

BMJ Open New South Wales Child Development Study (NSW-CDS): an Australian multiagency, multigenerational, longitudinal record linkage study

Vaughan J Carr,^{1,2,3} Felicity Harris,^{1,2} Alessandra Raudino,^{1,2} Luming Luo,^{1,2} Maina Kariuki,^{1,2} Enwu Liu,^{1,2} Stacy Tzoumakis,^{1,2} Maxwell Smith,⁴ Allyson Holbrook,⁴ Miles Bore,⁵ Sally Brinkman,^{6,7,8} Rhoshel Lenroot,¹ Katherine Dix,⁹ Kimberlie Dean,^{1,10} Kristin R Laurens,^{1,2} Melissa J Green^{1,2}

To cite: Carr VJ, Harris F, Raudino A, *et al*. New South Wales Child Development Study (NSW-CDS): an Australian multiagency, multigenerational, longitudinal record linkage study. *BMJ Open* 2016;**6**: e009023. doi:10.1136/bmjopen-2015-009023

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-009023>).

Received 9 June 2015
Revised 3 November 2015
Accepted 30 November 2015



CrossMark

For numbered affiliations see end of article.

Correspondence to

Professor Vaughan Carr;
v.carr@unsw.edu.au

ABSTRACT

Purpose: The initial aim of this multiagency, multigenerational record linkage study is to identify childhood profiles of developmental vulnerability and resilience, and to identify the determinants of these profiles. The eventual aim is to identify risk and protective factors for later childhood-onset and adolescent-onset mental health problems, and other adverse social outcomes, using subsequent waves of record linkage. The research will assist in informing the development of public policy and intervention guidelines to help prevent or mitigate adverse long-term health and social outcomes.

Participants: The study comprises a population cohort of 87 026 children in the Australian State of New South Wales (NSW). The cohort was defined by entry into the first year of full-time schooling in NSW in 2009, at which time class teachers completed the Australian Early Development Census (AEDC) on each child (with 99.7% coverage in NSW). The AEDC data have been linked to the children's birth, health, school and child protection records for the period from birth to school entry, and to the health and criminal records of their parents, as well as mortality databases.

Findings to date: Descriptive data summarising sex, geographic and socioeconomic distributions, and linkage rates for the various administrative databases are presented. Child data are summarised, and the mental health and criminal records data of the children's parents are provided.

Future plans: In 2015, at age 11 years, a self-report mental health survey was administered to the cohort in collaboration with government, independent and Catholic primary school sectors. A second record linkage, spanning birth to age 11 years, will be undertaken to link this survey data with the aforementioned administrative databases. This will enable a further identification of putative risk and protective factors for adverse mental health and other outcomes in adolescence, which can then be tested in subsequent record linkages.

Strengths and limitations of this study

- The sample is a multigenerational, population cohort of approximately 87 000 Australian children, representative of 99% of children in the state of New South Wales entering their first year of formal education in 2009.
- The use of record linkage methodology to combine multiagency administrative data collections limits selection and participation bias and loss to follow-up, but may also be limited in depth and accuracy of information.
- The available data on parental history of mental and physical illness and criminal offending permit the investigation of children at familial risk of developing mental illness and other adverse health and social outcomes, as well as resilience to these outcomes.
- This large sample size offers opportunities to identify different early developmental pathways of risk and resilience, and affords sufficient power to determine the relationships between relatively rare exposures and outcomes.

INTRODUCTION

The relatively high prevalence in childhood of both clinical and subclinical mental health difficulties in Australia, alongside low service utilisation,¹ calls for a population-based approach to childhood mental health promotion. This should be augmented by early intervention and prevention programmes that target vulnerable children, but which are not limited to those presenting with overt clinical symptoms or established diagnoses. Recent estimates indicate that major depressive disorder, self-harm, anxiety disorder and violence are 4 of the top 10 causes of global burden of disease and injury among individuals aged 15–24 years,² with a

quarter of global disability attributable to mental health, and substance use disorders in individuals aged 0–24 years.³ Among Australians of this age, psychotic and mood disorders contribute almost two-thirds of the total burden of disease due to mental illness, and violence against self or others contributes one-third of the total burden of injury.⁴ Between a quarter and two-fifths of these disorders in adulthood could be prevented by effective early intervention for juvenile mental health problems.⁵ Preventative interventions are therefore necessary as soon as, or even before, identifiable risk characteristics in childhood emerge.

The central questions to be addressed at the population level in this context include: (1) what is the most reliable and efficient way of identifying childhood patterns of risk (and resilience) for adverse mental health and related outcomes in later childhood and/or adolescence; (2) what universal prevention and early intervention policies most effectively reduce or mitigate high risk for later adverse outcomes; (3) what targeted interventions are most effective for groups at high risk for mental ill health, and how can they be deployed in a way that avoids stigmatisation and damage to self-esteem? The present study aims to address the first of these questions, and provide a foundation to help inform the second and third. There are two types of factors that affect risk, those that increase risk or likelihood of adverse outcomes, which are referred to as vulnerability factors, and those that reduce risk, namely protective factors. The New South Wales Child Development Study (NSW-CDS) seeks to identify both of these at a population level so that interventions designed to reduce vulnerability can be considered in combination with those that increase protection.

