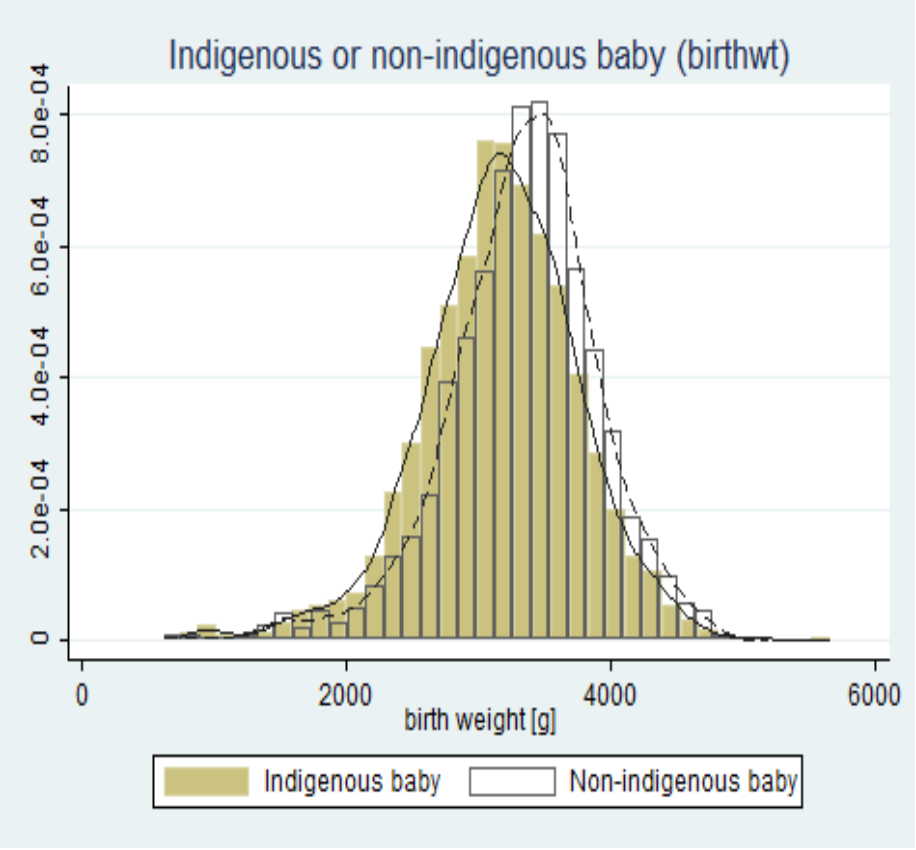


Ranked correlations forscale_score	Numeracy	Spelling
PEA	0.6093	0.5469
agemum	0.3330	0.3229
birthwt	0.1878	0.1532
gest_age	0.1147	0.0976

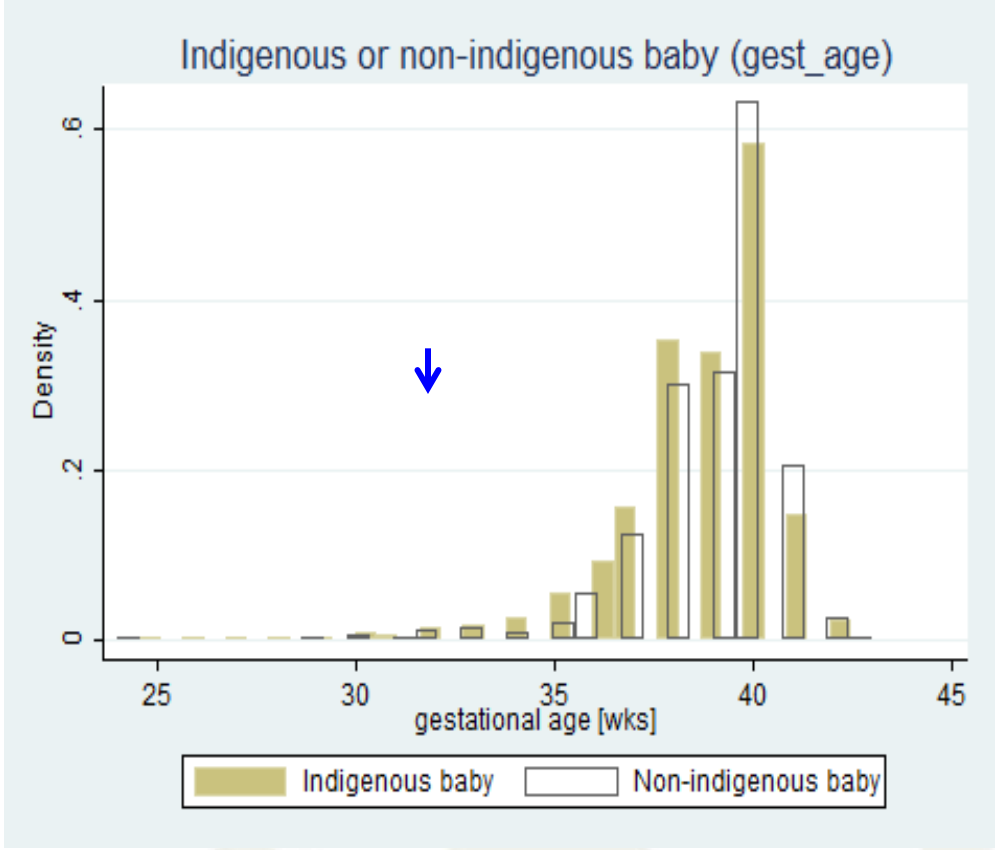
* indicates statistically significance at p < 0.001

