The Bayley Scales of Infant and Toddler Development- 3rd Edition: A Quantitative Analysis for Application in an Australian Population

INGRID HONAN

Bachelor of Psychology (Honours)

Macquarie University

Faculty of Human Sciences

Department of Psychology

Thesis submitted in fulfilment of the requirements for the combined degrees of Master of Clinical Neuropsychology and Doctor of Philosophy

August 2017

Table of Contents

List of Tables and Figures

Thesis Abstract

The Bayley Scales of Infant and Toddler Development- 3rd Edition (BSID-III) is widely used in Australia. Despite this, there are no Australian normative data and scant quantitative evidence regarding its clinical and predictive utility in an Australian context.

This thesis used prospective, longitudinal data from a large cohort of general population 1- and 3-year-old Australian children and their mothers to investigate the clinical and predictive utility of the BSID-III in Australia. Specifically, this thesis consisted of four empirical studies that aimed to: (1) explore differences between 1-year-old Australian infant performance on the BSID-III and the US normative sample; (2) investigate the utility of the BSID-III to detect differences in infant performance based on indirect perinatal risk factors for neurodevelopmental delay; (3) explore differences in performance on the BSID-III between 3-year-old Australian children and the US normative sample; and, (4) examine the predictive utility of the BSID-III from 1-year to 3-years, as well as by sex.

Information on maternal socio-demographics, birth and infant health was collected via structured maternal interview antenatally, and at 8-weeks, 1-year and 3-years (subsample) postnatally. The BSID-III was administered by trained assessors at 1- and 3-years.

Study 1 found that at 1-year of age Australian children performed significantly higher than the US BSID-III normative sample on the cognitive domain and significantly lower on the gross motor domain. Study 2 showed that the BSID-III was able to detect some statistically significant, but not clinically relevant differences between infants at indirect risk of neurodevelopmental delay based on perinatal factors, and those at low risk/no risk. Study 3 demonstrated that Australian 3-year-old children performed significantly higher on the language (expressive and receptive) and motor domains (fine and gross) of the BSID-III compared to the US normative sample, but significantly lower on the cognitive domain. With regards to Study 4, results suggested that the BSID-III held greatest predictive utility from 1-

to 3-years on the receptive language domain, and that predictive utility of the BSID-III improved once examined by sex.

This thesis contributes to our knowledge of the clinical and predictive utility of the BSID-III for Australian infants and pre-schoolers. Taken together, results suggest that the current clinical practice in Australia of utilising US normative data to interpret infant performance is likely to result in sub-optimal identification of developmental delay. Clinical implications and directions for future research are discussed.

Candidate Declaration

I, Ingrid Honan, Candidate for the combined degrees of Masters in Neuropsychology/Doctor of Philosophy at Macquarie University, Sydney, Australia do hereby certify that:

i) The papers and thesis contained herein comprise my original work towards the degree;

ii) This work has not been submitted to any other university or institution for a higher degree;

iii) This thesis is less than 75,000 words in length, excluding table, figures, reference and appendices; and

iv) Ethics approval for research conducted as part of the current thesis was obtained from:

   a. NSW Health Human Research Ethics Committees (approval numbers: X11-011; HREC/11/RPAH/153)

   b. 5201400106

Signed: _____ Date: 02/08/2017

Ingrid Honan

Co-Author Declaration

We the undersigned acknowledge the following statements:

i) This thesis principally represents the work of Ingrid Honan.

ii) Dr Delyse Hutchinson and A/Prof Melanie Porter provided supervisory support for each of the papers, and for the preparation of the Introduction and General Discussion of this thesis.

iii) The broader Triple B: NHMRC grant was obtained by: Professor Richard P. Mattick; Delyse Hutchinson; Professor Craig A. Olsson; Professor Steve Allsop; Professor Elizabeth J. Elliott; Associate Professor Lucinda Burns; Emeritus Professor Jake Najman and Dr Sue Jacobs.

iv) Listed authors of each study contributed to the publications by critically reviewing written drafts prior to journal submission.

Delyse Hutchinson Signature:

Melanie Porter Signature:

Richard P. Mattick Signature:

Craig A. Olsson Signature:

Steve Allsop Signature:

Elizabeth J. Elliott Signature:

Lucinda Burns Signature:

Jake Najman Signature:

Sue Jacobs Signature:

Contribution to Data Collection for Thesis

The conceptualisation of the thesis topic stemmed from my previous role as a Research Officer on the Triple B Pregnancy Cohort Study (Bumps, Babies and Beyond). Triple B is a longitudinal pregnancy cohort study examining factors that influence child development antenatally to 1-year of age. I was responsible for establishing ethics and recruitment procedures at one of the major hospital sites, interviewing parents, and training a team of staff on the administration of the Bayley Scales of Infant and Toddler Development-3rd Edition. Through this work, my interest in examining the psychometric properties of the BSID-III in Australia developed; this was a unique research question beyond the aims of the broader study.

While the 1-year data included in this thesis were derived from the Triple B cohort, I significantly contributed to data collection and entry throughout my candidature period. I undertook approximately 1,700 in-kind hours of data collection and data entry time. I also conducted the 3-year sub-sample follow-up, manually cross checked all 1-year Pearson Psychcorp software scored BSID-III assessments for accuracy and conducted formal inter-rater reliability checks. Moreover, I was responsible for merging, cleaning, coding, and writing syntax for all data relevant to this thesis, as well as running analyses with advice sought from statistical advisors.

Acknowledgements

This thesis would not have been possible without the support and encouragement of all those around me; and for that I will be forever grateful.

First and foremost, I would like to thank my supervisors Associate Professor Melanie Porter and Dr Delyse Hutchinson for your continued support, guidance and patience throughout my candidature. You both believed in me even when I doubted myself, and for that I am truly grateful. Mel, your academic and clinical knowledge and expertise is inspirational and has shown me what is possible beyond my candidature. Delyse, you have been there for me both academically and professionally for the past 6 years. Without you, this project would not have been possible. Thank you for all the opportunities you have created for me and your guidance over the years. I would also like to thank Professor Richard Mattick, not only for your support professionally but for your exceptional editing and word-cutting skills.

To Macquarie University, the National Education Trust Fund and the National Drug and Alcohol Research Centre, UNSW, thank you for the financial support that you have provided me which has enabled me to complete this thesis.

I wish to thank all my colleagues on the Triple B project. This thesis would not have been possible if it weren't for the tireless effort of the team: driving around Sydney; rushing to assessments; and, missing toilet breaks. To Hannah and Chiara, you both went above and beyond, celebrating my every small success and putting up with my w(h)ining well beyond when you both completed your higher degrees. Thank you for truly understanding what I was going through.

To my current colleagues at Cerebral Palsy Alliance, thank you for your continued understanding and support, particularly through this last difficult stretch.

I would also like to thank all my friends, especially the waterpolo group, the ‘Italians’ and my school friends. Thank you for your patience and understanding throughout my candidature. Your ability to take my mind off things and lift my spirits has been a welcome reprieve over the past 4-years.

To my family, Andrew, Jenny, Lisa, Anna and Tai, thank you for your endless love and support. Thank you for the “checking-up” phone calls making sure I was on track, your refusal to accept my “fine” responses to your thousands of questions, for reading drafts, for the home-cooked meals and for at least pretending to understand to what I have been talking about for the past 4-years. Your encouragement and belief in me has been unwavering not only throughout my studies, but all aspects of my life, and for that, words cannot express my gratitude.

Finally, to my loving husband Samuele Avallone, from the bottom of my heart I thank you. For not only putting up with my procrastination, stress, sleepless nights, and mood swings, but for marrying me despite it. You are the reason I kept my sanity throughout both the Masters and the PhD. Thank you for understanding and supporting me, and for taking care of all aspects of our lives when I put everything on hold until “after my PhD”; it’s finally finished!

1. General Introduction.

Standardised developmental assessment tools serve two important functions: (1) to assist clinicians to accurately detect developmental delays; and, (2) to inform the provision of appropriate and timely interventions. The Bayley Scales of Infant and Toddler Development-3rd Edition (BSID-III) (Bayley, 2006a) is one of the most widely reported standardised developmental assessment tools in research and clinical practice both in Australia and internationally (Anderson, De Luca, Hutchinson, Roberts & Doyle, 2010; Luttikhuizen dos Santos, de Kieviet, Konigs, van Elburg & Oosterlaan, 2013). Despite the long-standing use of the Bayley Scales in Australia, no Australian normative data are available. Australian clinicians rely solely on normative data from the United States (US) to interpret Australian infants' performance on the BSID-III.

Research from the broader fields of psychometric assessment (Wechsler, 2008; Wechsler, 2014), paediatric assessment (Walker, Badawi, Halliday & Laing, 2010; Chinta, Walker, Halliday, Loughran-Fowlds & Badawi, 2014) and international research on the BSID (Steenis, Verhoeven, Hessen & Van Baar, 2015; Krogh, Væver, Harder & Køppe, 2012), suggests that utilisation of US normative data for non-US populations is sub-optimal, and may result in inaccurate interpretation of results. For example, research from the Netherlands has demonstrated differences in interpretation of child performance on the BSID-III when using country specific normative data compared to the US normative data (Steenis et al., 2015; Westera, Houtzager, Overdiek & Van Wassenaer, 2008). Furthermore, Australian normative data have been developed for many gold standard assessment tools, such as the Wechsler Scales of Intelligence, to ensure normative data are representative of the Australian population (Weschler, 2008; Weschler, 2014). However, scant research to date has examined the psychometric properties of the BSID-III in an Australian context, and no Australian normative data are available. This has important implications for clinicians that use the

BSID-III in Australia. Moreover, available research in this field has highlighted the need for a greater understanding of the application of the BSID-III in an Australian population (Walker et al., 2010; Anderson et al., 2010).

The BSID-III is one of the most widely used assessment tools in Australia yet research examining its clinical and predictive utility for use in an Australian population is limited. Thus, the broad aim of this thesis was to conduct a detailed quantitative examination of the BSID-III in an Australian cohort. More specifically, the four aims of the current thesis were: (1) to explore differences between 1-year-old Australian infant performance on the BSID-III and the US normative sample; (2) to investigate the utility of the BSID-III to detect differences in infant performance based on indirect perinatal risk factors for neurodevelopmental delay; (3) to explore differences in performance on the BSID-III between 3-year-old Australian children and the US normative sample; and, (4) to examine the predictive utility of the BSID-III from 1-year to 3-years, stratified by sex.

In this general introduction, literature relevant to the aims of this thesis, gaps in the literature, and the importance of addressing these gaps, will be discussed.

1.1 Prevalence and Assessment of Delay in Early Childhood in Australia

1.1.1 Prevalence and Burden of Developmental Delays in Australia

Developmental delay in infancy, toddlerhood and the preschool years indicates those children most at risk of ongoing disability (Illingworth, 2013; Aylward 2002; Walker, Holland, Halliday & Badawi, 2012). In Australia, 3.4 percent of children aged 0- to 4-years have a developmental delay or disability (sensory, physical, mental), of which mental and behavioural conditions are the most common, making up 40 percent (The Australian Bureau of Statistics (ABS), 2012). Furthermore, intellectual disability alone in Australia is estimated to cost approximately 14,720 billion dollars annually, with the majority of the expense

absorbed by families (Doran et al., 2012). Targeted early intervention has demonstrated ameliorative potential for many types of developmental delay (Ramey & Ramey, 2004; Bratton, Ray, Rhine & Jones, 2005; Peters-Scheffer, Didden, Korzilius & Sturmey, 2011; Morgan, Novak & Badawi, 2013), reducing the individual, familial, societal and economic burden associated with ongoing disability. However, the first step necessary for the provision of appropriate and timely early intervention is the accurate detection of developmental difficulties.

1.1.2 Detection of Developmental Delays and Disability in Australia

While some developmental conditions such as Trisomy 21 and Fragile X Syndrome can be detected through genetic testing, many other common conditions rely on clinical judgment to detect delay. Furthermore, the extent of delay varies within conditions; as such, developmental assessments are required to determine level of delay. As it is not feasible to monitor the developmental milestone attainment of all children, those children identified as most at risk of neurodevelopmental delay (NDD) are referred to *developmental follow-up clinics* (Walker et al., 2012). NDD refers to substantial reduction in functioning resulting from specific or non- specific damage to the developing brain or central nervous system. Examples of these developmental follow-up clinics in Australia include: *Grace Centre for Newborn Intensive Care at The Children's Hospital at Westmead; Royal Hospital for Women- Follow Up Clinic; Monash Newborn; ACT Health- NICU Growth and Development Clinic; Government of Western Australia- King Edward Memorial Hospital Neonatal Follow-Up Program; and the Outpatient Clinics Tasmanian Health Service- Neonatal Follow Up Clinic*. At these clinics development is closely monitored across the first years of life, with the aim of accurately detecting early developmental difficulties, and informing early intervention.

1.1.3 Key Periods for Developmental Assessment in Australia

Two of the most common developmental follow-up ages in Australia where the BSID-III is administered are 1- and 3-years of age. Clinics across Australia, including the *Grace Development Clinic at The Children's Hospital at Westmead; ACT Health- NICU Growth and Development Clinic; Government of Western Australia- King Edward Memorial Hospital Neonatal Follow-Up Program; and, the Outpatient Clinics Tasmanian Health Service- Neonatal Follow up Clinic* administer the BSID-III at 1- and/or 3-years of age for children referred to their services. Therefore, understanding the clinical and predictive utility of the BSID-III at these ages is critical.

One year of age is a significant developmental stage, where infants begin to walk and talk (Bayley, 2006a). Although infant development is rapid and variable at this age and therefore some children may be expected to 'catch-up' or outgrow delays through the natural maturation process (Anderson & Burnett, 2017), other delays can have a significant impact on a child's ability to learn and engage with the environment, resulting in increased impairment (Guralnick, 2011). As such, detection of delays at 1-year of age allows for early detection of delay and the provision of early intervention with the aim of maximising individual potential.

Three years of age is a period during which behaviour becomes more purposeful and individual aptitude stabilises. Developmental delays at this age are considered indicators of future delay, thus it is a common diagnostic age for many developmental conditions such as Autism Spectrum Disorder (Chakrabarti & Fombonne, 2001; Charman et al., 2005; Kleinman et al., 2008; Lloyd, MacDonald & Lord, 2013); Global Developmental Delay (Riou, Ghosh, Francoeur & Shevell, 2009; Shevell, Majnemer, Platt, Webster & Birnbaum, 2005) and Communication and Language disorders (Silva, Williams & McGee, 1987). Moreover, the rapid brain development documented to occur during infancy and early childhood increases

the risk of neurodevelopmental delay across these periods (Knickmeyer et al., 2008). Thus, accurate early detection of delays at 3-years of age is essential to inform diagnosis, early interventions and for appropriate access to funding and services.

Taken together, given the significance of clinical assessment for delay at 1- and 3-years of age, particularly in Australian health services, this thesis focused specifically on these two age groups.

1.2 Structure of the BSID-III

1.2.1 The Domains of the BSID-III

The BSID-III is a gold standard development assessment tool, and is widely used in Australia and internationally. It was developed to identify delay and inform planning for early intervention with children aged 0-to 42-months (Bayley, 2006b). It is comprised of five objectively assessed domains of function, as well as two parent self-report questionnaires measuring social-emotional functioning and adaptive behaviour (Bayley, 2006b). Table 1 describes the five objectively assessed domains of the BSID-III.

Table 1. Structure and content of objectively assessed BSID-III domains

Domain	No. of items	Types of abilities assessed
Cognition	91	Memory, concept formation, problem solving skills and learning abilities.
Receptive Language	39	Ability to orientate to sounds, understand spoken language and respond to verbal requests.
Expressive Language	48	Ability to make sounds, label items, use tense, adjectives and pronouns appropriately and communicate meaning.
Fine Motor	66	Manual speed and dexterity, and ability to integrate visual-perceptual information and plan motor responses.
Gross Motor	72	Core strength, extremity muscle control, balance and co-ordination.

1.2.2 BSID-III Scoring

Start points are determined based on the child's age and established basal level of competency, with credit automatically awarded for previous items. Items are pass/fail and awarded one point each. Once a ceiling is achieved, item scores are totalled per domain and then compared to age matched normative data in order to obtain scaled scores for each domain (Bayley, 2006a). Composite scores can also be obtained for cognition, language and motor, by combining performance on receptive and expressive language domains, and performance on the fine and gross motor domains (Bayley, 2006a). However, it is important to note that combining scores across related but different domains to obtain composite scores, can result in loss of sensitivity of results for children with specific developmental delays (Lezak, 2004). Thus, for the purpose of this thesis, scaled scores were utilised to report performance on all five objectively administered domains.

1.3 Development of the BSID-III

First developed in 1969 (Bayley, 1969), the BSID was founded on developmental theory and comprised of items adapted from a number of available assessment tools including the California First-Year Mental Scale (Bayley, 1933), the California Preschool Mental Scale (Jaffa, 1934), and the California Infant Scale of Motor Development (Bayley, 1936). Despite some criticism regarding test-retest reliability (Nellis & Gridley, 1994), the BSID established itself as the "best measure of infant development available" (Sattler, 1988, p. 916). On revision in 1993, the BSID-II retained the same domain structure and the majority of test items as its predecessor (Nellis & Gridley, 1994). Comprised of a Mental Development Index (MDI) and Psychomotor Development Index (PDI), the revision resulted in an expansion of the original age range to include infants aged 1- to 42-months. A number of new test items were also incorporated to assess additional aspects of development and ability not previously included in the BSID (Bayley, 1993). Furthermore, the former Infant Behaviour Record was replaced by the Behaviour Rating Scale, with the aim of improving reliability of scoring and interpretation.