The NSW-CDS (<http://www.nsw-cds.com.au>) adopts a life course epidemiological approach to examine associations at a population level between various indices of biological and environmental exposures (eg, perinatal complications, child maltreatment, parental mental illness or parental criminal history), and a range of indices of psychosocial adjustment in later childhood, adolescence and young adulthood. It combines multi-agency, multigenerational record linkage methodology with cross-sectional survey information obtained at ages 5 and 11 years, and takes a longitudinal perspective by means of successive waves of record linkage. The NSW-CDS cohort thus provides an unprecedented opportunity to examine the complex relationships between various exposures, individual characteristics and later development at multiple time points in a large population cohort.

COHORT DESCRIPTION

The State of NSW comprises 32% of the Australian population;⁶ it is the most populous state in Australia, with an ethnically diverse population of around 7 million inhabitants, of which the majority

(approximately 63%) reside in Sydney, the largest city in Australia.⁷ In 2009, teachers in government and private education sectors completed a national survey for the first time, the Australian Early Development Census (AEDC). This included all children entering their first year (Kindergarten) of full-time formal schooling at approximately 5 years of age (N=87 170), representing 99.9% of the eligible NSW children in 2009. The NSW-CDS child cohort (N=87 026) was defined from this original AEDC sample, with the exclusion of 0.9% of the NSW AEDC cohort for whom either a catch-up assessment was completed in 2010, or duplicate AEDC records existed.⁸

The AEDC (previously referred to as the Australian Early Development Index) was conducted using the Australian revision of the Canadian Early Development Instrument,⁹ and was completed by teachers on the basis of at least 1 month's knowledge of the child. It measures school readiness in five developmental domains: physical health and well-being, social competence, emotional maturity, language and cognitive development, and communication skills and general knowledge.⁹ The AEDC has satisfactory construct and concurrent validity,¹⁰ and the Australian Government has committed to collecting the census data on school entry every 3 years. Aggregated data are publicly available, and microdata for use in record linkage studies can be accessed at <http://www.AEDCdata.com.au>. A list of individual items available under each developmental domain can be found in Brinkman *et al.*¹¹

A summary of the sociodemographic characteristics of the NSW-CDS child cohort (N=87 026), defined by inclusion in the AEDC of 2009 in NSW, is presented alongside Australian Census data available for a comparable NSW and national age group (5–9 years) in [table 1](#). This demonstrates the comparability of the NSW-CDS cohort to the state and national population distributions of sex, socioeconomic index of areas, and areas of accessibility and remoteness.^{12 13} The NSW-CDS child cohort may thus be considered representative of the NSW and Australian populations of comparable age.

In 2013, the AEDC cohort was linked to several administrative data sets as detailed below. These included the children's birth, mortality, health, school and child protection records, their mothers' perinatal records, and both parents' mortality, health and criminal records. The record linkage was conducted by an independent agency, the Centre for Health Record Linkage (CHeReL: <http://www.cherel.org.au/>) using ChoiceMaker software (Choice Maker Technologies Inc.) to facilitate probabilistic record linkage methods that ensure strict privacy protocols are adhered to. Matching variables included name, date of birth, residential address and sex, and were obtained for each of the data sets. Definite and possible matches between these data sets were identified using 'blocking' and 'scoring', with 0.75 and 0.25 probability cut-off limits employed to ensure false positive links were minimised

(ie, all pairs of records with probabilities above the upper cut-off were designated as ‘true matches’, whereas all pairs of records with probabilities below the lower cut-off were designated as ‘false matches’, and clerical reviews were performed on all pairs with probabilities between the cut-off limits). At the completion of the linkage, a project-specific Person ID was assigned to allow linked records for the same individual to be identified and extracted. No content data (eg, health information) was used in the linkage process. Instead, each data custodian extracted the approved data and provided the researchers with a de-identified unit record file numbered by the project-specific Person ID, which allowed the researchers to combine the multiple data sets. In addition to the privacy protection afforded by the record linkage methodology, restrictions on the nature of data items available to the research team, as well as restrictions on the provision of geographical and calendar data, help ensure that individual participants cannot be identified.

Ethical approval for the research was obtained from the NSW Population and Health Services Research Ethics Committee (HREC/11/CIPHS/14), and the University of New South Wales Human Research Ethics Committee (HC11409), with data custodian approvals

granted by the relevant Government Departments. The Australian National Health and Medical Research Council (NHMRC) National Statement of Ethical Conduct in Human Research (Chapter 2.3) enables a waiver of consent to be enacted for the purpose of record linkage research, where stringent privacy and anonymity procedures are followed, and where there is a perceived public good; these guidelines are consistent with Australian and NSW privacy and information legislation.¹⁵

Child cohort

AEDC data were linked to: (1) birth and mortality data derived from the NSW Registry of Births, Deaths and Marriages—Birth Registrations and Mortality records, (2) education data from the NSW Department of Education Best Start Kindergarten Assessment records (public education sector only), (3) Case Management System (KiDS) provided by the NSW Department of Family and Community Services—including Child Protection Substantiations, Child Out of Home Care and Brighter Futures records and (4) health records from the NSW Ministry of Health’s Perinatal, Emergency Department, and Admitted Patients Data Collections. The linked data covered the period from birth to age 5 years.

Table 1 A comparison of demographic characteristics between the NSW-CDS cohort and Australian Census data

	NSW-CDS child cohort		General ‘comparable’ population*	
	n	Per cent	NSW, %	National, %
Age, years				
<5	4132	4.75		
5–6	68 275	78.45		
≥6	14 619	16.70		
Gender				
Male	44 729	51.40	51.50	51.40
Female	42 297	48.60	48.50	48.60
ARIA†				
Major cities	62 556	71.88	72.81	72.90
Inner regional	16 674	19.16	20.05	20.24
Outer regional	6771	7.78	6.48	6.33
Remote	890	1.02	0.51	0.46
Very remote	135	0.16	0.15	0.06
SEIFA Index for Relative Socio-Economic Disadvantage‡				
Quintile 1 (greatest disadvantage)	20 951	24.07	20.00	20.00
Quintile 2	19 336	22.22	20.00	20.00
Quintile 3	12 489	14.35	20.00	20.00
Quintile 4	12 201	14.35	20.00	20.00
Quintile 5 (least disadvantage)	22 034	25.02	19.00	20.00

*Comparative population of children aged 5–9 years in 2011 (NSW-CDS child cohort approximately 7 years in 2011) derived from the Australian Bureau of Statistics (ABS) and Australian Institute of Health and Welfare.^{6–8 12 13}

†This index was commissioned by the former Department of Health and Aged Care, and uses geographic information systems to summarise a community’s level of remoteness based on the accessibility of services (derived from measures of road distances between populated localities and service centres).¹³

‡The indices were developed by the ABS as a set of measures derived from census information that summarise the socioeconomic conditions of an area. The postcode of residence of the child recorded in the AEDC was matched to an ABS State Suburb (SSC), and the corresponding SEIFA score for the SSC. Quintiles for SEIFA scores reported in the 2009 AEDC data set are based on ABS SEIFA deciles for the 2006 census.¹¹ Quintiles are based on National AEDC data (ie, created on the basis of all children who participated in the 2009 AEDC nationally).¹⁴

ARIA, Accessibility/Remoteness Index of Australia; NSW-CDS, New South Wales Child Development Study; SEIFA, Socioeconomic Indexes for Areas.

Parents of the child cohort

Parents were identified through linkage of the children's AEDC records, with birth registration data held in the NSW Registry of Births, Deaths and Marriages; mothers were identified for 81.6% (N=72 796) and fathers for 81.5% (N=72 778) of the AEDC sample. For children who had an NSW birth registration record, linkage identified 98.09% of mothers and 98.07% of fathers. Child cohort members without a matched NSW birth registration record were born outside of NSW (16.9%), either elsewhere in Australia or overseas. The sample was then cleaned to remove duplicate AEDC records and any AEDC records which were part of a 'recovery' collection in 2010. Following data cleaning, mother and father records were available for 72 245 (83.0%) children in the child cohort, of which there were 71 076 individual mothers and 71 039 individual fathers. Having identified that a substantial proportion of the child cohort had not been born in NSW, and would thereby be excluded from future studies using linked parental records, we compared the sociodemographic characteristics of the whole cohort to those without an NSW registered birth, as well as Australian and NSW population estimates for children of the same age; no major group differences were detected (see online supplementary table 1-X).

Identified mothers and fathers were linked to records derived from the (1) Registry of Births, Deaths and Marriages—Mortality records, (2) health records provided by the NSW Ministry of Health's Mental Health Ambulatory, Emergency Department, and Admitted Patients Data collections and (3) criminal offending records derived from the NSW Bureau of Crime Statistics and Research reoffending records, including data from Drug, Local, District and Supreme Criminal Courts, and Corrective Services.

FINDINGS TO DATE

Optimal linkage rates were achieved (table 2), with a false positive rate of 0.03% and 0.5% for the child cohort and their parents, respectively. Table 2 outlines the linkage rates for each data set and the retained sample following data cleaning across child, mother and father subgroups.

Sociodemographic information

Table 3 presents sociodemographic and other characteristics of the 87 026 members of the child cohort (48.6% female). At the time of the AEDC in 2009, the mean age of participants was 5.75 years, with an SD of 0.39 (males: M=5.78, SD=0.40; females: M=5.71, SD=0.38). The top five countries of children's birth included: Australia 94.1%, England 0.7%, New Zealand 0.7%, India 0.6% and the USA 0.3%. The socioeconomic distribution of area of residence for the cohort members was similar to the distribution reported for the national AEDC sample.⁸

Child development

Scores on the AEDC provide an indication of early childhood development on five domains of functioning, as described above. For each domain, a child received a score between 0 and 10, with higher scores indicating better developmental functioning. Performance on each of the domains was also expressed categorically, with children falling in the bottom 10% of the distribution classified as developmentally 'vulnerable'; children who scored in the 10th–25th centile as developmentally 'at risk.' Individual domain scores in the cohort were comparable to the national distribution of scores, with 5.9–9.2% classified as 'developmentally vulnerable' (0–10th centile), 9.5–15.8% classified as 'developmentally at risk' (11th–25th centile), and 77.2–78.5% of the children classified as developmentally 'on track' (>25th centile).¹⁰ The distribution of scores in the language and cognitive skills domain was slightly higher than the national average, with 84.6% of the cohort classified as 'on track', compared with 77.1% of the national sample. AEDC domain and subdomain percentile and vulnerability distributions for the whole child cohort and the subcohort with linked parental records are provided in the online supplementary table 2-X. Those with linked parental records uniformly showed slightly lower rates of vulnerability on AEDC domain and subdomain scores. Because AEDC domain scores are not provided for children with special needs (ie, children who require special assistance in the classroom due to a chronic medical, physical, or intellectually disabling condition), we additionally present demographic data for the cohort with these children removed from the total cohort, and the subcohort with parental linked data (see online supplementary table 1-X).

Child educational attainment

The Best Start Kindergarten Assessment (BSKA) was available for 44.8% of the child cohort (the assessment was not conducted outside the public education system). Literacy included seven dimensions and numeracy included four dimensions, listed in table 3. Scores across dimensions were standardised to a range 0–3, in which 0 indicated normal or expected performance on school entry, and 1–3 indicated incremental performance increases above what is expected on school entry. In our cohort, the majority of children achieved an expected level of proficiency: 48.2% of children obtained a score of 0 in literacy and 43.1% scored 0 in numeracy, with 10% demonstrating very high proficiency in early literacy and numeracy competence (see table 3).

Child protection (2000–2009)

There were 3822 cohort members (4.4%) with a record of child protection involvement (see table 3). This included children with at least one report where actual harm or risk of significant harm was determined (N=3078; 80.5%). Additional data collections provided information on the number of cohort members who were

Table 2 Multiagency data collection record linkage rates and retained sample following cleaning

Data collection	Years	Children			Mothers			Fathers		
		Linkage rate	Retained		Linkage rate	Retained		Linkage rate	Retained	
		Per cent	N/n	N/n	Per cent	N	N/n	Per cent	N	N/n
Early development										
Australian Early Development Census	2009		89 268	87 026						
Vital events										
NSW Registry of Births, Deaths and Marriages: Birth Registrations data	2000–2006	83.1	74 293	72 245		72 796	72 245		72 778	72 245
NSW Registry of Births, Deaths and Marriages: Death Registrations data	2000–2009*	0.0	2	2	0.2	108	109	0.5	363	369
Australian Bureau of Statistics Mortality data	2000–2007	0.0	0	0	0.1	64	64	0.3	223	223
Health										
NSW Ministry of Health Perinatal Data Collection	2000–2006	83.9	74 930	73 056	99.2	72 213	71 663			
NSW Ministry of Health Emergency Department Data Collection	2005–2009	61.2	54 598	53 184	44.1	32 068	31 814	43.3	31 529	31 309
NSW Ministry of Health Admitted Patients Data Collection	2000–2009	86.6	77 313	75 391	99.4	72 376	71 824	49.9	36 341	36 170
NSW Ministry of Health Mental Health Ambulatory Data Collection	2000–2009				7.3	5308	4629	4.5	3246	2854
Education										
NSW Department of Education Best Start Kindergarten Assessment data (Government schools only)	2009	44.8	40 035	40 032						
Child protection										
NSW Department of Family and Community Services Case Management System (KiDS) (Child Protection, Out-of-home-care and Brighter Futures)	2003–2009†	4.4	3929	3822						
Criminal offending										
Bureau of Crime Statistics and Research data	1994–2009				10.1	7350	6180	29.3	21 329	18 540

*NSW Registry of Births, Deaths and Marriages Death Registrations data for the child cohort is for 2009 only.

†Brighter Futures data for the child cohort is available from 2004.

NSW, New South Wales.

Table 3 Selected characteristics of the NSW-CDS child cohort

	n	Per cent
Child developmental vulnerability (N=87 026)		
Physical health and well-being*	7166	8.6
Social competence*	7268	8.8
Emotional maturity*	6134	7.4
Language and cognition*	4848	5.9
Communication skills and general knowledge*	7589	9.2
Child educational attainment: expected level (N=40 032)		
<i>Literacy</i> : Phonics	16 846	42.8
<i>Literacy</i> : Phonemic awareness	23 251	59.6
<i>Literacy</i> : Comprehension	19 063	50.2
<i>Literacy</i> : Aspects of speaking	10 858	27.7
<i>Literacy</i> : Aspects of writing	30 867	79.9
<i>Literacy</i> : Reading texts	21 634	55.0
<i>Literacy</i> : Concepts about print	18 241	46.3
<i>Numeracy</i> : Pattern repeated unit	3668	9.2
<i>Numeracy</i> : Forward number word sequences	4130	10.4
<i>Numeracy</i> : Numerical identification	16 954	42.7
<i>Numeracy</i> : Early arithmetic strategies	16 956	42.8
Child protection (N=3822)		
Emotional/psychological abuse	1648	1.9
Neglect	1189	1.4
Physical abuse	704	0.8
Sexual abuse	386	0.4
Out-of-Home-Care	1143	1.3
Targeted intervention programme: Brighter Futures	987	1.1
Child perinatal health (N=73 056)		
Any smoking during pregnancy	10 870	14.9
Maternal hypertension	801	1.1
Pre-eclampsia	4130	5.7
Maternal diabetes mellitus	439	0.6
Gestational diabetes	3338	4.6
Apgar score at 1 min†7~10	65 606	90
Apgar score at 5 min†7~10	72 088	98.9
Birth weight (g)†		
<1500	523	0.7
1500–2899	11 156	15.3
2900–3499	29 638	40.6
3500–4499	30 490	41.7
4500 and over	1249	1.7
Maternal age (years)†		
≤24	13 311	18.2
25–29	20 537	28.1
30–34	24 694	33.8
≥35	1414	19.9

*Children who score in the 0–10th centile are classified as 'developmentally vulnerable', indicating much lower than average AEDC scores.

†The variable has been categorised from the original continuous variable in the data set.

AEDC, Australian Early Development Census; NSW-CDS, New South Wales Child Development Study.

programme for families with children at risk of abuse and/or neglect that started in 2004 (see table 3).

Child health

Perinatal health (1999–2006): Mean maternal age at birth, and mean birth weight, were in line with national norms.^{16 17} One in 10 cohort members were given a low Apgar score (<7) at 1 min after birth, whereas 1.1% had a low Apgar score at 5 min.¹⁸ The most common recorded maternal health problem during pregnancy was pre-eclampsia (5.7% of mothers) (see table 3).

Hospital admissions (2000–2009): The number of admissions for each child ranged from 1 to 255 (median (Mdn)=3.08). Among the 75 391 children admitted to hospitals, 30 336 (40.2%) had only a single record of admission comprising the birth event, with no additional admissions or diagnoses. Prior to 12 months of age, excluding the birth event, the most common diagnoses were conditions originating from the perinatal period (n=23 326, 31.9%), which include birth trauma, and disorders related to length of gestation and fetal growth. These accounted for 72.4% of all (non-birth event) diagnoses across hospital admissions during that period. After 12 months of age, the most common diagnoses were diseases of the respiratory system (n=12 850, 17.6%), accounting for 28.2% of all diagnoses. Further details regarding the most prevalent diagnoses under each ICD10-AM chapter block are provided in the online supplementary table 3-X.

Emergency Department presentations (2000–2009): The number of presentations to the hospital emergency departments for each child ranged from 1 to 76, with a total of 53 184 (61.1%) children presenting at least once. The most common reasons for emergency department presentation were otitis media unspecified (4.1%), open wound of other parts of head (7.1%), fever unspecified (7.6%), acute upper respiratory infection unspecified (9.6%), viral infection unspecified (10.1%), and special screening examination unspecified (49.2%). Further details regarding the most prevalent diagnoses under each ICD10-AM chapter block are provided in the online supplementary table 4-X.

Parental health

Hospital admission (2000–2009): There were 71 824 (99.4%) mothers and 36 170 (50.1%) fathers with a reported hospital admission, of which 70 501 (99.2%) of mothers had birth-related hospital admissions, and 38 079 (53.6%) of mother had a non-birth-related hospital admission. The number of admissions for mothers ranged from 1 to 302 (Mdn=4.7) per person, and for fathers from 1 to 1027¹ (Mdn=12) per person. The most frequent reasons for admission of mothers, aside from those related to pregnancy, childbirth and the puerperium, were diseases of the digestive and genitourinary

¹One male patient was admitted 1027 times for renal dialysis.

or had been placed in out-of-home care, and the proportion of children and their families who were enrolled in the Brighter Futures programme, an early intervention

systems. For fathers the most common reasons for admission were diseases of the digestive system, and injuries or poisonings. Further details regarding the most prevalent diagnoses under each ICD10-AM chapter blocks are provided in the online supplementary table 3-X.

Emergency department presentations (2000–2010): There were 31 814 mothers (44.0%) and 31 309 fathers (43.3%) with a reported emergency department presentation. The number of emergency department presentation events for mothers ranged from 1 to 294 (Mdn=2), and for fathers from 1 to 129 (Mdn=2). Classifications were assigned using both ICD9 and ICD10 diagnostic codes, with the migration timetable to V.10 differing across hospitals. In this paper, for descriptive purposes, both ICD versions are reported in the online supplementary table 5-X.

Mental Health Ambulatory (2001–2010): There were 4629 mothers (6.4%) and 2854 fathers (4.0%) with a Mental Health Ambulatory record.ⁱⁱ The number of contact events with the mental health ambulatory services for mothers ranged from 1 to 4044 (Mdn=96), and for fathers from 1 to 4938 (Mdn=94). The most frequent contact events for mothers were for depression (n=33 188 events; n=1104 mothers) and schizophrenia (n=18 958 events; n=184 mothers). The same pattern was evident for fathers: the most frequent contact events were for depression (n=19 158 events; n=547 fathers) and schizophrenia (n=15 485 events; n=202 fathers).ⁱⁱⁱ

Parental criminal offending (2000–2010)

There were 6180 mothers (8.6%) and 18 540 fathers (25.7%) with a report in the 'offense/appearance' records of the NSW Bureau of Crime Statistics. The type of offence was classified using the 16 categories in the Australian Standard Offence Classification (ASOC). Online supplementary table 6-X provides the number of children with a maternal and/or paternal history of offending, with all 16 ASOC categories represented. Online supplementary table 6-X shows, respectively, the number of children with a maternal and paternal history of the offences listed in the standard ASOC categories. The most frequent offence for both parents was 'traffic and vehicle regulation offences'.

Future directions

The NSW-CDS is a longitudinal study. A self-report survey of the children's mental health and well-being was implemented in the second half of 2015 when the

cohort was aged approximately 11 years. This was conducted in school class time, with assistance from 830 schools in NSW, and captured approximately 30.1% of the eligible child population. A second multiagency, multigenerational record linkage will be undertaken in early 2016 using the administrative data bases described above, spanning birth to age 11 years, and including the child mental health and well-being survey data. This will provide an opportunity to elucidate patterns of risk and resilience across early and middle child development, and will form the foundation upon which subsequent waves of record linkage will be conducted to provide information about health and other outcomes as the cohort moves into adolescence and early adulthood.

STRENGTHS AND LIMITATIONS

The main strengths of the NSW-CDS are the representative nature of the large population sample, the extensive linkage of multiagency, intergenerational (parent-child) data collections, and the use of independent informants (teachers' reports) of child functioning at approximately 5 years of age. The use of record linkage methodology enables an entire cross-section of the general population to be sampled with minimal selection bias, and allows for investigation of multiple factors contributing to risk and protection for outcomes of low prevalence and/or of relevance to minority groups (eg, indigenous Australians, remote communities, children with special needs). The capacity to map records from children to parents also provides a unique opportunity to conduct nested 'high-risk' substudies of the cohort where interactions between familial (eg, parental history of mental illness or criminal behaviours) and environmental risk and protective factors can be explored. The study thus affords a unique opportunity to investigate developmental pathways representing both risk of disorder and resilience to adversity, with respect to rare exposure and long-term outcomes that will be determined over time in future record linkages. This research is enabled by the investment of Federal and State Governments in Australia in providing the necessary record linkage infrastructure, ethical guidelines and specialist committee review, as well as privacy legislation to safeguard the use of individual data for research in a protected manner.

The use of sequential record linkages as the primary means of longitudinal follow-up is expected to minimise loss of participants in future phases, other than due to migration or mortality. Attrition rates are anticipated to reflect the average annual inward (interstate: n=162 535; international: n=144 100) and outward (interstate: n=267 907; international: n=93 000) overall migration rates in NSW,^{19 20} as well as loss within the education data associated with year-level repetition (8.4% of students over the typical 13 years of schooling in NSW),²¹ mortality (6.5/1000 for the total Australian population),²² and insufficient/incorrect identifiers for linkage.

ⁱⁱThe Mental Health Ambulatory data collection contains administrative data for public mental health services and does not incorporate services provided in the private sector, such as general practitioners and private psychiatrist and psychologists.

ⁱⁱⁱThe most frequent type of episodes/activities recorded in the Mental Health Ambulatory Data Collection for both mothers and fathers was: Mental Health Diagnosis not yet allocated or F99.1. This is a code provided in the ICD-10 and Mental Health Ambulatory Data Collection Dictionary when the diagnosis did not fall into the categories already identified, or the clinicians were unsure.

In terms of limitations, first, the research data obtained through record linkage involves information collected primarily for administrative purposes, potentially limiting the depth and accuracy of the information available. For example, there are no indicators of low (ie, 'below expected') performance on the BSKA, owing to the lack of requirement to meet any literacy or numeracy benchmarks at school entry; while this limits the utility of this indicator for studies of poor functioning at school entry, it may be useful in denoting children performing above expectation according to recent models of resilience.²³ Second, while comprehensive information is available within these repositories, it is possible that other important factors contributing to the development of risk and protective factors were not included. This limitation will be minimised in the second phase of the NSW-CDS, by supplementing administrative data with information gathered in the self-report survey of mental health and well-being at around 11 years of age. Third, intergenerational analyses in the future will not be possible for 16.9% of the cohort who were born outside of NSW, and for whom parents could not be identified from NSW birth records. Finally, a weakness of the first record linkage described here is the absence of information on the indigenous status of cohort participants. Permissions for accessing the indigenous indicator are being sought for future linkages.

COLLABORATION

Initial data analyses and publications will be generated primarily by those listed as authors on this paper, and others mentioned in the acknowledgements section as members of the scientific committee overseeing this project, together with their postgraduate research students. However, the research team is open to potential research collaborations with other scientists, with the proviso that analysis of linked data is currently authorised to occur at only one location, owing to ethical considerations in relation to relevant privacy legislation. In the first instance, potential researchers interested in collaboration should contact the first author (VC) with their expression of interest.

Author affiliations

¹School of Psychiatry, University of New South Wales, Sydney, New South Wales, Australia

²Schizophrenia Research Institute, Sydney, New South Wales, Australia

³Department of Psychiatry, Monash University, Melbourne, Victoria, Australia

⁴School of Education, University of Newcastle, Newcastle, New South Wales, Australia

⁵School of Psychology, University of Newcastle, Newcastle, New South Wales, Australia

⁶Telethon Kids Institute, Perth, Western Australia, Australia

⁷Centre for Child Health Research, University of Western Australia, Perth, Western Australia, Australia

⁸Australia Institute for Social Research, University of Adelaide, Adelaide, South Australia, Australia

⁹Principals Australia Institute, Flinders University, Adelaide, South Australia, Australia

¹⁰Justice Health & Forensic Mental Health Network, New South Wales, Australia

Acknowledgements This research was supported by the use of population data owned by the Department of Education and Training; NSW Department of Education; NSW Department of Family and Community Services; NSW Ministry of Health; NSW Registry of Births, Deaths and Marriages; the Australian Bureau of Statistics; and the NSW Bureau of Crime Statistics and Research. However the information and views contained in this study do not necessarily, or at all, reflect the views or information held by these Departments. The authors acknowledge the contributions of all members of the NSW-CDS scientific committee: Vaughan Carr (Chairman), Miles Bore, Sally Brinkman, Marilyn Chilvers, Kimberlie Dean, Katherine Dix, Sandra Eades, Stephanie Dick, Melissa Green, Felicity Harris, Allyson Holbrook, Maina Kariuki, Kristin Laurens, Rhoshel Lenroot, Luming Luo, Stephen Lynn, Caitlin McDowell, Alessandra Raudino, Maxwell Smith, Titia Sprague, Robert Stevens, Michael Tarren-Sweeney, Stacy Tzoumakis, and Anna Williamson.

Contributors In line with the ICMJE authorship guidelines, authors VJC, FH, LL, MS, AH, MB, SB, RL, KD, KRL and MJG made substantial contributions to the conception or design of the work. Authors VJC, FH, LL, KRL and MJG made substantial contributions to the acquisition of data. Authors VJC, FH, AR, LL, MK, EL, KD, KRL and MJG made substantial contributions the interpretation of data for the work. All authors contributed to the drafting of the manuscript, and/or the revising of the manuscript. All authors have given final approval of the version to be published, and agree to its accuracy.

Funding This work was supported by the Australian Research Council's Linkage Project funding scheme (project number LP110100150), with the NSW Ministry of Health, NSW Department of Education, and the NSW Department of Family and Community Services representing the Linkage Project Partners; and the Australian Rotary Health Project Grant funding scheme (project number RG104090). The 2015 survey of child mental health and well-being was supported by funding from an Australian National Health and Medical Research Council (NHMRC) Project Grant (1058652). MJG was supported by a NHMRC R.D. Wright Biomedical Career Development Fellowship (1061875); VC, KRL, AR, and FH were supported by the Schizophrenia Research Institute using an infrastructure grant from the NSW Ministry of Health. KD was supported by Justice Health & Forensic Mental Health Network, NSW.

Competing interests None declared.

Ethics approval NSW Population and Health Services Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The research team is open to potential research collaborations with other scientists, with the proviso that analysis of linked data is currently authorised to occur at only one location, owing to ethical considerations in relation to relevant privacy legislation. In the first instance, potential researchers interested in collaboration should contact the first author (VC) with their expression of interest.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Lawrence D, Johnson S, Hafekost J, *et al*. The Mental Health of Children and Adolescents. Report on the second Australian Child and Adolescent Survey of Mental Health and Wellbeing, Australia. Department of Health. 2015 (cited 2 Nov 2015). <http://www.health.gov.au/internet/main/publishing.nsf/Content/mental-pubs-m-child2>
2. Institute for Health Metrics and Evaluation [University of Washington]. *GBD 2010 arrow diagram: disability-adjusted life years 15–19 year and 20–24 years*. Washington, USA, 2013 (cited 30 Sep

- 2014). <http://www.healthmetricsandevaluation.org/gbd/visualizations/gbd-arrow-diagram>
3. Erskine HE, Moffitt TE, Copeland WE, *et al.* A heavy burden on young minds: the global burden of mental and substance use disorders in children and youth. *Psychol Med* 2015;45:1551–63.
 4. Begg S, Vos T, Barker B, *et al.* *The burden of disease and injury in Australia 2003*, Australia: Australia Institute of Health and Welfare, 2007 (cited 14 Sep 2014), Cat. no. PHE 82. <http://www.aihw.gov.au/publication-detail/?id=6442467990>
 5. Kim-Cohen J, Caspi A, Moffitt TE, *et al.* Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. *Arch Gen Psychiatry* 2003;60:709–17.
 6. Australian Bureau of Statistics. Australian Demographic Statistics, Sep 2014. 2013 (cited 13 Apr 2015). ABS Cat. no. 3101.0. <http://www.abs.gov.au/ausstats/abs@.nsf/mf/3101.0>
 7. Australian Bureau of Statistics. Regional Population Growth, Australia 2011–2012. 2013 (cited 12 Sep 2014). ABS Cat. no. 3218.0. <http://www.abs.gov.au/ausstats/abs@.nsf/mf/3218.0/>
 8. Centre for Community Child Health and Telethon Institute for Child Health Research [Australian Government]. A Snapshot of Early Childhood Development in Australia: AEDI National Report 2009, Australia. 2009 (updated May 2011; cited 12 Sep 2014). <http://www.aedc.gov.au/researchers/aedc-research/publications/publication/a-snapshot-of-early-childhood-development--australian-early-development-index-national-report-2009-%28re-issue---march-2011%29>
 9. Janus M, Offord DR. Development and psychometric properties of the early development instrument (EDI): a measure of children's school readiness. *Can J Behav Sci* 2007;39:1–22.
 10. Brinkman SA, Silburn S, Lawrence D, *et al.* Investigating the validity of The Australian Early Development Index. *Early Educ Dev* 2007;18:427–51.
 11. Brinkman SA, Gregory TA, Goldfeld S, *et al.* Data resource profile: the Australian Early Development Index (AEDI). *Int J Epidemiol* 2014;43:1089–96.
 12. Australian Bureau of Statistics. Socio-economic indexes for areas (SEIFA), Australia [Internet]. 2006 (cited 12 Sep 2014); ABS Cat. No. 2033.0.55.001. <http://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001/>
 13. Department of Health and Aged Care [Australian Government]. Measuring remoteness: Accessibility/Remoteness Index of Australia (ARIA), Australia [Internet]. Occasional Papers: New series No.14. 2001 (cited 2 Nov 2015) [http://www.health.gov.au/internet/main/publishing.nsf/Content/E2EE19FE831F26BFCA257BF001F3DFA/\\$File/ocpanew14.pdf](http://www.health.gov.au/internet/main/publishing.nsf/Content/E2EE19FE831F26BFCA257BF001F3DFA/$File/ocpanew14.pdf)
 14. Centre for Community Child Health and Telethon Institute for Child Health Research [Australian Government]. AEDI Technical Paper 2: Matching AEDI data to SEIFA, Australia [Internet]. 2009 (Updated May 2011; cited 12 Sep 2014). http://www.rch.org.au/aedi/Researchers/Technical_papers/
 15. National Health and Medical Research Council [Australian Government]. National Statement on Ethical Conduct in Human Research 2007, Australia. 2007 (Updated May 2015; cited 29 Oct 2015). https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72_national_statement_may_2015_150514_a.pdf
 16. Australian Institute of Health and Welfare. Aboriginal and Torres Strait Islander Health Performance Framework 2010 report, Australia. 2011 (cited 13 April 2015); AIHW Cat. No. IHW 66. <http://www.aihw.gov.au/publication-detail/?id=10737420657>
 17. Australian Bureau of Statistics. Australian Social Trends, 'Australia's Babies'. 2007 (cited 2014 September 12); ABS Cat. no. 4102.0. <http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4102.02007?OpenDocument>
 18. American Academy of Pediatrics, Committee on Foetus and Newborn, American College of Obstetricians and Gynaecologists. The Apgar score. *Pediatrics* 2006;117:1444–7.
 19. Karuppanan S. Population change and internal migration in Australia. State of Australian Cities Conference 2011, University of Melbourne, Melbourne, Australia. 2011 (cited 12 Sep 2014 Sep). http://soac.fbe.unsw.edu.au/2011/papers/SOAC2011_0116_final.pdf
 20. Australian Bureau of Statistics. State and Territory Statistical Indicators, Australia [Internet]. 2012 (cited 12 Sep 2014). ABS Cat. no. 1367.0. <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1367.0~2011~Main%20Features~Australia%20Home~1>
 21. Organisation for Economic Co-operation and Development. Program for International School Assessment: What Makes a School Successful 2009—Resources, Policies and Practices (Volume IV). France [Internet]. 2010 (cited 12 Sep 2014). <http://www.oecd.org/pisa/keyfindings/pisa2009keyfindings.htm>
 22. Australian Bureau of Statistics. Deaths, Australia 2012 [Internet]. 2013 (cited 12 Sep 2014). ABS Cat. no. 3302.0. <http://www.abs.gov.au/ausstats/abs@.nsf/mf/3302.0>
 23. Amstadter AB, Myers JM, Kendler KS. Psychiatric resilience: longitudinal twin study. *Br J Psychiatry* 2014;205:275–80.