Birthweight & gestational age by Indigenous status

Birth weight [g]

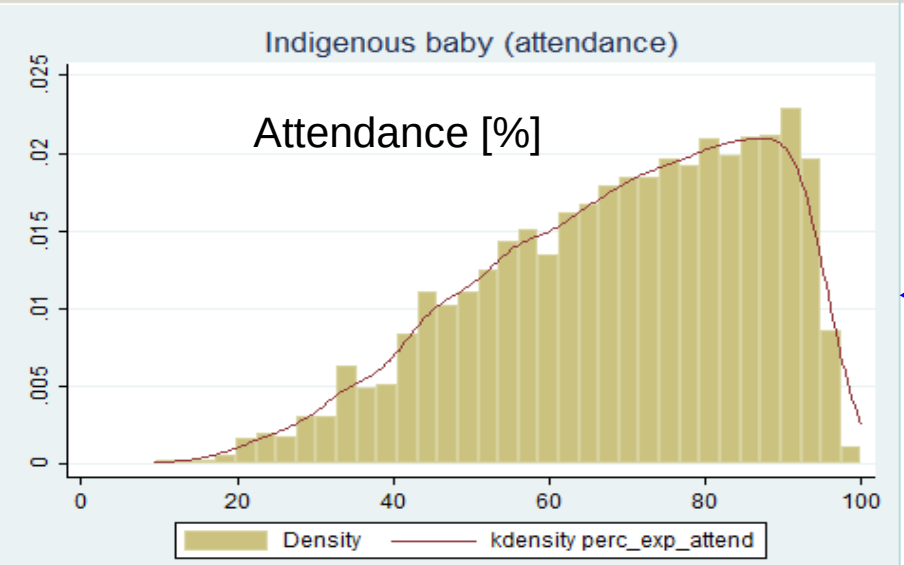


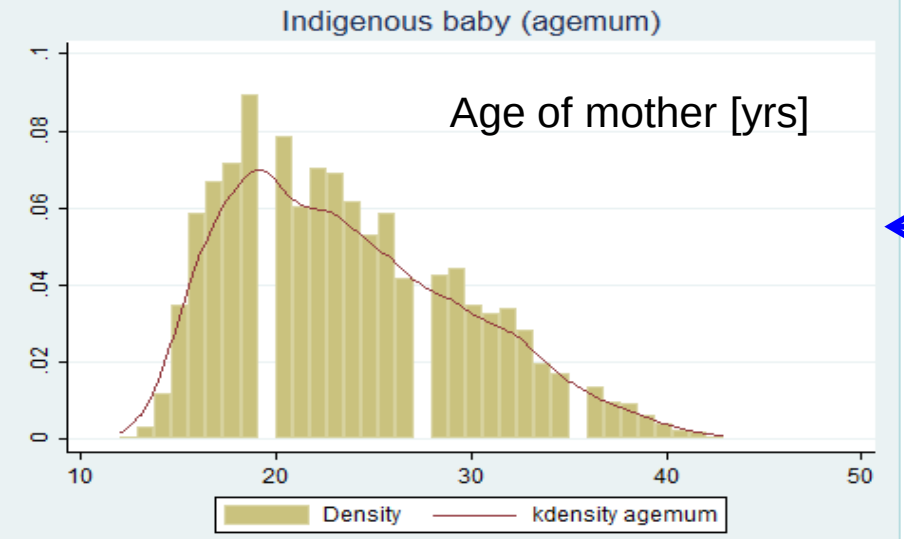
Gestational age [wks]



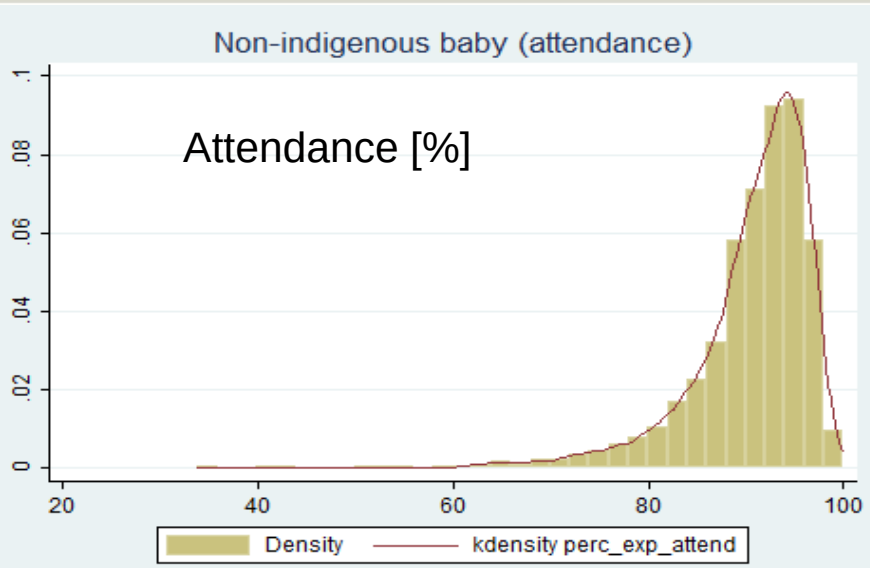
School attendance by mother's age at child's birth