In 2006, the BSID-II was again revised, resulting in the currently available BSID-III (Bayley, 2006a). The BSID-III underwent extensive changes when compared to its predecessor with the primary goals of updating items, stimuli and normative data, simplifying administration, incorporating new research findings on the major constructs and improving the psychometric properties and clinical utility of the tool (Albers & Grieve, 2007; Bayley, 2006b). This resulted in substantial changes, with approximately 50 percent of items being changed, deleted or added. The MDI and PDI domains were replaced with five objectively assessed domains of function: cognition; receptive language; expressive language; fine motor and gross motor (Albers & Grieve, 2007; Bayley, 2006b). Furthermore, the Behaviour Rating Scale was replaced with a Social-Emotional Scale and Adaptive Behaviour Scale, based on

the Greenspan Social Emotional Growth Chart (Greenspan 2004) and the Adaptive Behaviour Assessment System- 2nd Edition (ABAS-II) (Albers & Grieve, 2007; Bayley, 2006b; Harrison & Oakland, 2003).

1.4 BSID-III Normative Data

The BSID-III was normed on 1,700 US children. The normative group was divided into 17 age bands with 100 children (50 males and 50 females) in each age band. The mean scaled score of the US normative sample on each of the five objectively rated domains was 10, with a standard deviation of three, and a scaled score range of one to 19 (Bayley, 2006b). The normative sample was reportedly representative of the US population with regards to ethnicity and parental education. Attempts were also made to ensure geographical representativeness of included infants by dividing the US into four regions, and sampling from each (Bayley, 2006b). Inclusion criteria for the general normative sample were: gestational age between 36- to 42-weeks; typical development, as defined by no significant medical complications at birth; no history of medical complications, and not diagnosed nor receiving treatment for mental, physical or behavioural difficulties (Bayley, 2006b). However, group studies of children with specific developmental conditions or at risk for delay (e.g., cerebral palsy, trisomy 21, autism spectrum disorder, asphyxiation at birth and prenatal alcohol exposure) were conducted and children with these conditions were reintroduced into the normative sample, making up approximately 10 percent of the overall normative sample (Bayley, 2006b). Although this was conducted in an effort to avoid truncation of normative data, 10 percent re-inclusion appears higher than would be expected in the general population. Furthermore, selective sampling resulted in limited age ranges of children with conditions/risk factors. Therefore, re-inclusion of children in the US normative data does not appear consistent across age bands, and it is unclear how many children with

specific developmental conditions or risk factors were reintroduced into the US normative age bands relevant to this thesis.

As outlined, parent-rated domains were derived from existing questions. Methods used to derive normative data for the social-emotional and adaptive behaviour domains also differed from the objectively rated domains, resulting in differing age bands between objectively rated domains and parent-rated domains. For this reason, this thesis did not examine the parent-rated domains. The cohort follow-up ages were selected in line with the US normative data objectively rated domains and therefore cannot be appropriately matched to the parent-rated domains.

1.5 BSID-III Psychometric Properties

According to the BSID-III technical manual, the *reliability* of the BSID-III, that is the consistency, accuracy and stability of scores obtained, was reportedly high (Bayley, 2006)b. Internal consistency, obtained using the split-half method with the Spearman-Brown correction and Fisher's z transformations, ranged from 0.86 to 0.91 across domains (Bayley, 2006b). Standard error of measurement statistics used to estimate the amount of error in an individual's obtained score compared to their true score, ranged from 0.93 to 1.17 scaled score points (Bayley, 2006b). Test re-test reliability was also adequate, with average stability coefficients across 197 children ranging from 0.72 to 0.81 (Bayley, 2006b), with differences between time periods partially attributable to practice effects resulting from the short test re-test period. It should be noted that reliability statistics were generally weaker in younger age bands (2- to 4-months of age; Bayley, 2006b); these statistics are not reported here given the cohort of this thesis was 1- and 3-years of age.

The *validity* of a scale refers to the extent to which an assessment tool measures what it is designed to measure. The BSID-III is purported to have good construct validity (Albers

& Grieve, 2007; Bayley, 2006b). Theoretical construct validity was demonstrated through the inclusion of constructs theorised to be involved in development, such as play (Albers & Grieve, 2007; Bruner,1972; Piaget, 1952; Vygotsky, 1978) and with the addition of the construct of information processing (Colombo & Frick, 1999) within the cognitive domain. Furthermore, consultation with a wide variety of experts, piloting and re-piloting and standardisation informed the construct validity of the BSID-III (Albers & Grieve, 2007; Bayley, 2006b). Empirical evidence for construct validity is also provided in the BSID-III technical manual. Moderate to high correlations were found both between domains of the BSID-III, and with tests designed to examine similar constructs to the BSID-III (concurrent validity), such as: the Wechsler Preschool and Primary Scale of Intelligence- 3rd Edition (Wechsler, 2002); The Preschool Language Scale Fourth Edition (Zimmerman, Steiner & Pond, 2002); the Peabody Developmental Motor Skills Second Edition (Folio & Fewell, 2000); the ABAS-II and the Vineland Adaptive Behaviour Scale Interview Edition (Albers & Grieve, 2007; Bayley, 2006b; Sparrow, Balla & Cicchetti, 1984).

Whilst the BSID-III technical manual reports good psychometric properties, a number of studies have raised questions regarding the reliability of the BSID-III in Australia (Anderson & Burnett, 2017). Research suggests that the BSID-III, when compared to its predecessors, over-estimates ability and therefore under detects delay (Anderson & Burnett, 2017). Moreover, results regarding its predictive utility have been inconsistent.

In order to determine the *clinical utility* of the BSID-III in Australia, the appropriateness of the application of US normative data in Australian cohorts, and the ability of the BSID-III to differentiate between those at-risk of delay, and those at low-risk, should be determined. Furthermore, the *predictive utility*; the degree to which performance on a measure predicts future ability on tasks assessing the same or similar abilities, needs to be established. Developmental assessment tools are grounded on a premise that they objectively

assess skills that *are* reasonably predictive of future ability. Therefore, examining the predictive utility of the BSID-III is essential to understanding its clinical utility for tracking and interpreting child progress across time, and before and after interventions. As such, these areas were the focus of this thesis and are discussed in the ensuing sections.

1.6 The Application and Utility of US BSID-III Normative Data Internationally

1.6.1 Use of US BSID-III Normative Data Internationally

The importance of country specific normative data has been well established within the psychometric assessment field (Ardila, 2005; Geisinger, 1994). Normative data provide a reference group to which individual performance can be compared, in order to determine functioning relative to same age peers. Country specific differences such as primary language, ethnicity, cultural beliefs, socio-economic factors, and access to education and services may impact performance on psychometric tests. As such, accurate interpretation of individual functioning is dependent on the representativeness of the normative reference group to the population from which an individual is drawn. In recognition of the importance of representative normative data, and the acknowledged cultural differences between countries, country specific normative data have been developed for many gold standard assessment tools (e.g., Wechsler, 2008; Wechsler, 2014). However, despite the BSID-III being a gold standard developmental assessment tool, no Australian normative data are available.

The appropriateness of using US BSID normative data has been examined across BSID versions II and III and in a number of international populations, with results suggesting that utilisation of US normative data results in sub-optimal interpretation of individual ability. For example, Westera et al. (2008) studied the performance of 376 premature Dutch children on the BSID-II and demonstrated statistically and clinically significant differences in the

interpretation of ability when using Dutch normative data versus US normative data. Similarly, a population study comparing 1,912 Dutch children (0- to 42-months) to the US BSID-III normative sample, reported significant group differences in child performance on all domains of the BSID-III across a range of ages. Results indicated that Dutch children generally scored lower than US children on the gross motor domain, and scored higher on cognitive, fine motor and receptive and expressive language domains (Steenis et al., 2015). Moreover, a study comparing the performance of 45 Danish children at 4-, 7-, 10- and 13-months of age, to the US BSID-III normative data, revealed significant differences in performance (Krogh et al., 2012). Danish children performed consistently lower than US children on the receptive language domain and lower than US children on the motor domain at 10- and 13-months (Krogh et al., 2012). Conversely, higher scores were obtained by Danish children on the cognitive domain at 4-, 7-, and 13-months, when compared to the US normative sample (Krogh et al., 2012), highlighting the importance of developing country specific norms.

1.6.2 Use of US BSID-III Normative Data in Australia

It could be argued that US BSID-III normative data may be more appropriate to use in countries where English is the primary language. Yet the limited available research in English speaking countries, such as Australia, suggests otherwise. A study comparing scores obtained by Australian children on the original BSID to those obtained on BSID-II, reported a greater mean difference obtained by Australian children than US children, suggesting Australian children may be developing at a different rate to US children (Tasbihsazan, Nettelbeck & Kirby, 1997). Following the revision and development of the BSID-III, three studies examined the utility of the US BSID-III normative data in an Australian population, with inconsistent findings. A study of 211 premature and/or low birth weight 2-year-old Australian

children, and 202 full-term, normal birth, controls, found that Australian full-term children performed higher than the US normative sample on all objectively assessed domains of the BSID-III (Anderson et al., 2010). As such, use of US BSID-III normative data resulted in an underestimation of delay in extremely preterm/low birth weight children (Anderson et al., 2010). Similarly, a study by Chinta, Walker, Halliday, Loughran-Fowlds and Badawi (2014) compared 156 3-year-old Australian children on the BSID-III to the US normative sample. Results indicated that Australian children scored significantly higher than the US BSID-III normative sample on cognitive, receptive language, expressive language and fine motor subtests, but no differences were detected on the gross motor domain (Chinta et al., 2014).

Conversely, a study of 211 healthy, full-term, singleton 1-year-old children reported that Australian children obtained lower scores on the gross motor and expressive language domains, and higher scores on the cognitive and receptive language domains when compared to the US BSID-III normative sample (Walker et al., 2010). In interpreting results, Walker et al. (2010) inferred that cultural factors such as ethnicity and child rearing practices impact development and therefore normative data. However, ethnicity proportions were reportedly representative of the local recruitment area, rather than the Australian population, and no maternal demographic data were available. Given the known associations between maternal factors such as socio-economic status, level of parental education and employment status, and offspring ability (Guralnick, 2011), cohort information pertaining to these factors is required in order to determine the generalisability of results to an Australian population.

Taken together, while results support the need for Australian specific normative data, the directionality of results is inconsistent. Based on available evidence, it is plausible that Australian developmental trajectories differ from the US normative sample at different ages. For example, at 1-year of age Australian expressive language may be lower than the US, and at 3-years of age higher than the US. This may be due to any number of differences between

Australian and US populations, such as bilingualism, for example. However, with a lack of reported demographic data, it is also possible that inconsistent results across studies reflect sampling biases rather than population differences. Therefore, the inconclusive findings from the three available studies warrant the need for further research in this area. This was the focus of Studies 1 and 3 of this thesis.

1.7 At Risk and Low-Risk Populations

In order for the BSID-III to have clinical utility in an Australian population, not only is it important that the US normative data are appropriate to use in the Australian context, but the BSID-III should accurately distinguish between children with known risk factors for NDD and those at low/no risk.

Whilst some perinatal risk factors for NDD, such as severe hydrocephalus, encephalopathy and intracranial haemorrhage, result in detectable neurological signs, others do not. Factors such as prematurity (Bos & Roze, 2011; Greene, Patra, Nelson, & Silvestri, 2012), low birth weight (de Moura et al., 2010), small head circumference (Peterson, Taylor, Minich, Klein, & Hack, 2006), admission to neonatal intensive care unit (NICU) and/or special care units (SCU) post-birth (Walker et al., 2012), low Apgar scores (Odd, Rasmussen, Gunnell, Lewis, & Whitelaw, 2008), maternal substance use (Bandstra, Morrow, Mansoor, & Accornero, 2010; Huizink, 2014), and multiple birth infants (Wadhawan et al., 2011) are all associated with increased risk of NDD; yet may not independently result in detectable neurological signs. Infants with identifiable perinatal neurological signs are often monitored closely during the perinatal period, with multiple assessment methods used to establish the likelihood, and degree of, NDD. Conversely, infants with indirect perinatal risk factors, not resulting in detectable neurological signs, but associated with increased risk of NDD, are often referred for developmental follow-up clinics in Australia.

Infants characterised solely by indirect perinatal risk factors are less likely to experience profound NDD than infants with detectable neurological signs perinatally. As such, infants with indirect perinatal risk factors are arguably the infants most likely to benefit from standardised psychometric assessment. Here, sensitive standardised psychometric assessment holds the potential to detect delays that may not be readily detectable on observation alone, increasing the likely delivery of appropriate and timely early intervention.

To date, although research suggests that the BSID-III may overestimate ability and therefore underestimate delay when US BSID-III normative data are used to interpret performance, it has been shown to detect group differences in performance between at risk infants and control infants (Anderson et al., 2010). This suggests that with more representative normative data, the BSID-III would be able to detect differences in performance between at risk children and low risk children. However, much of the research examining the ability of the BSID-III to detect NDD in children with indirect risk factors has been conducted in samples with one primary risk factor (e.g. prematurity) (Lobo, Paul, Mackley, Maher & Galloway, 2014). Yet, risk factors co-occur in the general population. While primary risk factor studies provide important profiling information, homogeneity of samples may limit generalisability of results to real world clinical settings.

Moreover, much of the research to date has been conducted with children older than 18-months of age. For example, a study of 185 extremely pre-term children, with a mean corrected age of 33-months, found that using conventional recommended cut-offs the BSID-III under-identified developmental delay (Johnson, Moore & Marlow, 2014). Similarly, studies of 2-year-old children with congenital heart disease (Long, Galea, Eldridge & Harris, 2012) and extremely preterm/low birth weight and full-term infants (Anderson et al., 2010) both showed sub-optimal detection of delay by the BSID-III. To the author's knowledge, no studies evaluating the ability of the BSID-III to detect delays in children less than 18-months

of age and drawn from heterogenous samples characterised by indirect risk factors are available. As previously outlined, 1-year of age is a common developmental follow-up age in Australia. Therefore, understanding the ability of the BSID-III to detect differences in performance based on indirect risk status at this age, is critical. This was the focus of Study 2 of this thesis.

1.8 Predictive Utility

As noted earlier, *predictive utility* refers to the degree to which performance on an assessment tool predicts future ability on tasks assessing the same or similar abilities. Although the BSID-III was designed to assess current delay, rather than predict future delay (Bayley, 2006b), the predictive utility of the BSID-III is arguably just as important. There is little value in identifying that a child has a current delay, if that delay is not predictive of future ability. Moreover, developmental assessment tools are commonly used to monitor progress over time and determine efficacy and responsiveness to early intervention. As such, understanding the predictive utility of developmental assessment tools is fundamental to determining clinical utility.

Earlier versions of the BSID came under criticism for weak predictive utility, with studies reporting low to moderate predictive utility (Aylward, 2002; Crowe, Deitz & Bennett, 1987; Hack et al. 2005; Potharst et al., 2012; Roberts, Anderson, Doyle & Victorian Infant Collaborative Study Group, 2010). A more recent meta-analysis of 16 studies of the BSID, BSID-II and one study of the BSID-III in very preterm/low birth weight children, reported stronger predictive power of the Mental Development Index (MDI), than the Psychomotor Development Index (PDI) (Luttikhuizen dos Santos et al., 2013). Although some studies reported better predictive utility in at risk/delayed populations (Aylward, 2002; Harris, Megens, Backman & Hayes, 2005), results have nonetheless been inconsistent (Niccols &

Latchman, 2002). Furthermore, when examining the predictive utility of previous versions of the BSID by sex, results suggested differing predictive utility based on sex. For example, a longitudinal study of 70 children from 6- to 36-months of age examining the predictive utility of the BSID-II by sex, suggested male performance on the BSID-II to be more stable over time than female performance, a result that was attributed to a large growth in female language ability (Lung, Shu, Chiang, Chen & Lin, 2009).

Despite one of the primary aims of the revised BSID-III being improvement in the psychometric properties and clinical utility of the tool, predictive utility was not assessed by the test developers (Bayley, 2006b). Although stability and predictive validity studies have since aimed to address this gap, the majority of research has been conducted in preterm samples and has reported inconsistent results. For example, examination of the predictive utility of the motor scale of the BSID-III at 2-years of age, on motor ability at 4-years of age, in a sample of 96 preterm children, demonstrated high specificity but low sensitivity to detecting delays (Spittle et al., 2013). Moreover, Spencer-Smith, Spittle, Lee, Doyle and Anderson (2015) investigated the ability of the cognitive and language domains of the BSID-III at 2-years of age to predict outcomes on the Differential Ability Scale-II at 4-years of age in 105 preterm children. Results again suggested poor predictive utility of the BSID-III. Similarly, Lobo et al (2014) examined 24 low risk and 30 preterm infants across seven time-points between 3- and 24-months of age and reported poor stability in BSID-III scores over time.

Conversely, Bode, D'Eugenio, Mettelman and Gross (2014) compared 156 preterm and 155 term childrens' performance on the cognitive and language BSID-III domains at 2-years of age, with outcomes on the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) at 4-years of age, and reported strong predictive utility for preterm children and moderate predictive utility for term children. Moreover, a study of 131 preterm and medically

unstable infants' performance on the BSID-III from 8- to 20-months of age reported fair to moderate correlation coefficients (Greene, Patra, Silvestri, & Nelson, 2013).

To the author's knowledge, no studies are available to date examining the predictive utility of all domains of the BSID-III in a general population cohort of children, and inconsistent results in the extant literature warrant further investigation. Furthermore, no research examining the predictive utility of the BSID-III stratified by sex is available. To address these gaps in the literature Study 4 of this thesis examined the predictive utility of the BSID-III from 1- to 3-years of age, as well as predictive utility stratified by offspring sex.

1.9 Aims of the Current Thesis

The overarching aim of the current thesis was to provide a quantitative examination of the BSID-III in an Australian cohort of 1-year and 3-year old children, to inform the clinical and predictive utility of the tool in Australia. More specifically, the thesis consisted of four broad aims:

(1) To explore differences between 1-year-old Australian infant performance on the BSID-III and the US normative sample;

(2) To investigate the utility of the BSID-III to detect differences in infant performance based on indirect perinatal risk factors for neurodevelopmental delay;

(3) To explore differences in performance on the BSID-III between 3-year-old Australian children and the US normative sample; and,

(4) To examine the predictive utility of the BSID-III from 1-year to 3-years of age in a general population cohort, stratified by sex.

Based on the aforementioned research, it was hypothesised that the current cohort of Australian 1-year-old and 3-year-old children would perform significantly differently on the BSID-III, when compared to the US normative sample. However, given the inconsistencies in the directionality of previous results, no predictions were made regarding the direction of expected differences. Furthermore, given the mixed evidence regarding the utility of the BSID-III to detect neurodevelopmental delay, and the inconsistent results from research examining the predictive utility of the BSID-III, no predictions were made with regard to aims 2 and 4.

The above aims were addressed across four separate studies described in the next section.

1.10 Thesis structure

Study 1: A comparison between Australian infant performance and United States (US) normative data at 1-year on the Bayley Scales of Infant and Toddler Development-3rd Edition.

This study compared the characteristics of 998 Australian 1-year-old infants to the US normative sample, applying raked weights to investigate whether infant performance on the BSID-III differed between the two samples. Sub-group analyses of 12- and 13-month age groups was conducted to examine consistency of detected differences across these age groups.

Study 2: Utility of the Bayley Scales of Infant and Toddler Development-3rd Edition BSID-III to distinguish 1-year old infants at perinatal risk of neurodevelopmental delay.

This study prospectively examined 935 1-year-old infants to investigate the utility of the BSID-III in detecting differences in infant performance based on indirect perinatal risk factors associated with NDD. A factor analysis and examination of individual risk factors with BSID-III domains was also conducted to examine the unique contribution of individual risk factors on overall "indirect risk status", and the appropriateness of a single "indirect risk status" factor.

Study 3: A comparison between Australian and United States normative data at 3-years of age on the Bayley Scales of Infant and Toddler Development- 3rd Edition (BSID-III).

This study used a longitudinal, prospective cohort of 119 children aged 3-years, with raked weights, to investigate whether 3-year-old Australian child performance on the BSID-III differed from the US normative sample. Population differences between Australia and the US were also examined.

Study 4: Predictive utility of Bayley Scales of Infant and Toddler Development (BSID-III) from 1-year to 3-years of age.

This study used a longitudinal, prospective cohort of 122 children to examine the predictive utility of the BSID-III from ages 1-year to 3-years, and to investigate whether predictive utility differed by sex.

1.11 Methodology

The thesis used a prospective, longitudinal design, with data from four times points (antenatally and postnatally at 8-weeks, 1-year and 3-years of age) to capture important family and developmental data. Participants included in this thesis completed all objectively assessed domains of the BSID-III at 1- and/or 3-years of age. In an attempt to elicit optimal performance reflective of true ability, assessments were scheduled at the child's most alert time of day and rescheduled if the child was unwell. Over 95 percent of BSID-III assessments were undertaken in one sitting. In rare cases where it became evident early in the assessment that a child was unwell or "not themselves", the assessment was terminated and rescheduled, recommencing administration from the point of termination.

In line with standardised administration procedures, items were scored through observation where permitted. At 1-year of age, following establishment of rapport, the gross motor domain was administered first, followed by the cognitive domain, receptive language, expressive language and fine motor domain. At 3-years of age domains were administered in the following order: Cognition, receptive language, expressive language, fine motor, and gross motor. Selection criteria and methodology relevant to the studies included in this thesis are provided in the corresponding studies, with further details in Appendices A to E. Statistical consideration of potential confounders is a unique strength of the thesis compared to other available research. Statistical raking was used in Studies 1 and 3 to improve cohort representativeness. Here the purpose of statistical adjustment was not to hold covariates constant, but to improve cohort representativeness relative to population data. In Studies 2 and 4, multivariate regression models were used, to ensure potential confounders were held constant.

1.12 References

ACT Health- NICU Growth and Development Clinic [online website]. Retrieved on March 15, 2017, from http://www.health.act.gov.au/our-services/women-youth-and-children/neonatology-department/nicu-growth-and-development-clinic.

Albers, C. A. & Grieve, A. J. (2007). Test Review: Bayley, N. (2006). Bayley Scales of Infant and Toddler Development- Third Edition. San Antonio, TX: Harcourt Assessment. *Journal of Psychoeducational Assessment, 25*, 180-190. doi: 10.1177/0734282906297199.

Anderson, P. J., & Burnett, A. (2017). Assessing developmental delay in early childhood—concerns with the Bayley-III scales. *The Clinical Neuropsychologist, 31*(2), 371-381. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1080/13854046.2016.1216518

Anderson, P. J., De Luca, C. R., Hutchinson, E., Roberts, G., & Doyle, L. W. (2010). Underestimation of developmental delay by the new Bayley-III Scale. *Archives of Pediatric and Adolescent Medicine*, *164(4)*: 352-356. doi:10.1001/archpediatrics.2010.20.

Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review, 15*(4), 185-195. doi: 10.1007/s11065-005-9180-y

Australian Bureau of Statistics (ABS). *Australian Social Trends June 2012: Children with a Disability*. Catalogue no: 4102.0. Canberra: ABS.

Aylward, G. P. (2002). Cognitive and neuropsychological outcomes: more than IQ scores. *Mental Retardation and Developmental Disabilities Research Reviews, 8*(4), 234-240. doi: 10.1002/mrdd.10043.

Bandstra, E. S., Morrow, C. E., Mansoor, E., & Accornero, V. H. (2010). Prenatal drug exposure: infant and toddler outcomes. *Journal of Addictive Diseases, 29*(2), 245-258. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1080/10550881003684871.

Bayley, N. (1933). The California First-Year Mental Scale. Berkeley, CA: University of California Press.

Bayley, N. (1936). The California Infant Scale of Motor Development. Berkeley, CA: University of California Press.

Bayley, N. (1969). *Scales of Infant Development*. New York: Psychological Corp.

Bayley, N. (1993). *Bayley Scales of Infant Development–Second Edition*. San Antonio, TX: The Psychological Corporation.

Bayley, N. (2006a). *Bayley Scales of Infant and Toddler Development- Third Edition*. San Antonio, Texas: Pearson Psychcorp.

Bayley, N. (2006b). *Bayley Scales of Infant and Toddler Development, Third Edition Technical Manual.* San Antonio, TX: Pearson PsychCorp.

Bode, M. M., D'Eugenio, D. B., Mettelman, B. B., & Gross, S. J. (2014). Predictive validity of the Bayley, at 2 years for intelligence quotient at 4 years in preterm infants. *Journal of Developmental & Behavioral Pediatrics, 35*(9), 570-575. doi: 10.1097/DBP.0000000000000110.

Bos, A. F. & Roze, E. (2011). Neurodevelopmental outcome in preterm infants. *Developmental Medicine and Child Neurology; 53*(s4), 35-39. doi: 10.1111/j.1469-8749.2011.04062.x.

Bratton, S. C., Ray, D., Rhine, T., & Jones, L. (2005). The efficacy of play therapy with children: A meta-analytic review of treatment outcomes. *Professional Psychology: Research and Practice, 36(4),* 376-390. doi:10.1037/0735- 7028.36.4.376.

Bruner, J. S. (1972). Nature and uses of immaturity. American Psychologist, 27(8), 687-708.

Chakrabarti, S., & Fombonne, E. (2001). Pervasive developmental disorders in preschool children. *JAMA, 285*(24), 3093-3099. doi:10.1001/jama.285.24.3093.

Charman, T., Taylor, E., Drew, A., Cockerill, H., Brown, J. A., & Baird, G. (2005). Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry, 46*(5), 500-513. doi: 10.1111/j.1469-7610.2004.00377.x.

Chinta, S., Walker, K., Halliday, R., Loughran-Fowlds, A., & Badawi, N. (2014). A comparison of the performance of healthy Australian 3-year-olds with the standardised norms of the Bayley Scales of Infant and Toddler Development (version-III). *Archives of disease in childhood, 99*(7), 621-624. doi: 10.1136/archdischild-2013-304834

Colombo, J., & Frick, J. E. (1999). Recent advances and issues in the study of preverbal intelligence. *The Development of Intelligence,* 43-71.

Crowe, T. K., Deitz, J. C., & Bennett, F. C. (1987). The relationship between the Bayley Scales of Infant Development and preschool gross motor and cognitive performance. *American Journal of Occupational Therapy, 41*(6), 374-378. doi: doi:10.5014/ajot.41.6.374

de Moura, D. R., Costa, J. C., Santos, I. S., Barros, A. J., Matijasevich, A., Halpern, R., ... & Barros, F. C. (2010). Risk factors for suspected developmental delay at age 2 years in a

Brazilian birth cohort. *Paediatric and Perinatal Epidemiology*, *24*(3), 211-221. doi: 10.1111/j.1365-3016.2010.01115.x.

Doran, C. M., Einfeld, S. L., Madden, R. H., Otim, M., Horstead, S. K., Ellis, L. A., & Emerson, E. (2012). How much does intellectual disability really cost? First estimates for Australia. *Journal of Intellectual and Developmental Disability*, *37*(1), 42-49.

Folio, M. R &Fewell, R. R. (2000), Peabody Developmental Motor Skills, Second Edition. San Antonio, TX, The Psychological Corporation.

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, *6*(4), 304. doi: http://dx.doi.org/10.1037/1040-3590.6.4.304.

Government of Western Australia- King Edward Memorial Hospital Neonatal Follow-Up Program [online website]. Retrieved on March 15, 2017, from http://kemh.health.wa.gov.au/brochures/consumers/wnhs0256.pdf.

Grace Centre for Newborn Intensive Care at the Children's Hospital at Westmead [online website]. Retrieved March 15, 2017, from https://www.schn.health.nsw.gov.au/parents-and-carers/our-services/neonatal-intensive-care/chw.

Greene, M. M., Patra, K., Nelson, M. N., & Silvestri, J. M. (2012). Evaluating preterm infants with the Bayley-III: patterns and correlates of development. *Research in Developmental Disabilities*, *33*(6), 1948-1956. doi: http://dx.doi.org/10.1016/j.ridd.2012.05.024.

Greene, M. M., Patra, K., Silvestri, J. M., & Nelson, M. N. (2013). Re-evaluating preterm infants with the Bayley-III: Patterns and predictors of change. *Research in Developmental Disabilities*, *34*(7), 2107-2117. doi: https://doi.org/10.1016/j.ridd.2013.04.001

Greenspan, S. I. (2004). *Greenspan social-emotional growth chart: A screening questionnaire for infants and young children*. PsychCorp.

Guralnick, M. J. (2011). Why early intervention works: A systems perspective. *Infants and young children*, *24*(1), 6. doi: 10.1097/IYC.0b013e3182002cfe.

Hack, M., Taylor, H. G., Drotar, D., Schluchter, M., Cartar, L., Wilson-Costello, D., et al. (2005). Poor predictive validity of the Bayley Scales of Infant Development for cognitive function of extremely low birth weight children at school age. *Pediatrics, 116*(2), 333-341. doi: 10.1542/peds.2005-0173.

Harris, S. R., Megens, A. M., Backman, C. L., & Hayes, V. E. (2005). Stability of the Bayley II Scales of Infant Development in a sample of low-risk and high-risk infants. *Developmental Medicine & Child Neurology*, *47*(12), 820-823. doi: 10.1111/j.1469-8749.2005.tb01087.x

Harrison, P. L., & Oakland, T. (2003). Adaptive Behavior Assessment System–Second Edition. San Antonio, TX: The Psychological Corporation.

Huizink, A. C. (2014). Prenatal cannabis exposure and infant outcomes: overview of studies. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 52*, 45-52. doi: http://dx.doi.org/10.1016/j.pnpbp.2013.09.014.

Illingworth, R. S. (2013). *The Development of the Infant and the Young Child: Normal and Abnormal* 10th ed. New Delhi, India: Elsevier Health Sciences.

Jaffa, A. S. (1934). The California Preschool Mental Scale. Berkeley, CA: University of California Press.

Johnson, S., Moore, T., & Marlow, N. (2014). Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used?. *Pediatric Research*, *75*(5), 670-674. doi:10.1038/pr.2014.10.

Kleinman, J. M., Ventola, P. E., Pandey, J., Verbalis, A. D., Barton, M., Hodgson, S., ... & Fein, D. (2008). Diagnostic stability in very young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(4), 606-615. doi: 10.1007/s10803-007-0427-8.

Knickmeyer, R. C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J. K., ... & Gilmore, J. H. (2008). A structural MRI study of human brain development from birth to 2 years. *Journal of Neuroscience*, *28*(47), 12176-12182. doi: https://doi.org/10.1523/JNEUROSCI.3479-08.2008.

Krogh, M. T., Væver, M. S., Harder, S., & Køppe, S. (2012). Cultural differences in infant development during the first year: A study of Danish infants assessed by the Bayley-III and compared to the American norms. *European Journal of Developmental Psychology*, *9*(6), 730-736. doi: 10.1080/17405629.2012.688101.

Lezak, M. D. (2004). *Neuropsychological Assessment*. Oxford University Press, USA.

Lloyd, M., MacDonald, M., & Lord, C. (2013). Motor skills of toddlers with autism spectrum disorders. *Autism*, *17*(2), 133-146. doi: https://doi-org.simsrad.net.ocs.mq.edu.au/10.1177/1362361311402230.

Lobo, M. A., Paul, D. A., Mackley, A., Maher, J., & Galloway, J. C. (2014). Instability of delay classification and determination of early intervention eligibility in the first two years of life. *Research in developmental disabilities*, *35*(1), 117-126. doi: https://doi.org/10.1016/j.ridd.2013.10.017

Long, S. H., Galea, M. P., Eldridge, B. J., & Harris, S. R. (2012). Performance of 2-year-old children after early surgery for congenital heart disease on the Bayley Scales of Infant and Toddler Development. *Early human development*, *88*(8), 603-607. doi: https://doi.org/10.1016/j.earlhumdev.2012.01.007

Lung, F. W., Shu, B. C., Chiang, T. L., Chen, P. F., & Lin, L. L. (2009). Predictive validity of Bayley scale in language development of children at 6–36 months. *Pediatrics International, 51*(5), 666-669. doi: 10.1111/j.1442-200X.2009.02844.x.

Luttikhuizen dos Santos, E. S. L., de Kieviet, J. F., Königs, M., van Elburg, R. M., & Oosterlaan, J. (2013). Predictive value of the Bayley scales of infant development on development of very preterm/very low birth weight children: a meta-analysis. *Early Human Development, 89*(7), 487-496. dio:10.1016/j.earlhumdev.2013.03.008.

Monash Newborn (May, 2017) [online website]. Retrieved on June 5, 2017, from http://www.med.monash.edu.au/scs/paediatrics/monash-newborn/.

Morgan, C., Novak, I., & Badawi, N. (2013). Enriched environments and motor outcomes in cerebral palsy: systematic review and meta-analysis. *Pediatrics, 132*(3), e735-e746. doi:10.1542/peds.2012-3985.

Nellis, L., & Gridley, B. E. (1994). Review of the Bayley Scales of Infant Development—second edition. *Journal of School Psychology*, *32*(2), 201-209. doi: https://doi.org/10.1016/0022-4405(94)90011-6.

Niccols, A., & Latchman, A. (2002). Stability of the Bayley mental scale of infant development with high risk infants. *The British Journal of Development Disabilities, 48*(94), 3-13. doi: 10.1179/096979502799104328.

Odd, D. E., Rasmussen, F., Gunnell, D., Lewis, G., & Whitelaw, A. (2008). A cohort study of low Apgar scores and cognitive outcomes. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *93*(2), F115-F120. doi: 10.1136/adc.2007.123745.

Olusanya, B. O. (2011). Perinatal outcomes of multiple births in southwest Nigeria. *Journal of Health, Population and Nutrition*, 29(6), 639-647. PMC3259727.

Outpatient Clinics Tasmanian Health Service- Neonatal Follow up Clinic [online website]. Retrieved on March 15, 2017, from http://outpatients.tas.gov.au/clinics/neonatal_follow_up_clinic.

Peters-Scheffer, N., Didden, R., Korzilius, H., & Sturmey, P. (2011). A meta-analytic study on the effectiveness of comprehensive ABA-based early intervention programs for children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 5*(1), 60-69. doi:10.1016/j.rasd.2010.03.011.

Peterson, J., Taylor, H. G., Minich, N., Klein, N., & Hack, M. (2006). Subnormal head circumference in very low birth weight children: neonatal correlates and school-age consequences. *Early Human Development*, *82*(5), 325-334. doi: https://doi.org/10.1016/j.earlhumdev.2005.09.014

Piaget, J. (1952). The origins of intelligence in children. New York: International Universities Press.

Potharst, E. S., Houtzager, B. A., Van Sonderen, L., Tamminga, P., Kok, J. H., Last, B. F., & Van Wassenaer, A. G. (2012). Prediction of cognitive abilities at the age of 5 years using developmental follow-up assessments at the age of 2 and 3 years in very preterm children. *Developmental Medicine & Child Neurology*, *54*(3), 240-246. doi: 10.1111/j.1469-8749.2011.04181.x.

Ramey, C. T., & Ramey, S. L. (2004). Early learning and school readiness: Can early intervention make a difference? *Merrill-Palmer Quarterly, 50*(4), 471-491. doi:10.1353/mpq.2004.0034.

Riou, E. M., Ghosh, S., Francoeur, E., & Shevell, M. I. (2009). Global developmental delay and its relationship to cognitive skills. *Developmental Medicine & Child Neurology*, *51*(8), 600-606. doi: 10.1111/j.1469-8749.2008.03197.x

Royal Hospital for Women- Follow Up Clinics. NSW Health, South Eastern Sydney Local Health District [online website]. Retrieved March 15, 2017, from http://www.seslhd.health.nsw.gov.au/rhw/Newborn_Care/clinics.asp.

Roberts, G., Anderson, P. J., Doyle, L. W., & Victorian Infant Collaborative Study Group. (2010). The stability of the diagnosis of developmental disability between ages 2 and 8 in a geographic cohort of very preterm children born in 1997. *Archives of Disease in Childhood*, *95*(10), 786-790. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1136/adc.2009.160283

Sattler, J. (1988). *Assessment of children (3rd ed.)*. San Diego, CA: Jerome Sattler.

Shevell, M., Majnemer, A., Platt, R. W., Webster, R., & Birnbaum, R. (2005). Developmental and functional outcomes at school age of preschool children with global developmental delay. *Journal of Child Neurology*, *20*(8), 648-654. doi: https://doi-org.simsrad.net.ocs.mq.edu.au/10.1177/08830738050200080301.

Silva, P. A., Williams, S., & McGee, R. (1987). A longitudinal study of children with developmental language delay at age three: later intelligence, reading and behaviour problems. *Developmental Medicine & Child Neurology*, *29*(5), 630-640. doi: 10.1111/j.1469-8749.1987.tb08505.x.

Sparrow, S. S., Balla, D. A. & Cicchetti, D.V. (1984). The *Vineland Adaptive Behaviour Scales*: *Interview Edition*, Expanded Form. Circle Pines. MN: American Guidance Service.

Spencer-Smith, M. M., Spittle, A. J., Lee, K. J., Doyle, L. W., & Anderson, P. J. (2015). Bayley-III cognitive and language scales in preterm children. *Pediatrics*, *135*(5), e1258-e1265. doi: 10.1542/peds.2014-3039

Spittle, A. J., Spencer-Smith, M. M., Eeles, A. L., Lee, K. J., Lorefice, L. E., Anderson, P. J., et al. (2013). Does the Bayley-III Motor Scale at 2 years predict motor outcome at 4 years in very preterm children? *Developmental Medicine & Child Neurology, 55*(5), 448-452. doi: 10.1111/dmcn.12049.

Steenis, L. J., Verhoeven, M., Hessen, D. J., & Van Baar, A. L. (2015). Performance of Dutch children on the Bayley III: a comparison study of US and Dutch norms. *PloS one*, *10*(8), e0132871. doi:10.1371/journal.pone.0132871.

Tasbihsazan, R., Nettelbeck, T. & Kirby, N. (1997). Increasing mental development index in Australian children: A comparative study of two versions of the bayley mental scale. Australian Psychologist, 32 (2), 120-125.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Wadhawan, R., Oh, W., Vohr, B. R., Wrage, L., Das, A., Bell, E. F., ... & Higgins, R. D. (2011). Neurodevelopmental outcomes of triplets or higher-order extremely low birth weight infants. *Pediatrics*, *127*(3), e654-e660. doi: 10.1542/peds.2010-2646.

Walker, K., Badawi, N., Halliday, R., & Laing, S. (2010). Brief report: performance of Australian children at one year of age on the Bayley scales of infant and toddler development (version III). *The Educational and Developmental Psychologist*, *27*(1), 54-58. doi:10.1375/aedp.27.1.54.

Walker, K., Holland, A. J., Halliday, R., & Badawi, N. (2012). Which high-risk infants should we follow-up and how should we do it? *Journal of Paediatrics and Child Health, 48*(9), 789-793. doi: 10.1111/j.1440-1754.2012.02540.x.

Wechsler, D. (2002). Wechsler Preschool and Primary Scale of Intelligence, Third Edition. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*, *22*, 498.

Wechsler, D. (2014). *WISC-V*. NCS Pearson, Incorporated.

Westera, J. J., Houtzager, B. A., Overdiek, B., & Van Wassenaer, A. G. (2008). Applying Dutch and US versions of the BSID-II in Dutch children born preterm leads to different outcomes. *Developmental Medicine & Child Neurology*, *50*(6), 445-449. doi: 10.1111/j.1469-8749.2008.02067.x.

Zimmerman, I. L., Steiner, V. G. & Pond, R. E. (2002). The Preschool Language Scale, Fourth Edition. San Antonio, TX, The Psychological Corporation.

Study 1

A comparison between Australian infant performance and United States (US) normative data at 1-year on the Bayley Scales of Infant and Toddler Development-3rd Edition.

Ingrid Honan*, BPsych (Hons), Delyse Hutchinson, PhD, Melanie Porter, PhD, Steve Allsop, PhD, Jake Najman, PhD, Elizabeth J. Elliott, MD, Lucinda Burns, PhD, Sue Jacobs, MBBS, Craig A. Olsson, PhD, Richard P. Mattick, PhD.

Paper has been prepared in accordance with authorship guidelines for:

JAMA Pediatrics

Paper has been submitted for publication (July 2017) – See Appendix F

2.1 Abstract

Importance: The Bayley Scales of Infant and Toddler Development III (BSID-III) is widely used internationally. Individual infant performance is compared to United States normative data. Research suggests developmental trajectories differ across countries. As such, utilisation of United States normative data internationally may result in inaccurate interpretation of infant ability.

Objective: To determine whether Australian infant performance on the BSID-III differs from the United States normative sample at 1-year of age.

Design: Data were drawn from the Triple B Pregnancy Cohort Study - a prospective longitudinal birth cohort of women recruited between 2009-13. Structured interviews were completed by mothers during their pregnancy and postnatally at 8-weeks and 1-year. Infants were assessed on the BSID-III at 1-year (12- and 13-months).

Setting: Low risk, term infants were recruited in-utero through NSW public antenatal clinics.

Participants: Of 1,151 infants who completed the BSID-III, 998 infants met inclusion/exclusion criteria and attended all follow-up time points.

Main Outcome(s) and Measure(s): Independently assessed BSID-III scores on domains of cognition, expressive language, receptive language, fine motor and gross motor at 1-year.

Results: Overall, the current cohort obtained higher scores on the cognitive (effect size, 0.54 [M_{diff} 95% CI, 1.31–1.60]); expressive language (effect size, 0.08 [M_{diff} 95% CI, 0.08–0.32]) and fine motor domains (effect size, 0.19 [M_{diff} 95% CI, 0.37–0.68]) when compared to the United States normative sample, and the current cohort obtained significantly lower scores on the gross motor domain (effect size, -0.33 [M_{diff} 95% CI, -1.10 – -0.76]). Differences in cognitive and gross motor scores remained significant in both 12-month (effect size, 0.69 [M_{diff} 95% CI, 1.64 – 1.99]; effect size, -0.24 [M_{diff} 95% CI, -0.89 – -0.47] respectively) and 13-month age groups (effect size, 0.28 [M_{diff} 95% CI, 0.48 – 1.04]; effect

size, -0.54 [M_{diff} 95% CI, -1.80 – -1.19], respectively), and remained unchanged following application of raked weights.

Conclusions and Relevance: Australian infant BSID-III performance consistently differed to United States infant performance at 12- and 13-months in the domains of cognition and gross motor. Population demographic factors such as region of birth and maternal educational attainment may explain these differences. Nation-specific normative data are required to ensure the accurate interpretation of BSID-III test results in Australia.

2.2 Introduction

Developmental difficulties, common in infancy and early childhood, indicate risk for ongoing disability[1]. Some 3.4 percent of Australian children under 4-years of age have such a disability[2]. While some children overcome early difficulties through a maturation process (or 'catching up' their delays), others show ongoing difficulties. Symptom severity and prognosis of children with developmental delay/disability may be attenuated by timely intervention[3]. Accurate diagnosis is the first critical step to providing such intervention. Accurate diagnosis informs early intervention and may alleviate stress experienced by families when a child fails to meet developmental milestones; it also has the long-term potential to reduce the economic burden associated with disability[3].

The Bayley Scales of Infant and Toddler Development III (BSID-III)[4] are a widely used psychometric assessment tool for identifying developmental difficulties in infants aged 0- to 42-months[5,6,7]. In Australia, BSID-III administration is standard practice in many developmental follow-up clinics for infants and children at risk of developmental delay[5]. Despite this, no Australian normative data are currently available for the BSID-III. Normative data provide a comparative reference group of same aged peers through which to interpret individual performance and identify delays. Australian practitioners and researchers currently rely on United States (US) normative data for the BSID-III. As socio-demographic factors are known to influence developmental trajectories[8,9] the US BSID-III normative sample is representative of socio-demographic characteristics of the US population, including parent education level, geographic region and ethnicity (predominantly Caucasian, followed by Hispanic and African American)[4]. This differs from the Australian population's socio-demographic profile (predominantly Caucasian ethnicity followed by Asian)[10]. Furthermore, government policy and access to services such as health care and education differ by country

and have the potential to impact expected developmental trajectories. As such, utilisation of US normative data to interpret Australian infant performance is not ideal.

The importance of using country-specific norms to detect developmental difficulties in infants using the BSID-III has been highlighted recently. A Danish study (n=45) reported significant differences between BSID-III scores obtained and US normative data on a number of age ranges and developmental domains; notably Danish children had poorer receptive language scores[11]. Likewise, in a population study comparing 1,912 Dutch children to the US normative sample, differences were found on all domains of the BSID-III across a range of ages, with poorer performance on the gross motor domain and stronger performance on the cognitive, fine motor and receptive and expressive language domains compared to the US normative sample[12].

In Australia, studies comparing child performance on the BSID-III at ages 1-, 2- and 3-years to US normative data have been conducted[6,13,14]. Although the directionality of results across age groups has been somewhat inconsistent, similar to the Danish and Dutch studies, results suggest a need for country specific normative data. A study of 156 healthy 3-year-old children reported stronger performance on cognitive, receptive language, expressive language and fine motor domains when compared to the US normative group[13]. Similarly, in a study of 211 extremely preterm/low birth weight and 202 full-term, normal birth 2-year-old children, the BSID-III underestimated delay when US normative data were used[6].

Conversely, a report on 211 healthy children at 1-year of age reported poorer performance on gross motor and expressive language domains and stronger performance on cognitive and receptive language domains when compared to the US normative sample[14]. Ethnicity and child rearing practices were discussed as potential contributors to these differences. Whilst statistical and methodological rigour were considerable strengths of this latter study, limited maternal demographic data were provided. Socio-economic status, level

of parental education and employment status are known to correlate with offspring ability. Assessment of these factors is critical when examining Australian and US infant BSID-III performance to ensure representativeness of results and to understand factors contributing to performance.

This study aimed to: (1) describe the demographic characteristics of a large cohort (n=998) of 1-year-old infants with BSID-III data and draw comparisons between Australian and US data; (2) establish a cohort representative of Australian population demographic characteristics; (3) compare Australian infant performance on the BSID-III at 1-year of age to the US normative sample; and, (4) determine differences across 12- and 13-month age groups.

2.3 Method

2.3.1 Participants

Data were from The Triple B Pregnancy Cohort Study, a prospective longitudinal study of women, partners and their infant offspring[15]. The 1,305 pregnant women in this study were recruited (2009-2013) through public hospital antenatal clinics, birth centres, high-risk antenatal clinics and substance use antenatal clinics across Greater Metropolitan Sydney, NSW. Eligibility criteria were: being pregnant; ≥16 years of age; adequate mental ability and English literacy; intention to reside in Australia until the child's first birthday; no other siblings in the study; intention to remain a primary caregiver for the child; and no known major medical maternal or foetal complications. The BSID-III assessment was completed at 1-year post-partum with 1,151 infants (1,128 families; 86.4 percent retention rate). Further cohort response rate details are presented elsewhere.[15]

Of infants with complete BSID-III data, in order to match US normative sampling techniques exclusion criteria for the current study were: multiple birth infants (n=47);

gestational age <36 weeks at birth (n=19); and recruitment through specialist antenatal clinics dedicated to individuals with complex substance use/mental health needs (n=36). Additionally, 49 infants were excluded due to incomplete data, and two infants were considered outliers based on age at BSID-III completion (9-months and 17-months), yielding a cohort of 998 infants. Informed consent was obtained from mothers, and ethical approval granted by University and NSW Health Human Research Ethics Committees (X08-0127; X12-0232; HREC/08/RPAH/218).

2.3.2 Measures

Demographic data

Maternal demographic data (assessed in Trimester 3) are summarised in Table 1. National Socio-Economic Indexes for Areas (SEIFA) percentile rankings of social advantage and disadvantage[16] were obtained for each family as a measure of socio-economic status. Information on infant offspring was obtained at 8-weeks and 1-year via maternal interview, infant health record books (hospital recorded data), and clinical assessment.

Cohort data were categorised in line with current Australian Bureau of Statistics (ABS) population data for infant sex, maternal age at birth, SEIFA quartiles, maternal region of birth, maternal education, maternal employment status and Aboriginal and Torres Strait Islander identification. Please see Table 1 and relevant references for further details. ABS data for women aged 30- to 34-years were used for categorisation of region of birth, employment status and level of education, consistent with the median age of the current cohort[17].

Infant development assessment

The BSID-III was used to assess infant development at 1-year of age. The BSID-III comprises five objectively rated domains of development: cognition, receptive language, expressive language, fine motor and gross motor[4]. The BSID-III has robust psychometric properties with moderate to strong reliability and validity[4,18]. As raw data were not available for the US normative sample, raw scores obtained on each BSID-III domain were converted to standardised scaled scores (M=10, SD=3) based on the US normative sample using Pearson Psychcorp scoring software. Scores were also manually cross-checked for accuracy.

2.3.3 Procedure

The BSID-III was administered by trained assessors with a university undergraduate degree in social sciences. Assessments were either conducted in a quiet space in the child's home or in an interview room at The University of New South Wales (UNSW). To achieve reliability in assessments, staff administered at least 10 assessments under the supervision of a trained assessor and were required to achieve 100 percent scoring agreement on two consecutive assessments prior to independent administration. Two trained assessors independently attended and scored 27 (2.4 percent) of all 1-year assessments to determine inter-rater reliability. Ongoing reliability checks were also conducted (on approximately 15 percent of the cohort).

The final cohort of 998 infants was categorised into two groups according to age at time of BSID-III completion, consistent with US normative sampling techniques: (1) n=657 '12-month infants' aged 11-months-16-days to 12-months-15-days and (2) n=268 '13-month infants' aged 12-months-16-days to 13-months-15-days[4]. Seventy-three infants who completed the BSID-III assessment outside 11-months-16-days and 13-months-15-days

remained in the total cohort analyses, but were excluded from 12- and 13-month age group analyses.

2.3.4 Statistical Analysis

Statistical analyses were conducted using the software IBM SPSS[19] version 22, Stata version 8 SE was used to obtain raked weights, and GPower 3.0.10[20] was used to compute Cohen's '*d*' effect sizes. Demographic variables were summarised using frequency and descriptive statistics. Aim two was achieved by obtaining raked weights for participants, with a fixed maximum value of 5[21,22,23]. Percentages applied to obtain raked weights were based on Australian population data for all variables outlined in Table 1. Aims three and four were assessed using one sample t-tests. Mean scaled scores of the overall cohort, as well as 12- and 13-month age groups, were compared to the age matched US normative sample. This was repeated for each domain on the BSID-III. Bonferroni adjusted significance level was set at 0.003 for multiple *a priori* analyses[24].

2.4 Results

2.4.1 Cohort Characteristics

Maternal demographic characteristics did not differ between the two age groups (Appendix A), as such, Table 1 summarises characteristics for the full cohort (N=998). Two participants declined to answer questions about Aboriginal Torres Strait Islander status. Male infants made up 51.6 percent of the current cohort, compared to 50 percent in the BSID-III US normative sample. When comparing the current maternal cohort to Australian population data, the cohort was more: economically advantaged (SEIFA); likely to have a University or College degree; likely to be older; to be born in a different geographical region, namely the United Kingdom; and less likely to report Aboriginal Torres Strait Islander origins.

Table 1. Demographic characteristics of the study cohort, study cohort with weights applied, Australian population and US population

Demographic characteristics	**Unweighted cohort (N=998), No. (%)**	**Australian population, %**	**Weighted cohort (N=996), %**	**US population/ BSID-III US normative data**
Infant gender	Female = 483 (48.4) Male = 515 (51.6)	Female = 48.6 Male = 51.4 (2016[26])	Female = 48.6 Male = 51.4	Equal numbers males and females per age bracket (BSID-III).[4]
Maternal age at birth (yrs)	Mean age was 33yrs. Proportions by age range categorized below. ≤19 = 4 (0.4) 20-24 = 39 (3.9) 25-29 = 197 (19.7) 30-34 = 363 (36.4) 35-39 = 321 (32.2) 40-44 = 68 (6.8) 45+ = 6 (0.6)	Median age in 2015 was 31yrs. Proportions by age range categorized below[27]. ≤19 = 2.8 20-24 = 12.7 25-29 = 27.2 30-34 = 35.0 35-39 = 18.0 40-44 = 4.0 45+ = 0.3	≤19 = 2.0 20-24 = 12.8 25-29 = 27.4 30-34 = 35.3 35-39 = 18.1 40-44 = 4.0 45+ = 0.3	In 2014 the average age for women giving birth was 28yrs[28].
SEIFA quartiles	1st = 31 (3.1) 2nd = 65 (6.5) 3rd = 263 (26.4) 4th = 639 (64.0)	1st= 24.2 2nd=24.6 3rd=25.1 4th=26.1[16]	1st= 15.6 2nd=27.4 3rd=28.0 4th=29.1	NA

Demographic characteristics	Unweighted cohort (N=998), No. (%)	Australian population, %	Weighted cohort (N=996), %	US population/ BSID-III US normative data
Maternal region of birth[b]	Australia = 560 (56.1) Asia =109 (10.9) UK = 117 (11.7) NZ = 43 (4.3) Europe other = 72 (7.2) Other = 97 (9.7)	Australia = 66.4 Asia = 16.5 UK = 3.4 NZ = 3.1 Europe other = 3.7 Other = 7.0[18]	Australia = 66.5 Asia = 16.5 UK = 3.4 NZ = 3.1 Europe other = 3.8 Other = 6.8	According to BSID-III US 1-year infant data, approximately, 60% of the US population is white, followed by 19% Hispanic, 15% African American and 4% Asian[4].
Maternal education	School completion or less = 136 (13.6) Post school certificate, diploma or trade =134 (13.4) Bachelor degree or higher = 728 (72.9)	School completion or less =26.3 Post school certificate, diploma or trade = 31.1 Bachelor degree or higher = 42.6[29]	School completion or less = 26.3 Post school certificate, diploma or trade = 31.1 Bachelor degree or higher = 42.6	Approximately 60% of parents in the US BSID-III normative sample had completed formal education beyond the 12-years of schooling[4].
Maternal employment status	Employed = 705 (70.6)	Women aged 30-34 years in July, 2015[30]. Employed= 69.7	Employed= 69.7	69.9% of mothers with a child aged less than 18 years were employed. Rates ranged from 68.3-74.3% for women aged 25-44 years[31].
ATSI identification	13 (1.3)[a]	2.5[32]	2.5	NA

[a]n=996; [b]Region of birth was made up of the following ABS 2011 categories: Australia; United Kingdom; New Zealand; Southern, Eastern, Northern and Western Europe-"Europe other"; South-East, North-East, Southern and Central Asia-"Asia"; and "Other"[17]; ATSI= Aboriginal Torres Strait Islander.

Weighted results were obtainable for n=996 as raked weights could not be calculated for the two participants with missing Aboriginal and Torres Strait Islander identification data. Weighted demographic data showed that the weighted cohort was representative of the Australian population, with the exception of frequency of the lowest SEIFA category which remained below the national average.

When examining infant outcomes relative to Australian population data, 12.4 percent of infants in the unweighted cohort were transferred to a special care or neonatal intensive care unit; a rate that is slightly lower than the national population average of 15.4 percent[25]. However, this was not unexpected considering the cohort consisted only of term infants to match US sampling techniques, and thus weights were not applied for this variable. Appendix A provides detailed birth and early development information on the cohort offspring.

Comparison of the US and Australian population data show population differences for maternal education level, birth region and maternal age at birth.

2.4.2 BSID-III Scores

Inter-rater reliability for the BSID-III was assessed via intraclass correlation coefficients (ICC) for each of the five objectively rated domains of the BSID-III. The resulting ICCs fell in the *very good (>0.8)* to *excellent (>0.9)* range; *C*ognitive ICC = 0.988, Receptive Language ICC =0.950; Expressive Language ICC = 0.869; Fine Motor ICC =0.889; Gross Motor = 0.988, indicating a high degree of agreement between assessors[33,34].

BSID-III scaled scores for the total cohort and the 12- and 13-month-old infants are provided separately in Table 2. Of note, mean scaled scores in the current cohort were, on average, approximately one point higher than the US BSID-III normative sample on the cognitive domain, and approximately one point lower than the US BSID-III normative sample on the gross motor domain, in both the unweighted and weighted data.

Table 2. BSID-III scaled score distributions

BSID-III domain	Total cohort (N=998) Mean [SD]		12-month infants (N=657) Mean [SD]		13-month infants (N=268) Mean [SD]	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
Cognition	11.45 [2.32]	11.24 [2.34]	11.82 [2.26]	11.57 [2.37]	10.76 [2.34]	10.75 [2.26]
Receptive language	10.15 [2.55]	9.84 [2.62]	10.40 [2.54]	10.15 [2.60]	9.86 [2.48]	9.39 [2.50]
Expressive language	10.20 [1.97]	10.00 [2.01]	10.50 [1.82]	10.28 [1.94]	9.74 [2.07]	9.70 [1.99]
Fine motor	10.53 [2.48]	10.48 [2.36]	10.51 [2.31]	10.39 [2.28]	10.26 [2.77]	10.26 [2.47]
Gross motor	9.07 [2.69]	9.17 [2.73]	9.32 [2.74]	9.40 [2.90]	8.50 [2.55]	8.65 [2.39]

Note: US scaled scores mean=10 SD=3

2.4.3 Comparison of BSID-III with US Normative Sample

Table 3 shows the t-test results and Cohen *d* effect sizes for the total cohort compared to the US BSID-III normative data, and a break-down by 12- and 13-month age groups. Approximately 62 percent of infants in the unweighted Australian cohort performed better than the US normative sample mean on the cognitive domain, but only 33 percent performed consistent with, or higher than, the US normative sample mean on the gross motor domain.

Table 3. Relationships between BSID-III scaled scores for current cohort and US normative sample

BSID-III domain	Total cohort				12-month infants (N=657)				13-month infants (N=268)			
Unweighted Data (total cohort n=998)												
	p	**M_{diff}**	**M_{diff} 95% CI**	***d***	**p**	**M_{diff}**	**M_{diff} 95% CI**	***d***	**p**	**M_{diff}**	**M_{diff} 95% CI**	***d***
Cognition	<.001	1.45	1.31 – 1.60	0.54	<.001	1.82	1.64 – 1.99	0.69	<.001	0.76	0.48 – 1.04	0.28
Receptive language	.06	0.15	-0.01 – 0.31	0.05	<.001	0.40	0.21 – 0.60	0.14	.36	-0.14	-0.44 – 0.16	0.05
Expressive language	.001	0.20	0.08 – 0.32	0.08	<.001	0.50	0.36 – 0.64	0.20	.04	-0.26	-0.51 – -0.02	0.10
Fine motor	<.001	0.53	0.37 – 0.68	0.19	<.001	0.51	0.34 – 0.69	0.19	.12	0.26	-0.07 – 0.59	0.09
Gross motor	<.001	-0.93	-1.10 – -0.76	-0.33	<.001	-0.68	-0.89 – -0.47	-0.24	<.001	-1.50	-1.80 – -1.19	-0.54
Weighted Data (total cohort n=996)												
BSID-III domain	**Total cohort**				**12-month infants (N=657)**				**13-month infants (N=268)**			
	p	**M_{diff}**	**M_{diff} 95% CI**	***d***	**p**	**M_{diff}**	**M_{diff} 95% CI**	***d***	**p**	**M_{diff}**	**M_{diff} 95% CI**	***d***
Cognition	<.001	1.24	1.09 – 1.38	0.46	<.001	1.57	1.39 – 1.76	0.58	<.001	0.75	0.47 – 1.03	0.28
Receptive language	.06	-0.16	-0.32 – 0.00	0.06	.14	0.15	-0.05 – 0.36	0.05	<.001	-0.61	-0.92 – -0.30	0.22
Expressive language	.96	0.00	-0.12 – 0.13	0.00	<.001	0.28	0.13 – 0.44	0.11	.02	-0.30	-0.55 – -0.06	0.12
Fine motor	<.001	0.48	0.34 – 0.63	0.18	<.001	0.39	0.21 – 0.57	0.15	.10	0.26	-0.05 – 0.56	0.09
Gross motor	<.001	-0.83	-1.00 – -0.66	-0.29	<.001	-0.60	-0.83 – -0.37	-0.20	<.001	-1.35	-1.65 – -1.06	-0.50

Note. Bonferroni significance correction for multiple a priori analyses=0.003; 'M_{diff}' = Mean difference between the current cohort mean scaled score and the US BSID-III normative sample mean scaled score of 10.

Although significant differences were detected on the receptive language, expressive language and fine motor domains at 12-months, differences were not found on these domains at 13-months; and overall, effect sizes were small.

Weighted results remained unchanged with the exception of receptive language where, upon weighting, scores were significantly lower than the US normative sample in the 13-month-old group and no significant difference between the weighted cohort and the US normative sample was detected in the 12-month-old group. Cognitive and gross motor scores remained the only domains demonstrating consistent significant differences when compared to the US normative sample, across both 12- and 13-month groups.

2.5 Discussion

The BSID-III is widely used in Australia, but no Australian normative data are available. Determining whether differences are evident between Australian and US infant performance on the BSID-III in a cohort representative of the Australian population, would address an important gap in the extant literature. This study used a longitudinal, prospective cohort of 998 infants aged 11-months 16 days to 13-months 15 days, with statistical weighting, to provide the first large scale, representative cohort of Australian infant performance on the BSID-III.

Results indicated that, on average, Australian infants scored significantly *higher* than US infants on the cognitive domain and significantly *lower* than US infants on the gross motor domain; a result that was consistent among the 12- and 13-month-old groups and after cohort weighting was applied. Taken together, these results suggest that utilitisation of US BSID-III normative data in an Australian population is inappropriate and likely contributing to inaccurate detection of delay. Significant differences between the current weighted cohort and US normative sample were also detected on the expressive language and fine motor

domains at 12-months and on the receptive language domain at 13-months. However, the results were not consistent across both age groups, effect sizes were small, and mean scaled score differences ranged from 0.28 to -0.61, suggesting limited clinical significance, and that statistical differences on these domains may instead be a bi-product of the large sample size.

Examination of population demographic data suggested that Australian mothers are likely to be older than US mothers when giving birth, more likely to have completed formal education post-school, and that the common regions of birth of mothers differs between Australia and the US. These documented differences between the two populations may provide one explanation for the differences detected between Australia and US infants on the gross motor and cognitive domains. The US BSID-III normative sample was reportedly matched to population characteristics of parental education and ethnicity[4]. ABS data demonstrated that approximately 74 percent of Australian women have completed a post-school qualification, compared to 60 percent of parents in the US BSID-III normative sample[4]. Given the known association between parental education and offspring intelligence, this population difference may account for detected differences on the cognitive domain[35].

Furthermore, infants of Hispanic and African American ethnicities form 31 percent of the overall US BSID-III normative sample and infants of Asian ethnicity are represented by 4 percent [4]. Conversely, in Australia, Asian ethnicities make up two of the top five countries of overseas birth[36], and are the most frequent region of overseas birth. Whilst direct comparisons between specific ethnicities are impossible to assess empirically in a cohort representative of the Australian population (due to the low numbers of people of African American and Hispanic origins), ethnicity may play a role in the detected differences in cognition and gross motor. African Americans possess superior gross motor abilities compared to Caucasians, and East Asians possess weaker gross motor abilities[8,37]. As such, the higher proportion of infants of Asian born mothers and low proportion of infants of

Hispanic and African American born mothers in the Australian cohort may account for lower scores obtained on the gross motor domain[38]. Additionally, a US study reported that at 24-months of age infants of African American and Hispanic origins performed between two thirds and three quarters of a standard deviation below white Americans on the cognitive scale of the BSID-III[39]. Furthermore, a study by Rushton and Ankney (1996) of cognitive abilities of children aged 0- to 7-years reported Asians to have highest cognitive abilities, followed by Caucasians, and then Africans[40]; and converging literature suggests that Asians possess strong academic skills[41]. While a variety of socio-economic factors and child rearing practices likely contribute to differing performances across ethnic groups, the disparate composition of the US normative sample and the Australian population may also contribute to the differences detected in cognitive scores.

Another factor that may account the detected differences between the current sample and the US normative sample on the gross motor domain is item order. Anecdotally, many children in the current cohort who did not yet pull to standing or did not bounce were able to cruise around furniture once placed in a standing position. However, bouncing and cruising was required in order to meet basal criteria at 1-year for the gross motor domain. Children were, therefore, regressed to the previous section if they were unable to complete these items (Bayley, 2006a). In the previous section, four items required crawling skills. If the child 'bum shuffled' instead of moving to hands and knees, they could not receive credit for these items. These children therefore lost a considerable number of credit points despite being able to age appropriately cruise furniture, suggesting potentially inappropriate test item order for Australian children.

2.5.1 Strengths and Limitations

This study is the largest cross-cultural examination of the BSID-III in a cohort of 1-year-old infants to our knowledge. Strengths of the current study include the measurement of a range of maternal socio-demographic factors associated with infant development, and the application of raked weights to improve cohort representativeness. Furthermore, this study represents the first Australian study to match infants on two US normative sample age groups: 12- and 13-months of age.

Some limitations of the current study are important to consider when interpreting the results. First, the number of infants that completed the BSID-III in the 12-month age group was considerably greater than those who completed the BSID-III in the 13-month age group. As such, 12-month infants were over-represented in the total cohort calculations. However, differences in cognitive and gross motor performance were consistently detected across both 12- and 13-month groups, suggesting robust results.

Second, although the demographic characteristics of the current cohort were consistent with the Australian population in terms of employment status and infant sex (Table 1), when unweighted, some important differences were noted. SEIFA percentiles and maternal educational attainment suggest that the current cohort was more socio-economically advantaged and educated than the Australian population[10, 29]. Whilst raked weights were applied, weighting was unable to fully remedy frequency discrepancies between the unweighted cohort and Australian population data for the lowest quartile of SEIFA scores. Considering the known connection between socio-economic status and intelligence, this may contribute to the high performance of Australian infants detected on the cognitive domain[35].

Third, although obtaining raked weights is good practice amongst epidemiological studies where cohorts are unrepresentative[21-23], raking is not without risk. If under-represented participants in the cohort are not representative of the category of the population

from which they were drawn, raking will increase bias by attributing greater weight to these participants. However, this is unlikely in the current cohort, as relatively unchanged results following weighting suggests robust results.

Fourth, differences between US normative sampling and the current sampling techniques may provide an explanation for differences detected on the cognitive domain. As outlined, the inclusion/exclusion criteria of the current study were devised to match the US BSID-III normative sample approach as closely as possible. However, following sampling, individuals with specific conditions such as cerebral palsy, trisomy 21, and pervasive developmental disorder were selectively re-introduced into the US BSID-III normative sample, making up approximately 10 percent of the sample. Conversely, the current cohort was healthy, term, singleton infants. While this was unavoidable as the proportion of children with expected difficulties in each age group of the US BSID-III normative sample was not clearly reported[4], and therefore not replicable, it is possible that sampling differences may account for the higher scores obtained by the current cohort on the cognitive domain when compared to the US normative sample (Bayley, 2006b). However, it does not account for the lower gross motor scores obtained by the current cohort when compared to the US normative sample.

2.5.2 Implications

Differences between infant performance in the current cohort and the US BSID-III normative sample have important implications for the utility of the BSID-III in clinical practice in Australia. Results of our study suggest that utilisation of US normative data to interpret Australian infant performance may result in under identification of cognitive delay and over identification of motor delay. As a result, infants may not receive appropriate and timely interventions for cognitive delays that are known to reduce the social and economic

burden of disability. Conversely, they may undergo unnecessary treatments for incorrectly diagnosed motor delays.

Importantly, this study provides the first attempt to ensure representativeness of cohort demographic characteristics through raked weights. While the potential contribution of two important population factors have been discussed, a range of additional population factors, such as access to public health services, neonatal practices, Gross Domestic Product per capita, and national wealth distribution, may also be contributing to the detected, between country, differences on the gross motor and cognitive domains. Regardless of the reasons for differences, through ensuring the representativeness of the current cohort with regards to Australian population demographic information, this study allows us to conclude that detected differences on cognitive and gross motor domains likely result from true population differences.

As such, weighted means and standard deviations reported in this paper may be used by clinicians to calculate z-scores, as a form of Australian normative data to assist in the interpretation of Australian 12- and 13-month infant performance on the BSID-III (see Table 2.). Results support the need for future research examining the appropriateness of international normative group utilisation for country specific interpretation, and suggest benefits in establishing full age range, country-specific BSID-III normative data.

2.6 References

1. Illingworth RS. *The Development of the Infant and the Young Child: Normal and Abnormal*. 10th ed. New Delhi, India: Elsevier Health Sciences; 2013.

2. Australian Bureau of Statistics (ABS). *Australian Social Trends June 2012: Children with a Disability*. Catalogue no: 4102.0. Canberra: ABS. http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442455787. Accessed November 10th, 2016.

3. Johnson S, Marlow N. Developmental screen or developmental testing? *Early Hum Dev.* 2006;*82*(3):173-183.

4. Bayley N. *The Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III).* 3rd ed. San Antonio, Texas: Pearson; 2006.

5. Walker K, Holland AJ, Halliday R, Badawi N. Which high-risk infants should we follow-up and how should we do it? *J Paediatr Child Health*. 2012;48(9):789-93.

6. Anderson PJ, De Luca CR, Hutchinson E, Roberts G, Doyle LW. Underestimation of developmental delay by the new Bayley-III Scale. *Arch Pediatr Adolesc Med*. 2010;164(4):352-356.

7. Luttikhuizen dos Santos ES, de Kieviet JF, Konigs M, van Elburg RM, Oosterlaan J. Predictive value of the Bayley Scales of Infant Development on development of very preterm/very low birth weight children: A meta-analysis. *Early Hum Dev.* 2013;89(7):487-496.

8. Angulo-Barroso RM, Schapiro L, Liang W, et al. Motor development in 9-month-old infants in relation to cultural differences and iron status. *Dev Psychobiol.* 2011;53(2):196-210.

9. Yu YT, Hsieh WS, Hsu CH, et al. A psychometric study of the Bayley Scales of Infant and Toddler Development - 3rd Edition for term and preterm Taiwanese infants. *Res Dev Disabil.* 2013;34(11):3875-3883.

10. Australian Bureau of Statistics. *2071.0 – Cultural Diversity in Australia. Reflecting a Nation: Stories from the 2011 Census.* http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/2071.0main+features902012-2013. Released May 21st, 2012. Accessed November 10th, 2016.

11. Krogh MT, Vaever MS, Harder S, Koppe S. Cultural differences in infant development during the first year: A study of Danish infants assessed by the Bayley-III and compared to the American norms. *Eur J Dev Psychol*. 2012;9(6):730-736.

12. Steenis LJ, Verhoeven M, Hessen DJ, Van Baar AL. Performance of Dutch children on the Bayley III: a comparison study of US and Dutch norms. *PloS One*. 2015;10(8):p.e0132871. http://dx.doi.org/10.1371/journal.pone.0132871. Accessed November 10th, 2016.

13. Chinta S, Walker K, Halliday R, Loughran-Fowlds A, Badawi N. A comparison of the performance of healthy Australian 3-year-olds with the standardised norms of the Bayley Scales of Infant and Toddler Development (version-III). *Arch Dis Child*. 2014;99(7):621-624.

14. Walker K, Badawi N, Halliday RJ, et al. Brief report: performance of Australian children at one year of age on the Bayley scales of infant and toddler development (version III). *Aust Educ Dev Psychol.* 2010;27:54–8.

15. Hutchinson D, Wilson J, Allsop S, et al. Cohort profile: The Triple B Pregnancy Cohort Study: A longitudinal study of the relationship between alcohol, tobacco and other

substance use during pregnancy and the health and wellbeing of Australian children and families. *Int J Epidemiol.* 2017; accepted manuscript.

16. Australian Bureau of Statistics. *2033.0.55.001 - Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2011.* http://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001. Released March 28th, 2013. Accessed November 10th, 2016.
17. Australian Bureau of Statistics. 34120DO006_201415 Migration, Australia, 2014-15. *Stat dataset: Estimated Resident Population by Country of Birth, 30 June 1992 to 2015.* http://stat.data.abs.gov.au/Index.aspx?Queryid=87. Accessed March 13th, 2017.
18. Albers CA, Grieve AJ. Review of Bayley Scales of Infant and Toddler Development. *J Psychoeduc Assess.* 2007; 25(2):180-190.
19. IBM Corp. *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: IBM Corp. Released 2013.
20. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;*39:*175-191.
21. Dal Grande E, Chittleborough CR, Campostrini S, Tucker G, Taylor AW. Health estimates using survey raked-weighting techniques in an Australian population health surveillance system. *Am J Epidemiol.* 2015;182(6):544-556.
22. Boboth Z, Zimmer A, Scott M, Hilarides E, Gromet D, Massey C. *Weighting for Coverage Bias in Internet Surveys.* 2007. http://www.sci.sdsu.edu/math-reu/2007-6.pdf Accessed March 13th, 2017.
23. Kolenikov S. Calibrating survey data using iterative proportional fitting (raking). *Stata J.* 2014;14(1):22-59.

24. Westfall PH, Johnson WO, Utts JM. A Bayesian perspective on the Bonferroni adjustment. *Biometrika.* 1997;*84*(2):419-427.

25. Australian Bureau of Statistics. *Labour force, Australia: Labour Force Status and Other Characteristics of Families, June 2012*. http://abs.gov.au/AUSSTATS/abs@.nsf/mediareleasesbyCatalogue/D2310F1B7D3020B2CA25801B0018B7F9?OpenDocument. Released May 1st, 2013. Accessed November 10th, 2016.

26. Australian Bureau of Statistics. *3101.0 - Australian Demographic Statistics, Jun 2016: Table 59- Estimated Resident Population By Single Year Of Age, Australia.* http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Jun%202016?OpenDocument. Released December 15th, 2016. Accessed, January 10th, 2017.

27. Australian Bureau of Statistics. *3301.0- Births, by Nuptiality, by Age of Mother.* http://stat.data.abs.gov.au/Index.aspx?QueryId=498#. Released November 8th, 2016. Accessed January 10th, 2017.

28. Mathews T, Hamilton BE. Mean age of mothers is on the rise: United States, 2000-2014. *NCHS data brief.* 2016;*232:*1-8.

29. Australian Bureau of Statistics. *6227.0 - Education and Work, Australia, May 2015.* http://www.abs.gov.au/ausstats/abs@.nsf/mf/6227.0/. Released November 27th, 2015. Accessed November 10th, 2016.

30. Australian Bureau of Statistics. *6291.0.55.001- Labour force, Australia, Table 01. Labour force status by Age, Social marital status, and Sex.* http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/6291.0.55.001Jan%202017?OpenDocument. Accessed Decmeber 12th, 2016.

31. United States Department of Labor. *Labor Force Participation rates. Labor force participation rate by age and sex, 2015 annual averages and 2024 projections.* https://www.dol.gov/wb/stats/Laborforce_par_rate_age_sex_2015_txt.htm. Accessed November 17th, 2016.

32. Australian Bureau of Statistics. *Census of Population and Housing - Counts of Aboriginal and Torres Strait Islander Australians, 2011.* http://www.abs.gov.au/ausstats/abs@.nsf/lookup/2075.0main+features32011. Released June 21st, 2012. Accessed November 17th, 2016.

33. Landers RN. Computing intraclass correlations (ICC) as estimates of interraterreliability in SPSS. *The Winnowner 2*. 2015. DOI: 10.15200/winn.143518.81744.

34. Shrout P, Fleiss J. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull.* 1979;86(2):420-428.

35. Forns J, Julvez J, García-Esteban R, et al. Maternal intelligence-mental health and child neuropsychological development at age 14 months. *Gac Sanit.* 2012;26(5):397-404.

36. Australian Bureau of Statistics. *3412.0 - Migration, Australia, 2014-15*. http://www.abs.gov.au/ausstats/abs@.nsf/mf/3412.0/. Released March 30th, 2016. Accessed November 17th, 2016.

37. Onis M. Assessment of sex differences and heterogeneity in motor milestone attainment among populations in the WHO Multicentre Growth Reference Study. *Acta Paediatr*. 2006;95(S450):66-75.

38. Greene MM, Patra K, Silvestri JM, Nelson MN. Re-evaluating preterm infants with the Bayley-III: Patterns and predictors of change. *Res. Dev. Disabil.* 2013;34(7):2107-17.

39. Halle T, Forry N, Hair E, et al. *Disparities in Early Learning and Development: Lessons from the Early Childhood Longitudinal Study – Birth Cohort (ECLS-B).* Washington, DC: Child Trends; 2009.

40. Rushton JP, Ankney CD, Brain size and cognitive ability: Correlations with age, sex, social class, and race. *Psychon Bull Rev.* 1996;3:21-36.

41. Kao G, Thompson JS. Racial and ethnic stratification in educational achievement and attainment. *Annu Rev Sociol.* 2003;29:417-442.

2.7 Appendix

Appendix A: Unweighted maternal demographic characteristics by age group

Demographic characteristics	**Total cohort, No. (%) N=998**	**12-month infants, No. (%) N=657**	**13-month infants, No. (%) N=268**
Infant gender	Female = 483 (48.4)	Female = 327 (50.2)	Female = 127 (47.4)
	Male = 515 (51.6)	Male = 330 (49.8)	Male = 141 (52.6)
Maternal age at birth (yrs)	Mean age = 33yrs.	Mean age = 33yrs.	Mean age = 33yrs.
	Proportions by age range:	Proportions by age range:	Proportions by age range:
	≤19 = 4 (0.4)	≤19 = 2 (0.3)	≤19 = 2 (0.7)
	20-24 = 39 (3.9)	20-24 = 27 (4.1)	20-24 = 5 (1.9)
	25-29 = 197 (19.7)	25-29 = 133 (20.2)	25-29 = 48 (17.9)
	30-34 = 363 (36.4)	30-34 = 231 (35.2)	30-34 = 101 (37.7)
	35-39 = 321 (32.2)	35-39 = 216 (32.9)	35-39 = 92 (34.3)
	40-44 = 68 (6.8)	40-44 = 45 (6.8)	40-44 = 18 (6.7)
	45+ = 6 (0.6)	45+ = 3 (0.5)	45+ = 2 (0.7)
SEIFA quartiles	1st = 31 (3.1)	1st= 20 (3.0)	1st= 6 2.2)
	2nd = 65 (6.5)	2nd= 40 (6.1)	2nd=18 (6.7)
	3rd = 263 (26.4)	3rd= 175 (26.6)	3rd=68 (25.4)
	4th = 639 (64.0)	4th= 422 (64.2)	4th= 176 (65.7)

Demographic characteristics	**Total cohort, No. (%) N=998**	**12-month infants, No. (%) N=657**	**13-month infants, No. (%) N=268**
Maternal region of birth[b]	Australia = 560 (56.1) Asia =109 (10.9) UK = 117 (11.7) NZ = 43 (4.3) Europe other = 72 (7.2) Other = 97 (9.7)	Australia = 383 (58.3) Asia = 61 (9.3) UK = 80 (12.2) NZ = 24 (3.7) Europe other = 53 (8.1) Other = 56 (8.5)	Australia = 140 (52.2) Asia = 37 (13.8) UK = 31 (11.6) NZ = 18 (6.7) Europe other = 14 (5.2) Other = 28 (10.4)
Maternal education	School completion or less = 136 (13.6) Post school certificate, diploma or trade =134 (13.4) Bachelor degree or higher = 728 (72.9)	School completion or less = 86 (13.1) Post school certificate, diploma or trade = 77 (11.7) Bachelor degree or higher = (494 (75.2)	School completion or less = 39 (14.6) Post school certificate, diploma or trade = 40 (14.9) Bachelor degree or higher = 189 (70.5)
Maternal employment status	Employed = 705 (70.6)	Employed = 456 (69.4)	Employed = 197 (73.5)
ATSI identification	13 (1.3)[a]	9 (1.4)[a]	2 (0.7)

[a]n=996; [b]Region of birth was made up of the following ABS 2011 categories: Australia; United Kingdom; New Zealand; Southern, Eastern, Northern and Western Europe-"Europe other"; South-East, North-East, Southern and Central Asia-"Asia"; and "Other"[17]; ATSI= Aboriginal and Torres Strait Islander.

Appendix B: Infant cohort characteristics

	Total cohort			12-month infants			13-month infants		
	n	**Mean [SD]**	**Range**	**n**	**Mean [SD]**	**Range**	**n**	**Mean [SD]**	**Range**
Gestation (wks)	998	39.53 [1.33]	36-43	657	39.54 [1.32]	36-43	268	39.54 [1.39]	36-42
Birth weight (kg)	997	3.48 [0.05]	2-5	656	3.47 [0.05]	2-5	268	3.52 [0.05]	2 -5
Birth length (cm)	987	51.01 [2.68]	40-62	650	51.00 [2.61]	43-62	265	51.15 [2.75]	40-58
Birth head circumference (cm)	860	34.75 [1.53]	30-43	561	34.70 [1.49]	30-43	236	34.92 [1.64]	30-40
Apgar score 1 min	923	8.54 [1.26]	1-10	613	8.54 [1.30]	1-10	246	8.55 [1.19]	1-10
Apgar score 5 min	922	8.95 [0.66]	1-10	612	8.94 [0.66]	2-10	246	8.99 [0.70]	1-10
Weight 1-year (kg)	970	10.13 [0.12]	7-15	637	10.06 [0.12]	7 -15	260	10.25 [0.12]	7-14
Length 1-year (cm)	963	76.29 [3.47]	64-90	633	76.19 [3.48]	64-90	257	76.29 [3.46]	65-87
Head circumference 1-year (cm)	945	46.56 [1.67]	39-56	619	46.51 [1.65]	41-56	253	46.58 [1.80]	39-56
BSID-III completion age (days)	998	380 [18]	324-509	657	371 [5]	352-380	268	390 [8]	381-410

Note. The n differs across groups based on available data. Unobtained data were treated as missing.

Study 2

Utility of the Bayley Scales of Infant and Toddler Development-3rd Edition (BSID-III) to distinguish 1-year-old infants at perinatal risk of neurodevelopmental delay.

Ingrid Honan*, BPsych (Hons), Delyse Hutchinson, PhD, Melanie Porter, PhD, Steve Allsop, PhD, Jake Najman, PhD, Elizabeth J. Elliott, MD, Lucinda Burns, PhD, Sue Jacobs, MBBS, Craig A. Olsson, PhD, Richard P. Mattick, PhD

Paper has been prepared in accordance with authorship guidelines for:

Child Neuropsychology

3.1 Abstract

The aim of this study was to investigate the utility of the Bayley Scales of Infant and Toddler Development-III (BSID-III) to detect group differences in infant performance at 1-year of age, based on indirect perinatal risk factors associated with neurodevelopmental delay (NDD). Infants were drawn from a population-based cohort (N=935) and were dichotomously categorised as 'low risk' or 'at risk' according to seven antenatal and/or perinatal risk factors commonly associated with indirect NDD. The BSID-III was administered to infants at 1-year of age and scaled scores calculated.

Results indicated that infants at indirect perinatal risk were more likely to be male and born to mothers of lower socio-economic status. At risk infants scored statistically significantly lower than low risk infants on the cognitive and gross motor domains of the BSID-III, a result that remained significant after controlling for potential covariates. However, mean scaled score differences were negligible (mean scaled scores on the cognition and gross motor domains were 0.57 and 0.44, respectively); likewise, the proportion of variance explained by risk status was small (1.5 percent and 0.5 percent), suggesting limited clinical significance. Between group differences were not detected on the receptive language, expressive language or fine motor domains following adjustment. As such, the BSID-III appears to hold little clinical utility in detecting delay when administered to 1-year-old infants at indirect perinatal risk of NDD. Further research is necessary to determine whether language and fine motor delays documented in at risk samples are expressed in infancy or become evident as higher order skills develop in later childhood.

3.2 Introduction

A range of perinatal factors are known to place infants at increased risk of developmental delay. Risk factors such as hydrocephalus (Futagi, Suzuki, Toribe, & Morimoto, 2002), severe encephalopathy (de Vries & Jongmans, 2010; Dixon, et al., 2002; Gonzalez & Miller, 2006; Van Handel, Swaab, De Vries, & Jongmans, 2007) and intracranial haemorrhage (Patra, Wilson-Costello, Taylor, Mercuri-Minich, & Hack, 2006) cause detectable neurological signs and an increased likelihood of neurodevelopmental delay (NDD). Infants with these perinatal risk factors for NDD are typically detected in the first weeks of life through rigorous screening and early detection procedures (e.g., imaging and neurological assessment).

In contrast, *indirect perinatal risk factors* are perinatal risk factors that do not necessarily result in detectable neurological signs, yet still place infants at an increased risk of NDD (Westrupp, Mensah, Giallo, Cooklin, & Nicholson, 2012). Indirect perinatal risk factors include prematurity (Bos & Roze, 2011; Greene, Patra, Nelson, & Silvestri, 2012), low birth weight (de Moura et al., 2010), small head circumference (Peterson, Taylor, Minich, Klein, & Hack, 2006), admission to neonatal intensive care unit (NICU) and/or special care units (SCU) post-birth (Walker, Holland, Halliday, & Badawi, 2012), low Apgar scores (Odd, Rasmussen, Gunnell, Lewis, & Whitelaw, 2008), maternal substance use (Bandstra, Morrow, Mansoor, & Accornero, 2010; Huizink, 2014), and multiple birth infants (Wadhawan et al., 2011). The extent of NDD in infants solely affected by indirect risk factors is typically subtler than for infants exposed to direct perinatal risks. Yet these infants are commonly referred to developmental follow-up clinics for monitoring of milestone attainment and potential developmental delay.

The Bayley Scales of Infant and Toddler Development III (BSID-III) is widely used in paediatric hospitals and developmental follow-up clinics internationally to detect infants

with NDD and to inform intervention (Anderson, De Luca, Hutchinson, Roberts, & Doyle, 2010; Luttikhuizen dos Santos, de Kieviet, Königs, van Elburg, & Oosterlaan, 2013; Walker, Badawi, Halliday, & Laing, 2010). The ability of the BSID-III to accurately detect differences in performance among infants affected by indirect perinatal risk factors, compared to infants with no known risk factors, is therefore crucial. Yet, the clinical utility of the BSID-III to detect delays has come into question (Anderson & Burnett, 2017).

There is now a substantive body of research on the use of the BSID-III to examine neurodevelopmental outcomes associated with specific perinatal risk factors (e.g. De Jesus et al., 2013; Jarjour, 2015; Logan, Brown, & Hayes, 2013; Skiöld et al., 2012). Whilst the BSID-III reportedly overestimates ability and therefore underestimates delay, at risk infant performance has been shown to significantly differ from low risk infant performance (Anderson et al., 2010). Yet, many of these studies have been conducted in sub-populations with one primary risk factor, such as very preterm infants (Greene et al., 2012). Although these studies provide important information about relationships between the BSID-III and the specific risk factor studied, risk factors frequently co-occur (de Moura et al., 2010; Greene et al., 2012). Examination of primary risk factor sub-populations may provide an overly simplified view of real world experiences and lead to inaccurate conclusions. Few general population-based cohort studies are available that have used the BSID-III to assess heterogenous, population-based perinatal risks. Investigating whether the BSID-III can identify differences in performance between those at indirect risk compared to those not at risk/low risk of NDD, in a population-based cohort, will provide important information pertaining to its value as a detection and intervention planning tool in the community.

Further, much of the research on neurodevelopmental outcomes for perinatal risk has been conducted on cohorts of infants 18-months of age or older; there is scant research available on younger infants. Yet, early interventions administered between birth and 18-

months demonstrate positive outcomes for infants (Blauw-Hospers & Hadders-Algra, 2005), suggesting that early infancy is an important time for assessment (Velikos et al., 2015; American Academy of Pediatrics, 2004). One study of 120 preterm infants examined the relationship between risk factors for NDD and BSID-III outcomes. This study identified a range of biological (e.g., sex), medical (e.g., ventilation post birth) and environmental (e.g., parent education) factors associated with poorer BSID-III outcomes at 1-year of age (Velikos et al., 2015). Similarly, an investigation of 85 preterm infants admitted to neonatal intensive care post-birth, who completed the BSID-III at 8- to 12-months of age, demonstrated that specific medical and environmental risk factors were able to predict BSID-III outcome (Greene et al., 2012). Furthermore, a study of 227 very preterm children examining predictors of language ability at 5-years of age reported associations between both perinatal (e.g., birth weight) and environmental factors (e.g., maternal education) and 5-year language ability, demonstrating the importance of controlling for environmental risk factors associated with development when examining perinatal risk factors (Howard et al., 2011). However, to date, no study of which we are aware has used the BSID-III to examine multiple perinatal risk factors in a general population-based cohort of young infants.

The aims of this study were: (1) to provide BSID-III outcome data for a general population-based cohort of 1-year-old infants with indirect perinatal risk factors for NDD, and those without; (2) to examine the relationship between BSID-III scores of infants with indirect perinatal risk factors for NDD compared to those without, controlling for biological, medical and environmental covariates; and, (3) to examine which (if any) perinatal risk factors are individually associated with poorer BSID-III outcomes, and the degree to which each factor contributes to overall indirect risk status.

3.3 Method

3.3.1 Participants

Data were from The Triple B Pregnancy Cohort Study, a prospective longitudinal study of pregnant women, partners and their infant offspring (post-birth) (Hutchinson et al., 2017; Tay et al., 2017). The 1,305 pregnant women were recruited (2009-13) via three public hospitals in metropolitan NSW, Australia. Women were approached at general antenatal clinics, birth centres, high-risk antenatal clinics and substance use antenatal clinics. Eligibility criteria included: being pregnant; ≥16 years of age; adequate mental ability and English literacy; intention to reside in Australia until the child's first birthday; no previous siblings enrolled in the study; intention to remain one of the primary caregivers for the child; and no known major medical complications for the mother or fetus (e.g., chromosomal abnormalities). The BSID-III was completed at 1-year post-partum by 1,151 infants (1,128 families) (86.4 percent retention rate). See Hutchinson et al. (2017) for more details on the cohort assessments and response rates. Ethical approval was granted by relevant University and NSW Health Human Research Ethics Committees (approval number: X12-0232; HREC/08/RPAH/218), and written informed consent was obtained from each participant.

Of the infants with BSID-III scores available, 33 were excluded due to missing baseline or 1-year mother interview data. Multiple birth infants made up n=43 of the remaining cohort (21 twin sets and one triplet set). In order to meet assumptions of independence of observations, 50 percent of first born and 50 percent of second+ born twins were randomly selected to remain in the cohort (n=21). A further 159 infants who had missing data for at least one risk factor but had been classified as 'low risk' based on other available risk data, were excluded from analyses, as overall risk status could not be confirmed (note: infants that were missing data on a risk factor but, could be defined as 'at risk' based on other available data, remained in the study). Two additional children were excluded due to

complex health needs requiring in excess of 30 hospitalisations since birth (one being a multiple birth infant), resulting in a cohort of n=935.

3.3.2 Measures and Procedure

Demographics

Maternal demographic information included: age; maternal education; employment status; relationship status; country of birth; Aboriginal and Torres Strait Islander origin (yes/no); number of biological children and postcode. These were assessed via structured interview in Trimester 3. Postcode was used to determine National Socio-Economic Indexes for Areas (SEIFA) percentile rankings of social advantage and disadvantage (Australian Bureau of Statistics [ABS], 2011) for each family as a measure of socio-economic status. Number of biological children was dichotomously categorised according to whether the child in the study was an only child or had siblings. Country of birth data were categorised by the official language of that country into a dichotomous 'English speaking country of birth' variable. Employment status and relationship status were dichotomised by yes/no responses to the following questions: "employed full-time, part-time or casually"; "in a married or defacto like relationship".

Infant hospitalisation since birth data were collected at 1-year of age via structured maternal interview and was dichotomously categorised yes/no according to responses to "was your child admitted hospital for at least one night, for any reason, in the past 12-months?".

Infant Risk Status

Data on infant risk status were collected via a structured maternal interview in Trimester 3, at 8-weeks postnatally, and from infant birth record books (hospital-recorded data). The following indirect risk factors for developmental difficulties were included.

(1) *Mother recruited through substance use clinic* (i.e., opioid replacement therapy or identified as using high levels of tobacco or alcohol at initial antenatal appointment) (Bandstra et al., 2010; Huizink, 2014).
(2) *Multiple birth infant* (Olusanya, 2011; Wadhawan et al., 2011).
(3) *Preterm* (<36 gestation) (Bos & Roze, 2011; Greene et al., 2012; McGowan, Alderdice, Holmes, & Johnston, 2011; Soria-Pastor et al., 2009).
(4) *Low birth weight* (<2.5kg) (de Moura et al., 2010; Shenkin, Starr, & Deary, 2004).
(5) *Low Apgar score* (<7 at either 1 or 5 minutes) (Odd et al., 2008)
(6) *Small head circumference* (≤3rd percentile) (Peterson et al., 2006; Chung, 2009).
(7) *Admission to special care or neonatal intensive care unit (NICU) post birth* (Walker et al., 2012; Aylward, 2003; Fallah, Islami, & Mosavian, 2011)

Each of these factors has been associated with increased risk of NDD. As such, infants who experienced one or more risk factor were categorised in the 'at risk' group (n=265); all other infants were classed as 'low risk' (n=670).

BSID-III

The BSID-III was used to assess infant development at 1-year of age (Bayley, 2006a). The BSID-III comprises five objectively rated domains of development: cognition, receptive language, expressive language, fine motor and gross motor, as well as a caregiver rated questionnaire assessing social-emotional development and adaptive behaviour (Bayley, 2006a). The BSID-III has robust psychometric properties with moderate to strong reliability and validity (Albers & Grieve, 2007; Bayley, 2006b). Raw scores obtained on each objectively rated BSID-III domain were converted to standardised scaled scores (M=10, SD=3) based on the US normative sample using Pearson Psychcorp scoring software. Scores were also manually cross-checked for accuracy. Consistent with the sensitivity and specificity

literature on the BSID-III, scores greater than one standard deviation below the mean (scaled scores of <7) were classified as evidence of NDD (Anderson et al., 2010; Johnson, Moore, & Marlow, 2014).

The BSID-III was administered by trained assessors with a university undergraduate degree in social sciences. Assessments were either conducted in a quiet space in the child's home or in an interview room at The University of New South Wales (UNSW). To achieve reliability in assessments, staff administered at least ten assessments under the supervision of a trained assessor and were required to achieve 100 percent scoring agreement on two consecutive assessments prior to independent administration. Two trained assessors independently attended and scored 27 (2.4 percent) of all 1-year assessments to determine inter-rater reliability. Ongoing reliability checks of approximately 15 percent of assessments occurred through the assessment timeframe. Inter-rater reliability was assessed via intraclass correlation coefficients (ICC) for each of the five objectively rated sub-domains of the BSID-III with results falling in the *very good (>0.8)* to *excellent (>0.9)* range (Landers, 2015; Shrout & Fleiss, 1979).

3.3.3 Statistical Analyses

Statistical analyses were conducted using the software IBM SPSS Statistics for Windows, version 23 (IBM, 2014). The primary outcome variable was infant performance, as measured by scaled scores on the five objectively assessed BSID-III domains at 1-year. Demographic variables were summarised using frequency and descriptive statistics. Pearson chi-squared (χ^2**)** and independent sample t-test analyses were conducted to compare 'at risk' and 'low risk' infants.

Regression models were fitted to determine whether cognitive, receptive language, expressive language, fine motor and gross motor scaled scores differed based on "risk" status,

controlling for: infant sex and hospitalisations since birth; and, maternal age at birth; linguistic background; employment status; educational attainment; socio-economic status and parity. Assumptions of homoscedasticity, normality of errors and multicollinearity were assessed by examination of Q-Q plots, skewness and kurtosis statistics (+/-1.5) (Tabachnick & Fidell, 2013), histogram and scatterplots of standardised residuals to fitted values, and by examining tolerance (>0.10) and variance inflation factors (<10) (SPSS Web Books Regression with SPSS Chapter 2 – Regression Diagnostics). Where assumptions were violated, statistical outliers were removed, necessary transformations were performed and analyses rerun. Results remained unchanged. Full cohort, untransformed data were reported for ease of interpretation.

Independent sample t-tests were conducted between individual risk factors and BSID-III domains in order to examine which risk factors were associated with differences in BSID-III performance. Exploratory and confirmatory factor analyses were performed using Maximum Likelihood and Varimax Rotation to determine the amount of explained variance in overall indirect risk status accounted for by the seven indirect risk factors included in the study. An eigenvalue cut-off of 1.0 was adopted for the exploratory analysis (Yong & Pearce, 2013) and extracted factors were restricted to one for the confirmatory analysis.

3.4 Results

3.4.1 Sample Characteristics

The overall sample was relatively high functioning compared to the Australian population, with a mean SEIFA percentile score of 77 (compared to a population mean of 50) and a high proportion of participants with university or college degrees (see Table 1.) (ABS, 2011; ABS, 2015). Consistent with expectations considering cohort selection criteria, slightly fewer infants in the current sample were re-hospitalised in the first year of life (13 percent),

compared to the NSW population (16.5 percent) (Lain, Roberts, Bowen, & Nassar, 2014), with number of hospitalisations ranging from zero to four and the most common causes being infections and croup.

Consistent with expectations, at risk infants were more likely to be male and to have been hospitalised overnight since birth than low-risk infants. Furthermore, mothers of at risk infants reported lower socio-economic status, educational attainment, paid employment and were less likely to have been married or in a defacto relationship, compared to mothers of low risk infants (see Table 1).

Table 1. Cohort characteristics by risk status with between group correlations/associations.

Sample Characteristics	**Total** ***n*** **= 935**	**Low risk** ***n*** **= 670**	**At risk** ***n*** **= 265**	
		M (*SD*)		**t (*p*)**
SEIFA percentile	77.17 (21.17)	78.28 (20.21)	74.34 (23.22)	6.71 (.01)
Maternal age	33.03 (4.81)	32.96 (4.78)	33.21 (4.89)	-0.70 (.49)
		***n* (%)**		**χ^2 (*p*)**
Married/Defacto	881 (94.2)	647 (96.6)	234 (88.3)	23.84 (<.001)
Employed	663 (70.9)	489 (73.0)	174 (65.7)	4.94 (.03)
English speaking birth country	720 (77.0)	506 (75.5)	214 (80.8)	2.94 (.09)
Level of Education				
School completion or below	144 (15.4)	89 (13.3)	55 (20.8)	8.14 (.004)
TAFE or technical college	125 (13.4)	85 (12.7)	40 (15.1)	0.95 (.33)
University/College	666 (71.2)	496 (74.0)	170 (64.2)	9.04 (.003)
Aboriginal /Torres Strait Islander*	15 (1.6)	8 (1.2)	7 (2.6)	2.53 (.11)
Infant sex male	477 (51.0)	320 (47.8)	157 (59.2)	10.02 (.002)
Infant hospitalised	117 (12.5)	73 (10.9)	44 (16.6)	5.65 (.02)
Only child	558 (59.7)	402 (60.0)	156 (58.9)	0.10 (.75)

Note. χ^2= Chi squared statistic; t=independent t-test statistic. *3 participants did not respond.

3.4.2 BSID-III Scaled Scores and Frequency of NDD

BSID-III scaled scores for the total sample, as well as for the at risk and low risk groups are provided in Table 2. Of note, low risk infants obtained significantly higher scores on the cognitive, receptive language and gross motor domains of the BSID-III than at risk infants. Furthermore, significantly more at risk infants had NDD (scaled score <7) on the receptive language and gross motor domains compared to low risk infants. No between group differences in rates of detected NDD were observed on the cognitive domain.

Table 2. Mean BSID-III scaled scores, NDD frequency and associations by risk status.

	Total *n* = 935	**Low risk *n* = 670**		**At risk *n* = 265**			
BSID-III Domains	**M (*SD*)**	**M (*SD*)**	**NDD *n* (%)**	**M (*SD*)**	**NDD *n* (%)**	**t (*p*)**	**χ^2 (*p*)**
Cognition	11.40 (2.23)	11.58 (2.27)	6 (.90)	10.94 (2.30)	6 (2.26)	3.89 (<.001)	2.81 (.10)
Receptive language	10.12 (2.57)	10.24 (2.46)	35 (5.22)	9.81 (2.80)	32 (12.08)	2.33 (.02)	13.40 (<.001)
Expressive language	10.18 (1.96)	10.26 (1.96)	16 (2.39)	9.99 (1.95)	11 (4.15)	1.85 (.07)	2.10 (.15)
Fine motor	10.52 (2.50)	10.60 (2.49)	6 (.90)	10.31 (2.52)	3 (1.13)	1.62 (.11)	0.11 (.74)
Gross motor	9.02 (2.67)	9.14 (2.59)	59 (8.81)	8.70 (2.84)	41 (15.47)	2.30 (.02)	8.83 (.003)

Note. χ^2= Chi squared statistic; t=independent t-test statistic.

3.4.3 Linear Regression Output

As seen in Table 3, risk status significantly predicted BSID-III cognitive and gross motor scores, with at risk infants performing significantly worse than low risk infants, after controlling for: infant sex; hospitalisations since birth; and, maternal age at birth; linguistic background; employment status; educational attainment; socio-economic status and parity. The between group differences detected on the receptive language domain in the unadjusted model were no longer significant in the adjusted model.

Next, to examine the clinical significance of our results, scaled score group differences and the percentage of overall variance in performance explained by risk status, was investigated. After controlling for covariates, at risk infants scored, on average, 0.57 and 0.44 scaled score points lower than low risk infants on the cognitive and gross motor domains respectively, with risk status explaining 1.2 percent and 0.5 percent of variation in infant cognitive and gross motor scores, respectively (see Table 3).

Table 3. Linear regression comparing at risk and low risk infant scaled scores on the BSID-III, controlling for covariates.

BSID-III Domains	**F**	***p***	η_p^2	η^2
Cognitive	11.53	.001	.012	.043
Receptive language	1.88	.17	.002	.075
Expressive language	1.64	.20	.002	.044
Fine motor	1.35	.25	.001	.045
Gross motor	5.02	.03	.005	.025

Note. df = 1, 924

3.4.4 Associations Between the Seven Individual Risk Factors and BSID-III Scaled Scores

Examination of the unique contribution of each indirect risk factor to overall risk status was then assessed. Each factor was significantly associated with poorer outcomes on one or more BSID-III domain with the exception of prematurity and low Apgar scores. However, it should be noted that the proportion of premature infants in the current sample was well below the national average, likely due to the inclusion of only one infant from multiple births, and selection criteria excluding known major medical problems antenatally. Mean scale score differences ranged from 0.37 to 1.70 points. On domains where differences were detected, maternal substance use, multiple birth infants, low birth weight and small head circumference were associated with the strongest clinical differences, with mean between group differences of approximately one to two scale score points (see Table 4). Low Apgar scores were associated with higher performance on the fine motor domain compared to infants with average Apgar scores, with a mean scale score difference of 0.67 points. Repetition of the analysis with a cut-point for low Apgar score of <7 at five minutes only (where infants with low Apgar score at one minute but Apgar >6 at five minutes were reclassified as low risk) (de Moura et al., 2010), resulted in no significant association between risk status on any of the BSID-III domains. Finally, linear regression analyses were repeated on the full available cohort with risk status recoded (see Appendix C). Similar results were produced, although expressive language and fine motor approached significance.

Table 4. Comparison between individual risk factors and scaled scores on the BSID-III

	n	Cognitive		Receptive language		Expressive language		Fine motor		Gross motor	
Risk factors	**(n at risk)**	**t (*p*)**	**M $_{diff}$**	**t (*p*)**	**M $_{diff}$**	**t (*p*)**	**M $_{diff}$**	**t (*p*)**	**M $_{diff}$**	**t (*p*)**	**M $_{diff}$**
Maternal substance use	935 (34)	2.56 (.01)	1.02	2.11 (.04)	0.95	0.99 (.32)		1.02 (.31)		1.41 (.16)	
Multiple birth infant	935 (20)	1.63 (.12)		2.95 (.003)	1.70	2.04 (.04)	0.90	1.12 (.27)		-0.14 (.89)	
Preterm	935 (21)	1.28 (0.20)		0.38 (.70)		-1.49 (.14)		0.69 (.49)		1.52 (.13)	
Low birth weight	926 (34)	2.33 (.02)	0.94	0.56 (.57)		0.44 (.15)		3.00 (.005)	0.85	1.59 (.11)	
Low Apgar score	898 (78)	0.65 (.52)		-0.50 (.62)		-1.21 (.23)		-2.26 (.02)	0.67	-1.35 (.18)	
Small head circumference	890 (44)	2.97 (.003)	1.05	1.73 (.08)		1.82 (.07)		3.26 (.002)	0.97	2.20 (.03)	0.91
Special care or NICU	929 (156)	1.807 (.07)		1.60 (.11)		2.14 (.03)	0.37	1.85 (.07)		2.68 (.007)	0.63

Note. Comparing at risk and low risk infant performance on the BSID-III when risk status is categorised by individual risk factor; 'M $_{diff}$' = Mean scaled score difference, on average, between at risk and low risk groups

3.4.5 Factor Analysis of Risk Status Construct

All Kaiser-Meyer-Olkin (KMO) values for the individual items were ≥0.5 and the KMO measure was 0.699, indicating appropriateness for exploratory factor analysis (Yong & Pearce, 2013). The Bartlett's test of sphericity $\chi^2(21)= 719.37$, $P<0.001$ showed that items were related. Using an eigenvalue cut-off of 1.0, three factors were identified which explained a cumulative variance of 33.64 percent. The scree plot, however, indicated that one factor may provide the best fit. This was supported by initial eigenvalues for factors two and three approaching the cut-off limit of 1.0 (the eigenvalue value for factor two was 1.119 and for factor three was 1.040). As such, confirmatory factor analysis was conducted in order to determine the percent of variance explained when restricting the number of extracted factors to one. Confirmatory factor analysis when restricted to one extracted factor explained 22.89 percent of variance. Table 5 shows the factor loadings for the confirmatory factor analysis. Low birth weight and small head circumference loaded on indirect risk status; low Apgar scores and maternal substance use in pregnancy contributed less to the risk model.

Table 5. Confirmatory factor analysis restricting risk model to one extracted factor.

Individual risk factors	**Factor 1: Indirect risk status**
Low birth weight (<2.5kg)	0.816
Small head circumference (≤3rd percentile)	0.622
Preterm (<36 gestation)	0.509
Multiple birth infant	0.374
Admission to special care or neonatal intensive care unit (NICU) post birth	0.365
Mother recruited through substance use clinic	0.115
Low Apgar score (<7 at either 1 or 5 minutes)	0.064
Extracted Sum of Square Eigenvalue	1.602
% of variance	22.89

3.5 Discussion

The BSID-III is widely used in developmental follow-up clinics in Australia. However, there is scant evidence available on the ability of the BSID-III to detect differences in young infant performance based on indirect perinatal risk of NDD (i.e. perinatal risk factors that do not necessarily lead to detectable neurological signs, but are associated with increased risk of delay). This is of clinical import as the administration of the BSID-III requires considerable resources. Therefore, understanding whether the BSID-III can detect differences in performance between those at indirect perinatal risk and those at low risk will inform its utility in developmental follow-up clinics in Australia. To address this gap the current study used a prospective, longitudinal cohort of 935 infants aged 1-year to examine the extent to which the BSID-III could accurately detect differences in performance between infants at indirect perinatal risk of NDD compared to infants at low perinatal risk of NDD. This study specifically built on the available evidence base by including a heterogenous cohort of at risk infants to increase the generalisability of results, and by extending the age range to include a younger cohort of infants (1-year-old).

Results highlighted a number of important differences between the at risk and low risk groups. On average, at risk infants were more likely to be male, to have been hospitalised overnight since birth, and to be born to mothers of lower educational attainment, not undertaking paid employment and not married or in a defacto relationship, compared to 'low risk' infants. This is consistent with previous research examining individual antenatal/perinatal risk factors, which has demonstrated male sex and low parental socio-economic status are important markers for increased likelihood of indirect perinatal risk (Aylward, 2003; de Moura et al., 2010; Greene et al., 2012; Velikos et al., 2015).

Examination of the BSID-III scores as a function of risk status indicated that the gross motor and cognitive domains had the greatest utility differentiating between 1-year-old

infants at indirect perinatal risk and those at low risk. Infants at indirect perinatal risk scored significantly lower than low risk infants on these domains. Despite this, the clinical significance of detected differences was small. Considering one standard deviation on the US BSID-III normative data is 3 scale score points, the detected group differences of 0.57 and 0.44 in the current study, are unlikely to be clinically significant. Moreover, the proportion of variance accounted for by risk status was low. No differences were found on the language and fine motor domains following covariate adjustment.

Results indicated that a higher proportion of at risk infants scored below seven on the receptive language and gross motor domains, compared to low risk infants; indicating increased proportions of developmental delay based on risk status across these domains. However, the average gross motor score across both groups was well below the US normative mean, suggesting that interpretation of scaled scores of less than 7 as indicative of delay may not be appropriate.

There are a number of possible explanations for the limited ability for the BSID-III to clinically differentiate between infants at indirect perinatal risk and those at low perinatal risk. First, it is possible that the indirect risk factors included in this study do not result in delays to the skills assessed by the BSID-III. However, considering the wealth of research available demonstrating associations between these risk factors and delay as measured by the BSID-III (Baron, Ahronovich, Erickson, Larson, & Litman, 2009a; Foster-Cohen, Friesen, Champion, & Woodward, 2010; Kerstjens et al., 2011), this is unlikely. Further, factor analysis revealed that all risk factors included in this study, with the exception of low Apgar scores, were associated with lower scores on one or more BSID-III domains, and/or significantly loaded onto a common ‘indirect risk status’ factor. Therefore, this is unlikely to account for the obtained results.

Second, the heterogeneity of risk factors in the cohort may have 'diluted' the clinical significance of results. If for example a single risk factor is associated with delay to language abilities but not motor abilities, and another risk factor is associated with delay to motor abilities but not language abilities, examination of these risk factors together in an "indirect risk status group" may result in weaker between group differences than would be expected with individual examination of risk factors. Results from analyses of BSID-III scores by individual risk factor provide some support for this explanation. For example, infants with maternal substance use, low birth weight and small head circumference each scored approximately one scaled score point lower than low risk infants on the cognitive domain, and infants with small head circumference and those transferred to NICU or SCU scored between 0.5 to one scaled score point lower than low risk infants on the gross motor domain. While this may have attenuated the strength of the results, it does not fully account for the lack of differences based on indirect risk status on the language and fine motor domains.

There are two further possible explanations for the lack of between group differences detected on the language and fine motor domains. First, the risk factors included in this study may only impact higher order skills developed later in childhood. Results of a recent systematic review and meta-analysis reported that preterm infants performed consistently worse on higher order language skills, but not simple language skills, and that delays become more pronounced with increasing age in childhood (van Noort-van der Spek, Franken, & Weisglas-Kuperus, 2012). Similarly, a study comparing 95 late preterm infants with term infants found no significant difference in motor dexterity at 3 years of age (Baron et al., 2009b), whereas a study by the same author found that preterm/extremely low birth weight infants without intraventricular haemorrhage at 6-years of age performed poorer on motor dexterity tasks than non-preterm children (Baron et al., 2009a). As such, it might be the case that specific language and fine motor skills impaired by indirect risk factors only develop

later in childhood and thus cannot be detected through early assessment. Second, the BSID-III may not sensitive enough to detect differences in infant language and fine motor performance at 1-year of age. While significant group differences on the language and fine motor domains were detected in the current study based on maternal linguistic background and plurality (Appendix A), the BSID-III may not be sensitive to delay. Optimal assessment tools should be structured in such a way that they are sensitive to true markers of delay and relatively robust against other environmental factors. As such, the BSID-III items for these domains may not appropriately assess true markers of early delay.

The distinction between the two aforementioned issues is of clinical import. A growing body of evidence suggests that less severe risk factors, such as those included in the current study, result in high-prevalence-low-severity delay (Foster-Cohen et al., 2010). Here, mild yet clinically significant delays including specific delays, mild intellectual impairment and behavioural difficulties, occur frequently following risk exposure, but are often only detected at school age. Some researchers have suggested a need to develop more accurate assessment tools to detect these delays earlier (Foster-Cohen et al., 2010). However, if the early signs of these delays do not emerge in infancy, assessment at this time is unwarranted.

3.5.1 Strengths and Limitations

A number of limitations should be noted. First, mothers were relatively high functioning, and therefore may have had higher intellectual abilities relative to the general Australian population; traits which have known heritability and are considered protective factors for offspring NDD. As such, the present results may represent a conservative estimate of the ability of the BSID-III to differentiate between at risk and low risk infants. Second, infants at greatest risk of NDD resulting from known antenatal factors such as hydrocephalus and intracranial haemorrhage were excluded from this study. While this approach was

intended, results are only applicable to the risk factors included in our definition of indirect risk. Similarly, the inclusion criteria of "no known major medical complications for the mother or fetus (e.g., chromosomal abnormalities)" likely reduced the number of very medically high-risk infants included in the study. This is supported by the low rate of prematurity in the sample compared to national averages (i.e. approximately 2 percent in the current sample compared with approximately 8 percent nationally). As such, results again represent a conservative estimate of the ability of the BSID-III to differentiate based on risk status. Third, no brain imaging data were available for the cohort and, therefore, it is not possible to rule out cases of neurological impairment at birth. Fourth, t-test and factor analysis results between individual risk factors and BSID-III scores suggested that, although evidence-based (Odd et al., 2008), Apgar categorisation of ≤6 as a marker of indirect risk may have been too stringent. Nonetheless, the results of the fully adjusted regression models remained unchanged in post-hoc analyses when Apgar risk was recoded to include only infants with 5-minute scores of ≤6. Finally, our indirect risk status factor only explained 22.89 percent of overall risk. The addition of other antenatal and perinatal factors to the model, such as maternal mental health, antenatal bonding (Rossen et al., 2016; Rossen, et al., 2017), and complications in labour (i.e., meconium, foetal distress, breech birth), may be of benefit in future research.

3.5.2 Conclusion and Implications

This study suggests that the BSID-III has limited clinical utility to detect differences in infant performance at 1-year *based on indirect risk status*, in a large community cohort. Whilst infants at indirect risk of NDD performed significantly lower on the cognitive and gross motor domains of the BSID-III, and to some degree, on the receptive language domain, when compared to low risk infants, mean group differences were minimal, suggesting limited

clinical significance. Moreover, no differences were detected on the expressive language and fine motor domains. It is possible that our measure of ‘at risk’ status did not sufficiently capture infant risk behaviour, given the cohort was relatively well adjusted overall (McPhie et al., 2017). Alternatively, the BSID-III may not be a sufficiently sensitive tool to detect meaningful differences in infant development between low and at risk infants (i.e., it may be better suited to detect more severe risk and NDD, for example). A final alternative is that the developmental sequalae of infant risk exposures examined here are more readily detected in older childhood, which would warrant further prospective follow-up of this cohort.

3.6 References

Albers, C. A. & Grieve, A. J. (2007). Test Review: Bayley, N. (2006). Bayley Scales of Infant and Toddler Development- Third Edition. San Antonio, TX: Harcourt Assessment. *Journal of Psychoeducational Assessment, 25*, 180-190. doi: 10.1177/0734282906297199.

American Academy of Pediatrics. (2004). Follow-up care of high-risk infants. *Pediatrics, 114* (Supplement 5), 1377-1397. doi: 10.1542/peds.2004-0866.

Anderson, P. J., & Burnett, A. (2017). Assessing developmental delay in early childhood—concerns with the Bayley-III scales. *The Clinical Neuropsychologist, 31*(2), 371-381. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1080/13854046.2016.1216518

Anderson, P. J., De Luca, C. R., Hutchinson, E., Roberts, G., & Doyle, L. W. (2010). Underestimation of developmental delay by the new Bayley-III Scale. *Archives of Pediatric and Adolescent Medicine, 164(4)*: 352-356. doi:10.1001/archpediatrics.2010.20.

Australian Bureau of Statistics (ABS). *2033.0.55.001 - Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2011.* Retrieved from the Australian Bureau of Statistics: http://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001.

Australian Bureau of Statistics (ABS). 6227.0 - *Education and Work, Australia, May 2015.* Retrieved from the Australian Bureau of Statistics: http://www.abs.gov.au/ausstats/abs@.nsf/mf/6227.0/.

Aylward, G. P. (2003). Neonatology, prematurity, NICU, and developmental issues. *Handbook of Pediatric Psychology, 3*, 253-268. Retrieved from: http://users.phhp.ufl.edu/jhj/CH15AND20.pdf.

Bandstra, E. S., Morrow, C. E., Mansoor, E., & Accornero, V. H. (2010). Prenatal drug exposure: infant and toddler outcomes. *Journal of Addictive Diseases*, *29*(2), 245-258. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1080/10550881003684871.

Baron, I. S., Ahronovich, M. D., Erickson, K., Larson, J. C. G., & Litman, F. R. (2009a). Age-appropriate early school age neurobehavioral outcomes of extremely preterm birth without severe intraventricular hemorrhage: A single center experience. *Early Human Development*, *85*(3), 191-196. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1016/j.earlhumdev.2008.09.411.

Baron, I. S., Erickson, K., Ahronovich, M. D., Coulehan, K., Baker, R., & Litman, F. R. (2009b). Visuospatial and verbal fluency relative deficits in 'complicated' late-preterm preschool children. *Early Human Development*, *85*(12), 751-754. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1016/j.earlhumdev.2009.10.002.

Bayley, N. (2006a) *The Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III).* 3rd ed. San Antonio, Texas: Pearson Psychcorp.

Bayley, N. (2006b). *Bayley Scales of Infant and Toddler Development, Third Edition Technical Manual.* San Antonio, TX: Pearson PsychCorp.

Blauw-Hospers, C. H., & Hadders-Algra, M. (2005). A systematic review of the effects of early intervention on motor development. *Developmental Medicine & Child Neurology*, *47*(6), 421-432. doi: 10.1111/j.1469-8749.2005.tb01165.x.

Bos, A. F. & Roze, E. (2011). Neurodevelopmental outcome in preterm infants. *Developmental Medicine and Child Neurology; 53*(s4), 35-39. doi: 10.1111/j.1469-8749.2011.04062.x.

Chung, S. 2009. Neurology News: Small head size. *Neurology Now*, *5*(5), 15.

De Jesus, L. C., Pappas, A., Shankaran, S., Li, L., Das, A., Bell, E. F., et al. (2013). Outcomes of small for gestational age infants born at< 27 weeks' gestation. *The Journal of Pediatrics*, *163*(1), 55-60. doi: http://dx.doi.org/10.1016/j.jpeds.2012.12.097.

de Moura, D. R., Costa, J. C., Santos, I. S., Barros, A. J., Matijasevich, A., Halpern, R. et al. (2010). Risk factors for suspected developmental delay at age 2 years in a Brazilian birth cohort. *Paediatric and Perinatal Epidemiology*, *24*(3), 211-221. doi: 10.1111/j.1365-3016.2010.01115.x.

de Vries, L. S., & M. J. Jongmans (2010). Long-term outcome after neonatal hypoxic-ischaemic encephalopathy. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *95*(3), F220-F224. doi: http://dx.doi.org/10.1136/adc.2008.148205

Dixon, G., Badawi, N., Kurinczuk, J. L., Keogh, J. M., Silburn , S. R., Zubrick, S. R., & Stanley, F. J. (2002). Early developmental outcomes after newborn encephalopathy. *Pediatrics*, *109*(1), 26-33. doi: 10.1542/peds.109.1.26.

Fallah, R., Islami, Z., & Mosavian, T. (2011). Developmental status of NICU admited low birth weight preterm neonates at 6 and 12 months of age using Ages and Stages Questionaire. *Iranian Journal of Child Neurology*, *5*(1), 21-28. Retrieved from: http://journals.sbmu.ac.ir/ijcn/article/view/2120.

Foster-Cohen, S. H., Friesen, M. D., Champion, P. R., & Woodward, L. J. (2010). High prevalence/low severity language delay in preschool children born very preterm. *Journal of Developmental & Behavioral Pediatrics*, *31*(8), 658-667. doi: 10.1097/DBP.0b013e3181e5ab7e.

Futagi, Y., Suzuki, Y., Toribe, Y., Morimoto, K. (2002). Neurodevelopmental outcome in children with fetal hydrocephalus. *Pediatric Neurology*, *27*(2), 111-116. doi: http://dx.doi.org/10.1016/S0887-8994(02)00395-8.

Greene, M. M., Patra, K., Nelson, M. N., & Silvestri, J. M. (2012). Evaluating preterm infants with the Bayley-III: patterns and correlates of development. *Research in Developmental Disabilities*, *33*(6), 1948-1956. doi: http://dx.doi.org/10.1016/j.ridd.2012.05.024.

Gonzalez, F. F., & Miller, S. P. (2006). Does perinatal asphyxia impair cognitive function without cerebral palsy? *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *91*(6), F454-F459. doi: 10.1136/adc.2005.092445.

Howard, K., Roberts, G., Lim, J., Lee, K. J., Barre, N., Treyvaud, K., ... & Anderson, P. J. (2011). Biological and environmental factors as predictors of language skills in very preterm children at 5 years of age. *Journal of Developmental & Behavioral Pediatrics*, *32*(3), 239-249. doi: 10.1097/DBP.0b013e31820b7882

Huizink, A. C. (2014). Prenatal cannabis exposure and infant outcomes: overview of studies. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *52*, 45-52. doi: http://dx.doi.org/10.1016/j.pnpbp.2013.09.014.

Hutchinson, D., Wilson, J., Allsop, S., Elliott, E., Najman, J., Burns, L., et al. (2017) