Linkage rate between NSW Perinatal Data Collection birth records and government school NAPLAN educational records, by gestational age at birth

Daneeta Hennessy, Siranda Torvaldsen, Christine Roberts

Clinical and Population Perinatal Health Research, Kolling Institute, Northern Sydney Local Health District, St Leonards NSW 2065 Australia

Sydney Medical School Northern, University of Sydney

Introduction

Exposures to the fetus during pregnancy, and the newborn at birth, may have significant impacts on their long-term health and development. Traditional approaches to investigating long-term outcomes are often hampered by selection bias or small sample sizes. Record linkage of population data minimises these problems by using routinely collected data with population coverage.

The availability and facility to link population health data to standardised educational testing data provides an opportunity to follow infants from birth through to school age, and examine associations between fetal or infant exposures and school performance.

Gestational age at birth is a strong predictor of health and developmental outcomes in childhood and later in life [1], so an understanding of the availability and nature of educational records over the gestational age range is essential in planning and interpreting any record linkage study using these data.

Aim: to examine the linkage of birth records from the NSW Perinatal Data Collection with educational results from the National Assessment Program – Literacy and Numeracy (NAPLAN).

Methods:

This work utilised linked birth, death, and education datasets. Birth records from 1994 to 2005 were taken from the NSW Perinatal Data Collection, a population-based statutory data collection that includes all live births and stillbirths of at least 20 weeks gestation or 400 grams birth weight.

NAPLAN is an annual standardised testing program for all students in Australian schools in grades 3, 5, 7, and 9 [2]. Children's skills in numeracy, reading, writing and language conventions (spelling, grammar and punctuation) are assessed and scored. For each test, scores are nationally equated across grades and years to ensure that results are comparable between grades and over time. Children are then categorised into one of six achievement bands based on their equated scores and nationally defined cut-points for each grade. All children who are enrolled in school when the NAPLAN is conducted will have a record for that year whether they sat the test or not. Reasons for not sitting the test are recorded: absent on the day of testing, withdrawn from testing due to parents' wishes, or exempt from testing due to significant or complex disability. Additional support is available to assist children with disabilities to undertake the tests, including large-print or braille test papers, use of a scribe, and rest breaks. Exemptions may be granted to students with disabilities who are unable to participate in testing despite these adjustments [3]. NAPLAN records from government schools were available for the years 2009 to 2014.

Death information was obtained from the NSW Registry of Births, Deaths, and Marriages death registration data, which includes all deaths registered in NSW. Data were available for deaths registered between January 1994 and March 2014. Additional deaths were identified by searching hospital admission records.

The study population was live births that occurred between 24 and 41 weeks of gestation, which had no missing data for birthweight, or sex. Overall 0.1% of birth records had missing data, and no more than 0.3% of records for any gestation were excluded due to missing data. To minimise the inclusion of records with incorrectly recorded gestational age, records with implausible birthweights were removed (n=1,371, 0.1%), based on Australian birthweight percentiles [4] and the method of Tukey, using a cutoff of 2.5 times the interquartile range [5].

For each week of gestational age, live births, live births with a linked death record at or before six years of age, and live births that linked to at least one government school record were enumerated. The proportion of children at each gestation with a linked school record was calculated in two ways: firstly as a proportion of all live births, and secondly as a proportion of all live born infants who had not died (no record of death in the linked data) by six years of age.

The proportions of children at each gestational age that were present (sat the test), absent, exempt, or withdrawn were calculated as a proportion of all children with at least one linked government school record. Where a child's records were available from multiple years, the first available record was used. These calculations were performed for the numeracy and reading test portions of NAPLAN assessment, as these are the components of the standardised testing which remain the most consistent over time. The pattern of linkage rates and proportions of absent, exempt and withdrawn children across the gestational age range was similar for numeracy and reading, so only the results for numeracy are presented.

For the purposes of this study, a 'low score' was defined in two ways: a score below the national minimum standard, and a score less than one standard deviation below the mean for that grade level and year. For both definitions, a child who was exempt from testing due to serious or complex disability was considered to have a low score. The proportion of children who had a result (either present or exempt) that obtained a low score by each of the definitions was calculated.

Birth, death, and education data was probabilistically linked by the NSW Centre for Health Record Linkage, and anonymised data was provided to the researchers. Quality assurance procedures ensure that both false positive and false negative linkage results are each less than 5 in 1000 [6].

Ethics approval for the study was obtained from the NSW Population and Health Services Research Ethics Committee (#2012-12-430).

Results and discussion:

Between 1994 and 2005, there were 1,010,018 live birth records that met the inclusion criteria. The proportion of live births that linked to a government school NAPLAN record was lowest in the earlier gestational ages (20.5% at 24 weeks), then increased though to a maximum at 32 weeks (59.4%), and then showed a slight decreasing trend in the later gestational ages up to 55.0% at 40 weeks (Figure 1, Table 1). When the denominator was children who lived to school age (live births that did not link to a death record by six years of age), the trend of lower linkage

rates at earlier gestations was less marked, increasing from 54.0% at 24 weeks to 60.7% at 30 and 32 weeks, then decreasing to 55.1% at 40 weeks. Both linkage rates were higher in births that occurred at 41 weeks than 40 weeks.

The overall linkage rate in children who lived to age six was 56%. The majority of the unlinked birth records are likely to be due to children attending non-government schools. In NSW, government schools account for approximately 70% of primary school students in a given year [7], so around 30% of children who lived to school age would not be expected to link to a government school record. Other reasons for non-linkage are: the child moving out of state before school age, the child not attending school at all, or death that occurred outside of NSW which was therefore not in the deaths data linked in this study.

Figure 1. NAPLAN linkage rates by gestational age at birth, NSW births 1994-2005

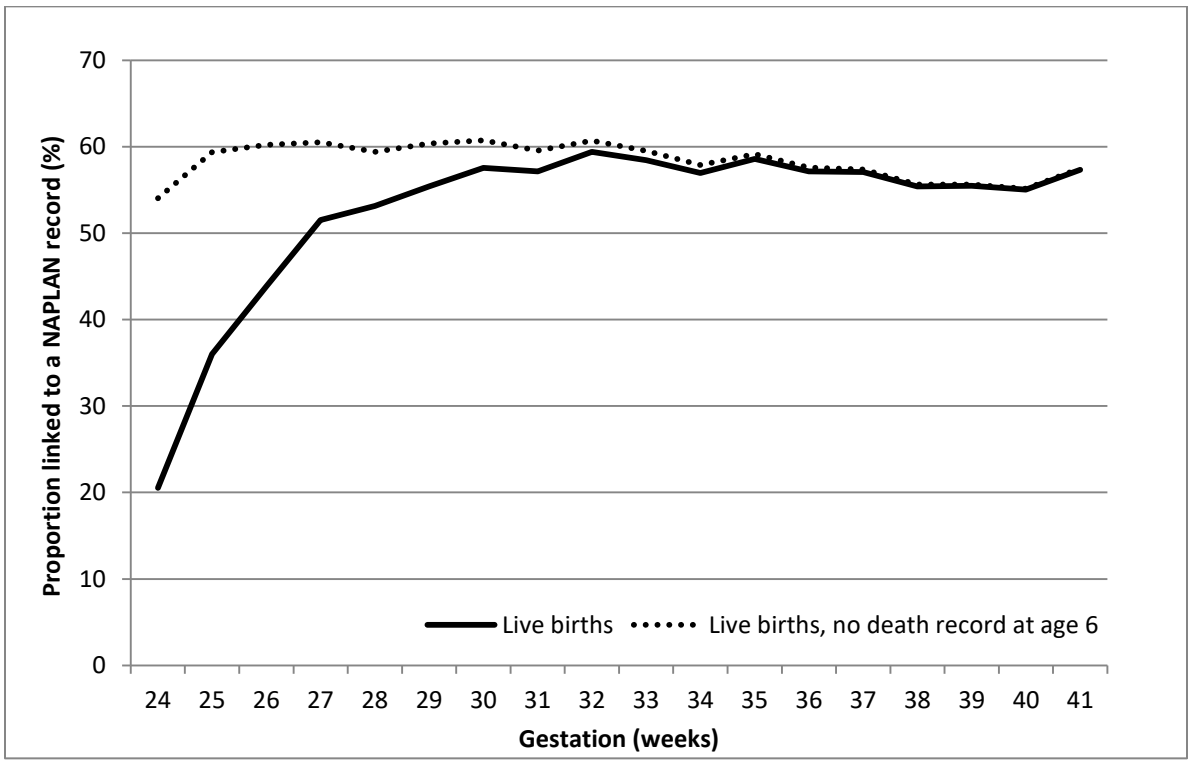

Table 1. NAPLAN linkage rates by gestational age at birth, NSW births 1994-2005

Gestational age at birth (weeks)	Live births n=1,010,018	Deaths before age 6 (% mortality) n=4,558	Live born, no death record n=1,005,460	Linked NAPLAN records	% of live births linked to NAPLAN record	% of children alive at 6 years linked to NAPLAN record
24	521	323 (62.0)	198	107	20.5	54.0
25	625	246 (39.4)	379	225	36.0	59.4
26	839	228 (27.2)	611	368	43.9	60.2
27	862	128 (14.8)	734	444	51.5	60.5
28	1195	126 (10.5)	1069	635	53.1	59.4
29	1303	107 (8.2)	1196	722	55.4	60.4
30	1786	93 (5.2)	1693	1028	57.6	60.7
31	2258	92 (4.1)	2166	1290	57.1	59.6
32	3472	74 (2.1)	3398	2062	59.4	60.7
33	4880	88 (1.8)	4792	2852	58.4	59.5
34	8370	133 (1.6)	8237	4767	57.0	57.9
35	13374	121 (0.9)	13252	7837	58.6	59.1
36	26537	213 (0.8)	26324	15161	57.1	57.6
37	55982	287 (0.5)	55695	31939	57.1	57.3
38	156378	572 (0.4)	155806	86656	55.4	55.6
39	227739	603 (0.3)	227136	126369	55.5	55.6
40	344617	748 (0.2)	343869	189628	55.0	55.2
41	159280	375 (0.2)	158905	91332	57.3	57.5

Of the children with a linked school record, the proportion who sat the numeracy test increased with increasing gestational age at birth (Table 2). At 24 weeks gestation, 72.9% of the linked records indicated that the child sat the test, and at 41 weeks it was 95.3%. This increase in the proportion of children who sat the test was explained by the tendency for the proportions of children who were exempt, withdrawn, or absent, to be higher at the earlier gestational ages (Figure 2). Exemptions were most common in children born at 24 and 25 weeks (15.9% and 17.4%). This proportion then dropped sharply to 8.7% at 26 weeks and from there decreased gradually to a minimum of 1.2% for children born at 40 and 41 weeks. This indicates that intellectual or physical disability severe enough to prevent a child from being able to participate in testing even with support was more common in children born at the earliest gestational ages.

The proportion absent from school on the day of testing was also highest at 25 weeks (6.5%) and similarly showed a generally decreasing trend over the gestational age range, to 3.0% at 41 weeks (Table 2). The proportion of children withdrawn from testing was generally low, but also showed a decreasing trend. The highest proportion of children withdrawn was seen in those born at 24 weeks (4.7%). The same pattern was seen for participation in reading testing (not shown).

There was a strong association between gestational age at birth and achievement on testing, with children born at earlier gestational ages more likely to have a low score, whether by performing poorly on the test or by being exempt due to disability (Table 1). This pattern held for both measures of low score: the proportion below the national minimum standard decreased from 28.9% to 5.7% for numeracy, and 32.3% to 7.1% for reading. The proportion with a score less than one standard deviation below the mean decreased from 50.5% to 15.1% for numeracy, and 47.5% to 16.0% for reading.

Table 2. Proportion of children that were present, absent, exempt, or withdrawn from testing on first linked government school NAPLAN record

Gestation at birth (weeks)	Children with linked government school NAPLAN result n=563412	Present n=533485 (94.7%)		Absent n=18691 (3.3%)		Exempt n= 8584 (1.5%)		Withdrawn n=2652 (0.5%)	
		n	%	n	%	n	%	n	%
24	107	78	72.9	7	6.5	17	15.9	5	4.7
25	225	173	76.9	11	4.9	39	17.3	2	0.9
26	368	311	84.5	17	4.6	32	8.7	8	2.2
27	444	399	89.9	9	2.0	30	6.8	6	1.4
28	635	556	87.6	27	4.3	43	6.8	9	1.4
29	722	625	86.6	35	4.8	53	7.3	9	1.2
30	1028	904	87.9	58	5.6	52	5.1	14	1.4
31	1290	1178	91.3	46	3.6	51	4.0	15	1.2
32	2062	1869	90.6	84	4.1	86	4.2	23	1.1
33	2852	2627	92.1	113	4.0	87	3.1	25	0.9
34	4767	4405	92.4	181	3.8	137	2.9	44	0.9
35	7837	7297	93.1	285	3.6	199	2.5	56	0.7
36	15161	14100	93.0	580	3.8	373	2.5	108	0.7
37	31939	29975	93.9	1147	3.6	648	2.0	169	0.5
38	86656	81932	94.5	2885	3.3	1442	1.7	397	0.5
39	126369	120127	95.1	3892	3.1	1800	1.4	550	0.4
40	189628	179897	94.9	6574	3.5	2355	1.2	802	0.4
41	91332	87040	95.3	2740	3.0	1141	1.2	411	0.5

Figure 2. Proportion of children with linked school records that were absent, exempt, or withdrawn from numeracy testing, by gestational age at birth

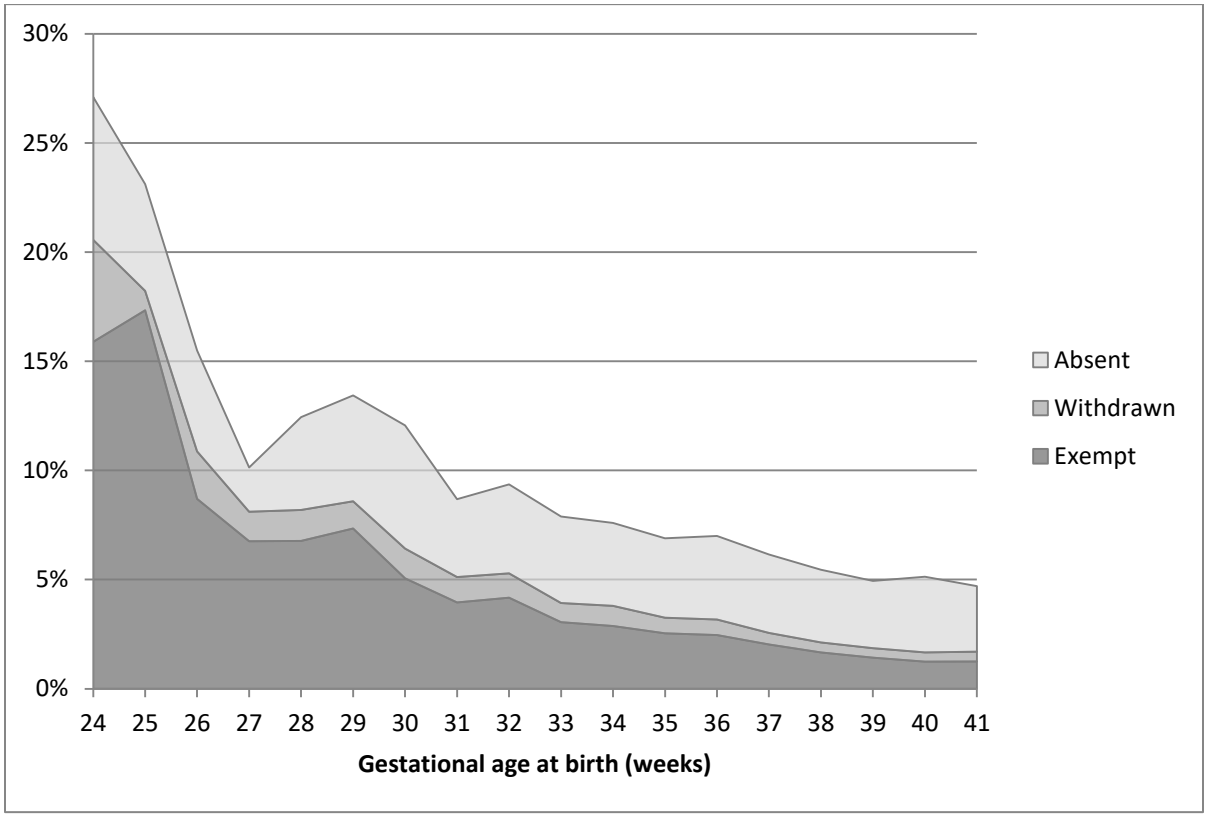

Table 3. Proportion of children who scored below the national minimum standard and below one standard deviation below the mean, by gestational age at birth

	Numeracy					Reading				
Gestational age (weeks)	Children with a result	Below NMS		Less than mean-1SD		Children with a result	Below NMS		Less than mean-1SD	
		n	%	n	%		n	%	n	%
24	97	28	28.9	49	50.5	99	32	32.3	47	47.5
25	219	65	29.7	102	46.6	216	59	27.3	83	38.4
26	355	81	22.8	155	43.7	355	76	21.4	129	36.3
27	439	77	17.5	144	32.8	439	74	16.9	138	31.4
28	613	112	18.3	212	34.6	619	98	15.8	170	27.5
29	699	126	18.0	217	31.0	702	131	18.7	217	30.9
30	982	133	13.5	266	27.1	992	136	13.7	256	25.8
31	1261	152	12.1	341	27.0	1264	174	13.8	325	25.7
32	2000	252	12.6	560	28.0	2013	301	15.0	564	28.0
33	2781	320	11.5	673	24.2	2788	333	11.9	671	24.1
34	4643	429	9.2	991	21.3	4659	497	10.7	1004	21.5
35	7625	675	8.9	1522	20.0	7650	755	9.9	1602	20.9
36	14751	1276	8.7	2940	19.9	14815	1554	10.5	3159	21.3
37	31161	2503	8.0	5869	18.8	31252	2839	9.1	6076	19.4
38	84665	5589	6.6	14024	16.6	84940	6822	8.0	14998	17.7
39	123782	7585	6.1	19071	15.4	124200	9219	7.4	20291	16.3
40	185123	11198	6.0	28460	15.4	185881	14239	7.7	31731	17.1
41	89536	5142	5.7	13512	15.1	89792	6342	7.1	14395	16.0

Summary:

- Linkage rate as a proportion of all live births increases with increasing gestational age at birth, due to higher mortality at lower gestational ages
- The overall linkage rate for children who lived to age six was 56%
- For children who lived to age six, linkage rates are highest in births occurring at 26-33 weeks gestation, and lowest in births occurring at 24 and 40 weeks
- Birth records that do not link to a government school NAPLAN record may be due to:
    - Death before school age
    - Child attending a non-government school
    - Child not attending school at all
    - Child moving out of the state before school age
- Of children who attended government schools, the proportion who sat the NAPLAN tests increased over the gestational age range

    - Around 95% of children born at term sat the tests, compared to 73% of those born at 24 weeks
    - Exemptions due to serious or complex disability were highest at 24-25 weeks and decreased with increasing gestational age
    - The proportion of children absent and withdrawn also decreased with increasing gestational age
- There was a strong association between gestational age and obtaining a low score on standardised testing for both numeracy and reading. A decreasing proportion of children received low scores as gestational age at birth increased, for both numeracy and reading.

References:

1. McCormick, M.C., et al., *Prematurity: an overview and public health implications.* Annu Rev Public Health, 2011. **32**: p. 367-79.
2. Australian Curriculum Assessment and Reporting Authority, *National assessment program literacy and numeracy: 2014 technical report*. 2015, ACARA: Sydney.
3. National Assessment Program. *NAPLAN - student participation*. 10/6/16]; Available from: http://www.nap.edu.au/naplan/school-support/student-participation.
4. Dobbins, T.A., et al., *Australian national birthweight percentiles by sex and gestational age, 1998-2007.* Med J Aust, 2012. **197**(5): p. 291-4.
5. Tukey, J.W., *Exploratory data analysis*. 1977, Reading, Massachusetts: Addison-Wesley.
6. Centre for Health Record Linkage. *Quality assurance*. [cited 2016 9 August]; Available from: http://www.cherel.org.au/quality-assurance.
7. Centre for Education Statistics and Evaluation, *2012 statistical bulletin: schools and students in NSW*, in *CESE Statistical Bulletin*, NSW government office of education, Editor. 2012.

Acknowledgements

We thank the Ministry of Health for provision of population data and the NSW Centre for Health Record Linkage for record linkage. This work was supported by an NHRMC project grant (APP1085775). Daneeta Hennessy, a NSW Biostatistics trainee, was funded through a NSW Ministry of Health 'Population Health and Health Services Research Support Program' grant. Christine Roberts is funded by a NHMRC Senior Research Fellowship (#APP1021025).