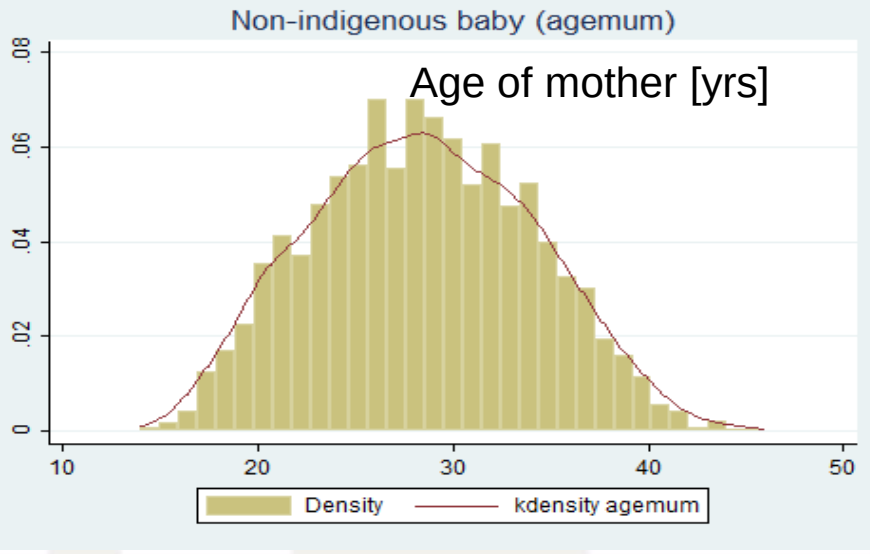
Indigenous baby





Non-indigenous baby

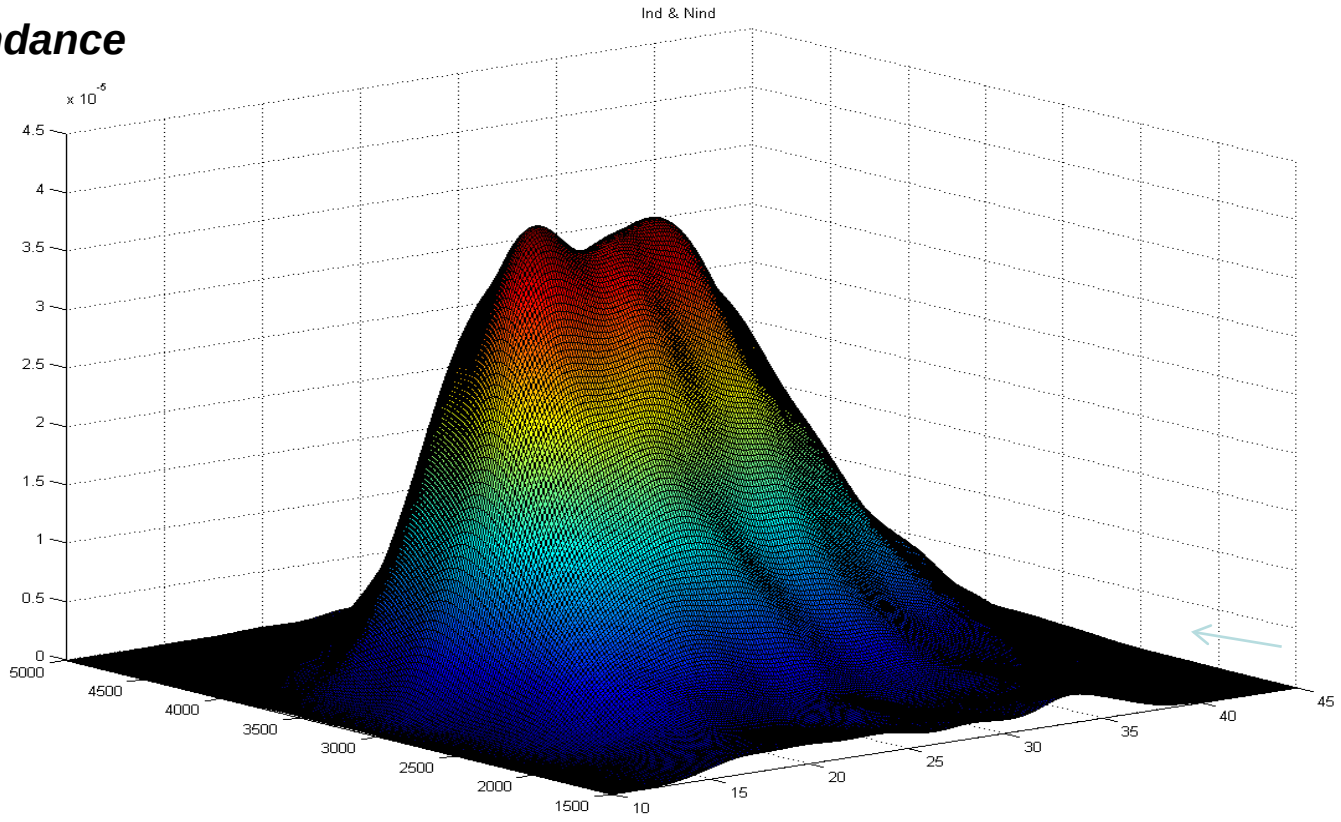




School attendance: mothers age at birth & birthweight

3D kernel densities for *agemum* and *birthwt*

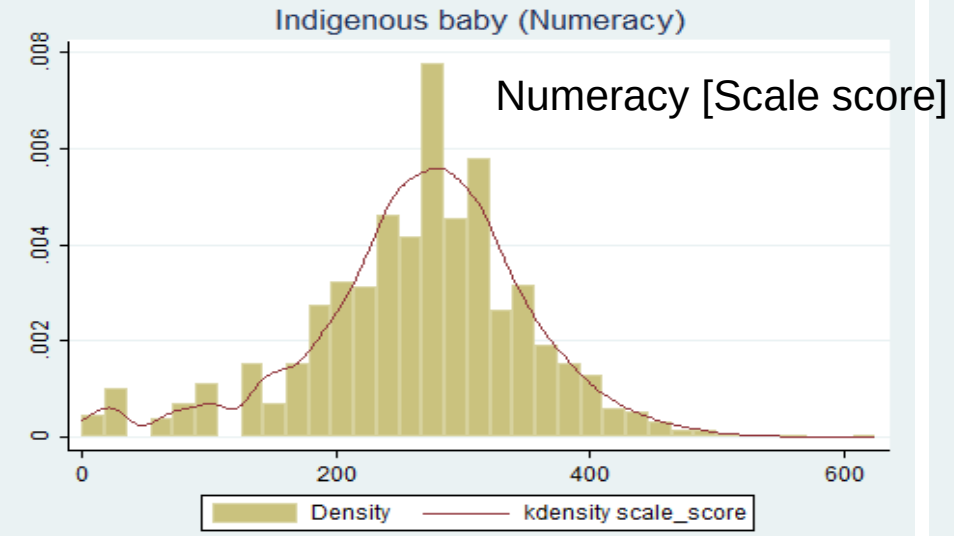
Attendance

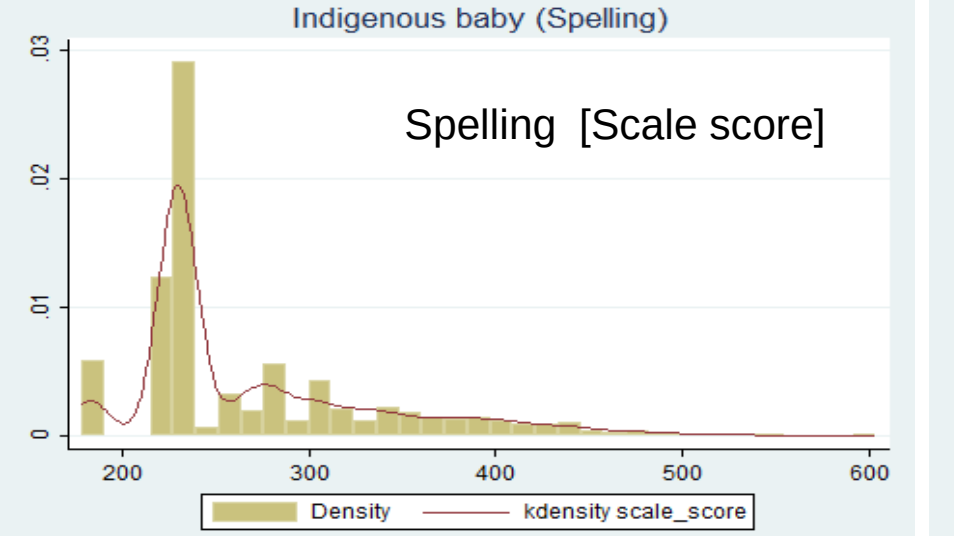


birthwt ***agemum***

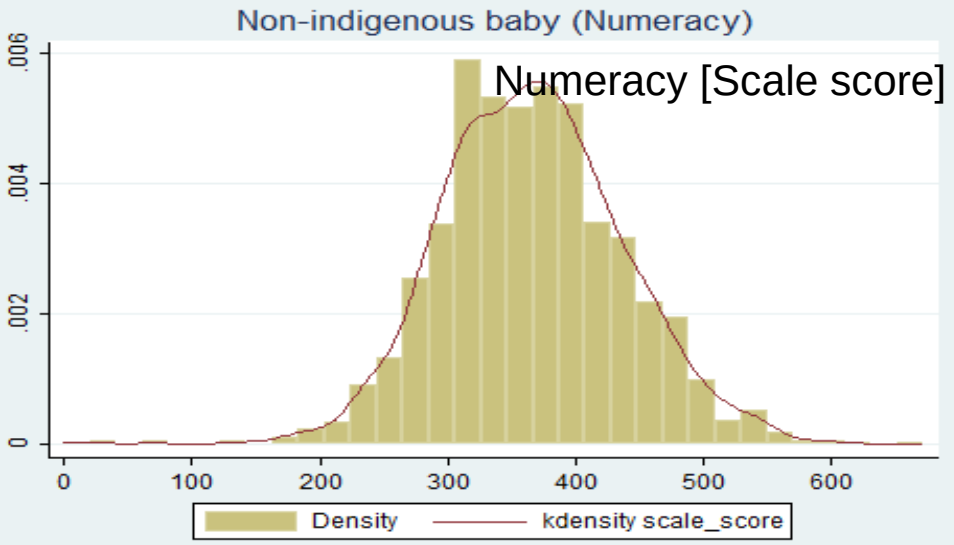
NAPLAN Yr 3 Numeracy and Spelling

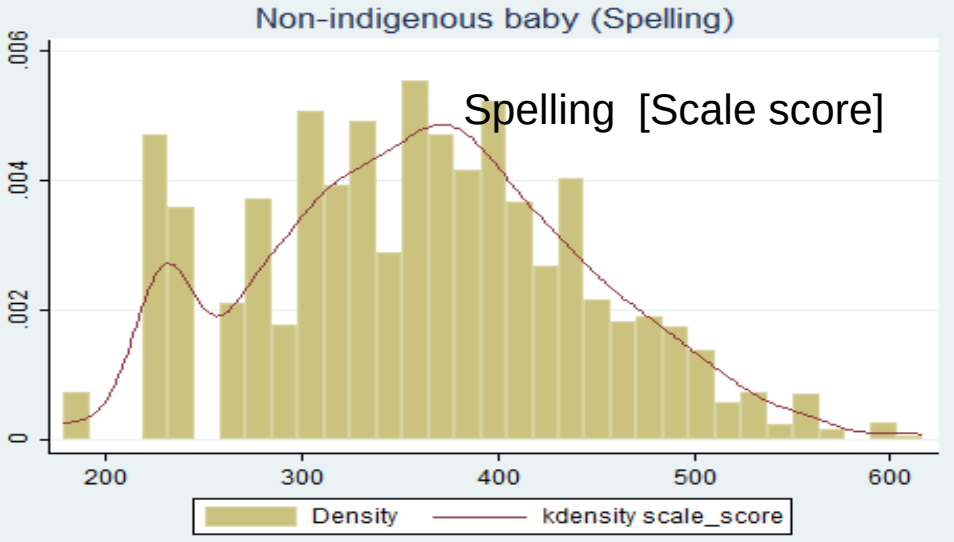
Indigenous baby





Non-indigenous baby





Imputation of unknown & missing data
e.g. Indigenous status of baby & mother

Key
frequency *cell percentage*

abormthr	aborbaby Indig	Non_indig	Unknown	.	Total
Indig	2,801 42.52	42 0.64	10 0.15	312 4.74	3,165 48.05
Non_indig	429 6.51	2,187 33.20	424 6.44	382 5.80	3,422 51.95
Total	3,230 49.04	2,229 33.84	434 6.59	694 10.54	6,587 100.00

No missing values
for mother's ethnicity
abormthr

⇨

- Strong association between *abormthr* and *aborbaby*

Apply a univariateimputation model for *aborbaby* which uses the ethnic information in *abormthr* to estimate missing child ethnicity values

aborbaby = f(abormthr, agemum, gender) (1)

where *f* is the multinomial logistic regression (logit) function (e.g. Raghunathan et al 2001).

Imputation of unknown & missing data

Child ethnicity data has 434 + 694 = 1128 absent values

aborbaby	Freq.	Percent	Cum.
Indig	3,230	49.04	49.04
Non_indig	2,229	33.84	82.88
Unknown	434	6.59	89.46
missing	694	10.54	100.00
Total	6,587	100.00	

Imputation Step 1

aborbaby	Freq.	Percent	Cum.
Indig	356	51.30	51.30
Non_indig	281	40.49	91.79
Unknown	57	8.21	100.00
Total	694	100.00	

Indig	3,586	54.44	54.44
Non_indig	2,510	38.11	92.55
Unknown	491	7.45	100.00
Total	6,587	100.00	

Imputation Step 2

aborbaby	Freq.	Percent	Cum.
Indig	71	14.46	14.46
Non_indig	420	85.54	100.00
Total	491	100.00	

Imputed child ethnicity data

aborbaby	Freq.	Percent	Cum.
Indig	3,657	55.52	55.52
Non_indig	2,930	44.48	100.00
Total	6,587	100.00	

No missing and unknown values

Imputation model estimates for absent values are therefore:

Indigenous 427 (=356 +71)

Non-indigenous 701 (=281+420)

Total 1128 imputed values

NAPLAN outcomes by antenatal alcohol and smoking

Are there any relative differences in child educational outcome if the data are partitioned by antenatal smoking or alcohol consumption?

- Alcohol consumption and smoking data are from records at weeks 1 and 36 only based on questionnaires
- Records were self reported by mothers ⇨ subject to error

Adopt a precautionary approach and assume alcohol consumption and smoking is under reported.

Do this by defining a derived variable *alch* which combines positive evidence of alcohol consumption in *alchvst1* and *alch36wk* according to the following truth table:

alchvst1	***alch36wk***	***alch***
yes	yes	yes
yes	no or unknown	yes
no or unknown	yes	yes
no	no	no
no	unknown	unknown
unknown	no	unknown
unknown	unknown	unknown

	alch	**smok**
Total unknown	1469 (22%)	1107 (16%)

Similarly define a derived variable *smok* from *smokvst1* and *smok36wk*

How can we estimate the unknown data on antenatal alcohol consumption and smoking?

Apply a multivariate imputation model for missing values in *alchvst1 , alch36wk, smokvst1, smok36wk (and aborbaby)* chained equations (implemented in Stata by Royston and White (2011) based on theory by Raghunathan et al 2001 and Rubin 1987):

aborbaby, alchvst1, alch36wk, smokvst1 smok36wk = f(abormthr,agemum,gender) (2)

where *f* is the multinomial logistic regression (logit) function.

This gives rise to 3 data sets with different levels of imputation as follows:

Data set 1

The unknown values for *alch and smok* (last 3 rows of table) listwise deleted

Data set 2

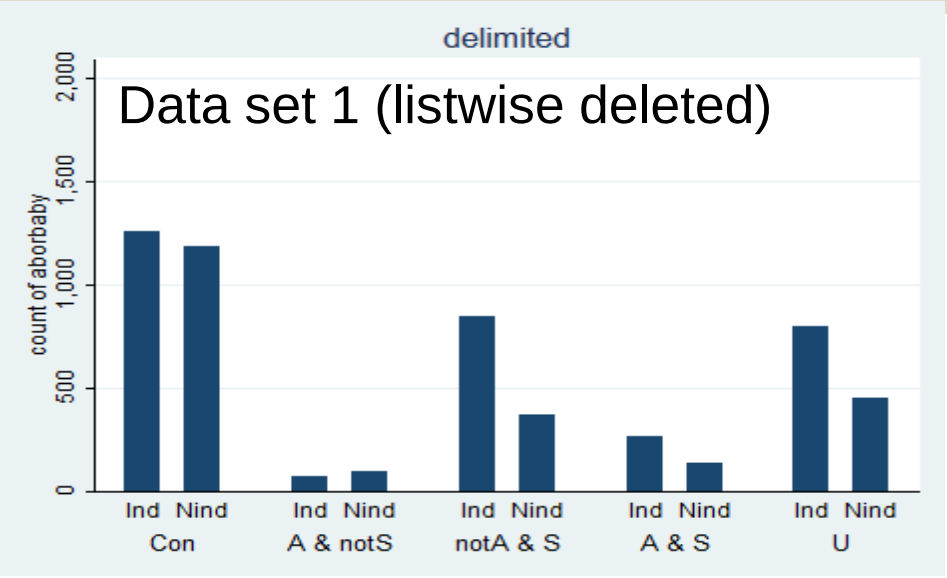
The values for *aborbaby* are imputed using the univariate imputation model and unknown values for *alch* and *smok* (last 3 rows of table) are listwise deleted

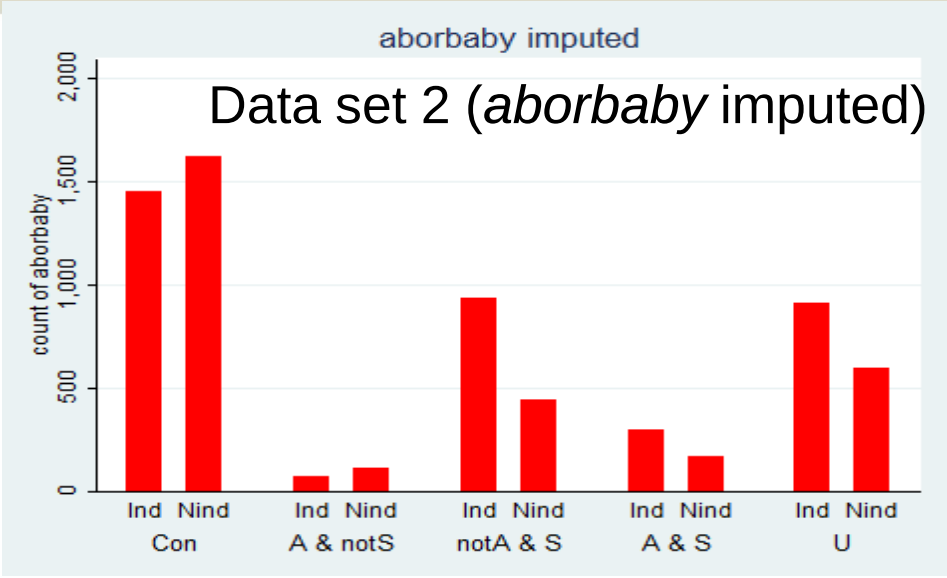
Data set 3

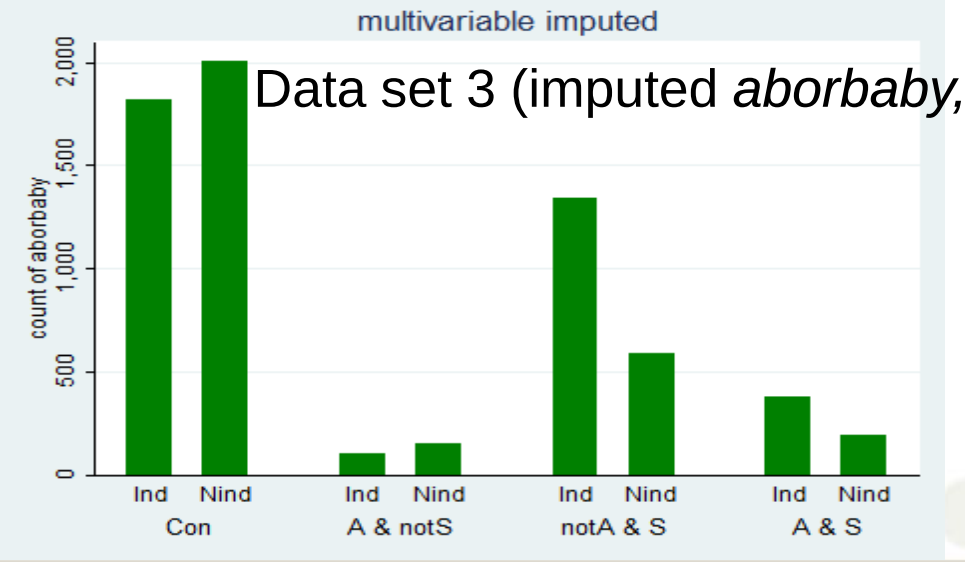
The unknown values for *alch and smok (and aborbaby)* are imputed using the multivariate imputation model (thus no omitted values for *alch* and *smok*)

Comparisons by antenatal alcohol and smoking

Distribution numbers of individuals by Indigenous status and category in each data set







Key:

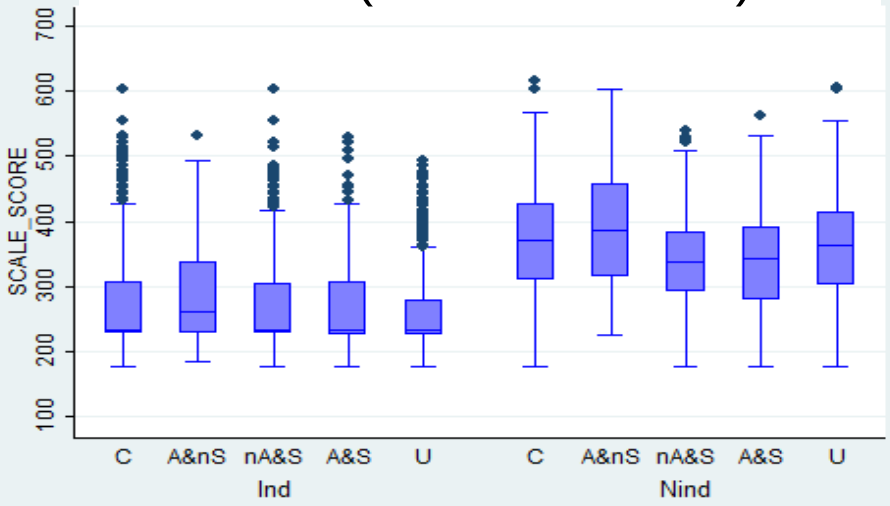
Con = no smoking & no alcohol
A & notS = alcohol and no smoking
notA & S = no alcohol and smoking
A & S = alcohol and smoking

Ind = Indigenous child
Nind= Non-indigenous child

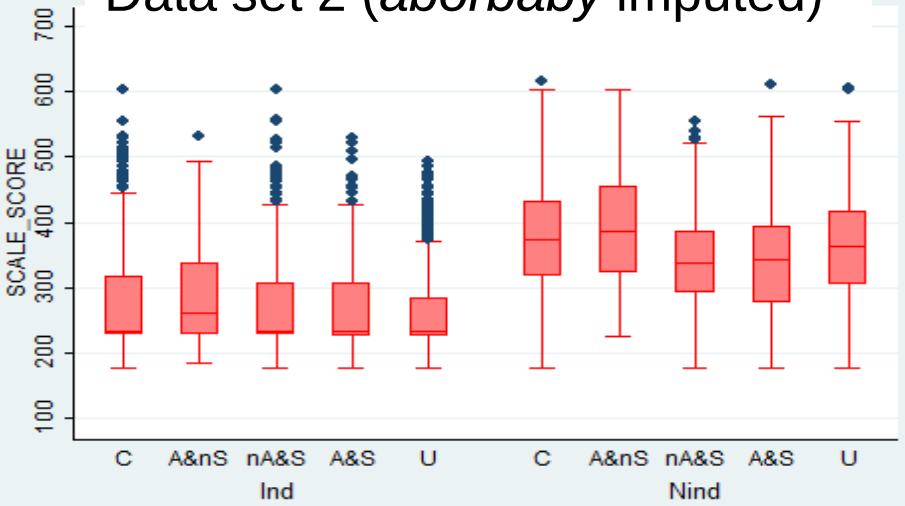
Antenatal alcohol & smoking and NAPLAN spelling by category nad Indigenous status

(*n*=6731)

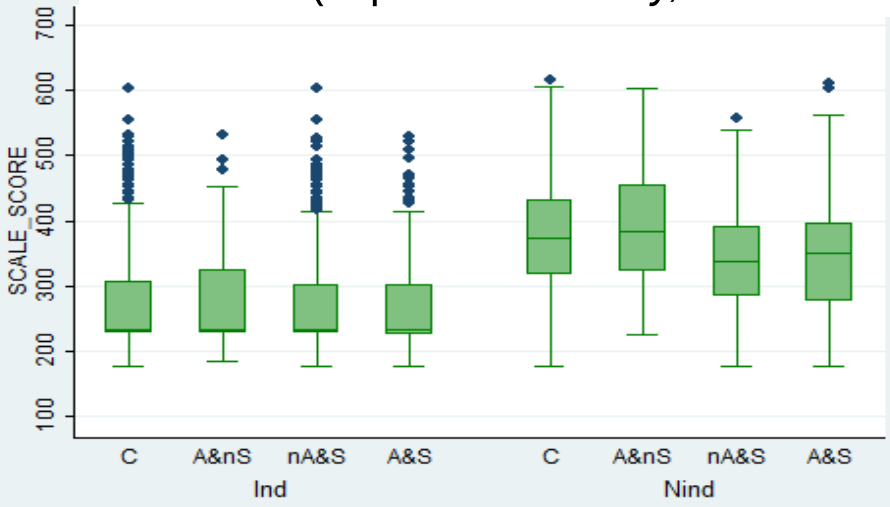
Data set 1 (listwise deleted)



Data set 2 (*aborbaby* imputed)



Data set 3 (imputed *aborbaby, alchs and smoks*)



Comparisons between the baseline (Control) with nA&S, A&S or U for indigenous and non-indigenous are statistically significant at $p<0.001$

However, it is not significant with A&nS (p=0.0358, 0.0358, 0.0685 respectively)

References

Raghunathan, T. E. et al. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models.
Survey Methodology 27: 85–95.

Royston P.R. and White I. (2011). Multiple Imputation by Chained Equations (MICE): Implementation in Stata .
Journal of Statistical Software 45(4): 1-20

Rubin, D. B. (1987) *Multiple Imputation for Nonresponse* in Surveys. New York: Wiley

Mr Sam Luddy

Replacing Mr David Engelhardt

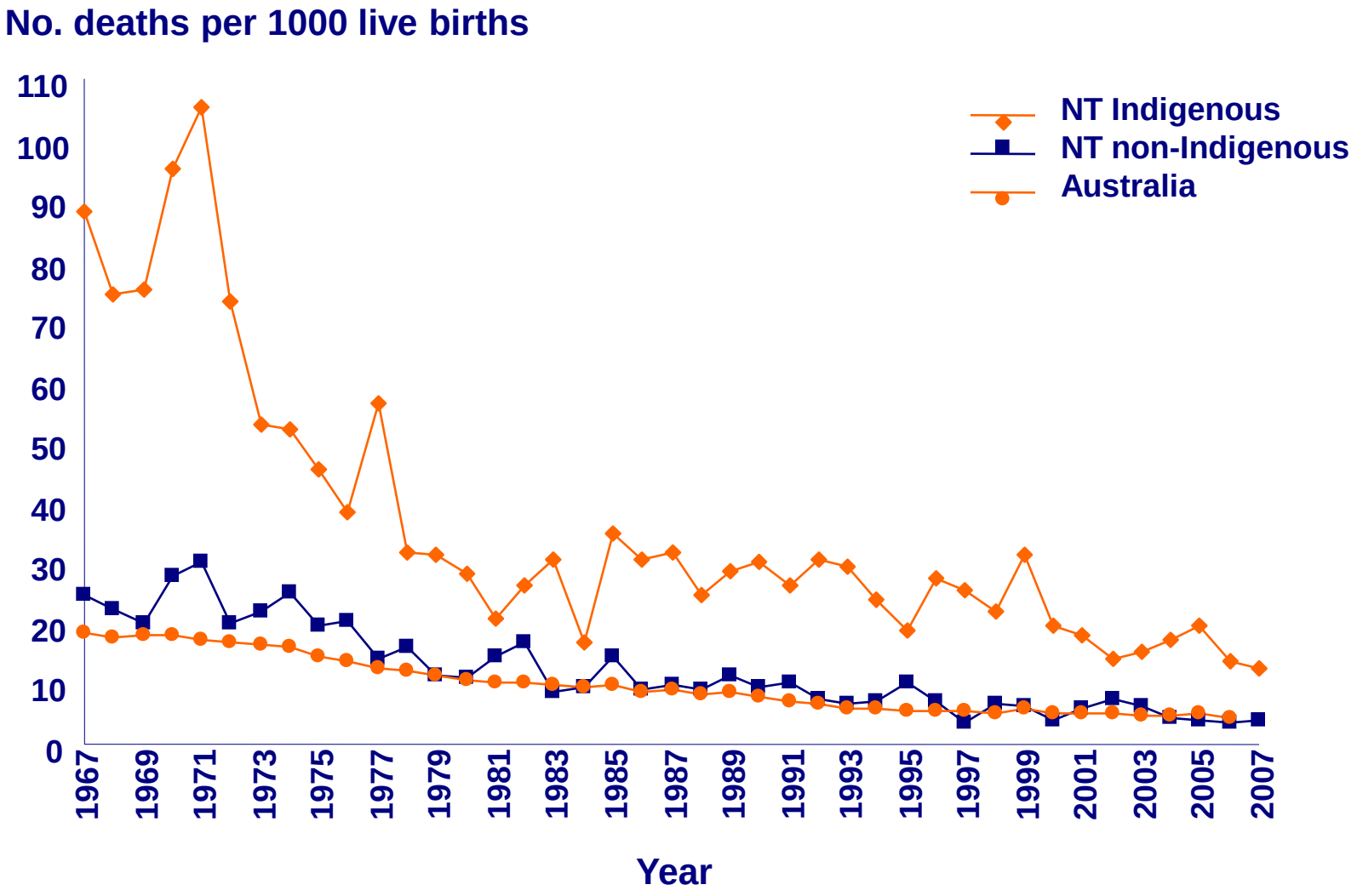
Dr Steve Guthridge

Enhanced Reporting on Closing the Gap Targets

Developing a data linkage infrastructure and demonstrating its utility

Steven Guthridge
Director, Health Gains Planning
NT Department of Health

Infant mortality, NT - 1967 to 2007



Phase 1: Infrastructure Development

Background

- This study was funded by the Department of Innovation, Industry, Science and Research (DIISR) as part of the Education Investment Fund – Super Science Initiative in 2011.
- Enhanced reporting against the six specific COAG closing the gap targets is part of the National Indigenous Reform Agreement to which the NT is a signatory

Phase 1 - Aim

- To establish the technical infrastructure and associated processes to create anonymised, researchable and linkable datasets containing NT Births and Deaths data

Phase 1: Infrastructure development

Data sources

- Births dataset
  - Obtained from the NT Department of Justice Office of Births, Deaths and Marriages
  - Contains data from 1868 – 2012 (175,250 records)
- Deaths dataset
  - Obtained from the NT Department of Justice Office of Births, Deaths and Marriages
  - Contains data from 1870 – 2012 (47,532 records)
- Client Master Index (CMI)
  - Obtained from the NT Department of Health
  - Contains data from 1976 – 2012 (758,818 records)

Phase 1: Infrastructure development

Progress so far

- Data Summary Document
  - Assessment of the quality of the individual datasets
  - Assessment of the quality of the linkage
  - Data dictionary
- Data Access Protocol
  - For internal and external users of the data
  - Includes Data Access Form
- Data Transfer Protocol
  - For internal use to guide data transfer from SANT DataLink to individual researcher requests

Estimated date of completion: December 2013

Phase 2: Demonstration study

Background

- This study was developed to demonstrate the utility of the anonymised, researchable datasets made linkable in Phase 1

Phase 2 - Aim

- To utilise anonymised, linkable Births, Deaths and CMI data alongside NT perinatal data to establish a period cohort of the NT Indigenous population from 1992 to 2012 and evaluate:
  - The quality and accuracy of recording of Indigenous identification of NT birth and death registrations
  - The quality and accuracy of infant mortality and life expectancy reporting
  - The current study findings and methods against current methods used by the ABS

Estimated date of completion: March 2014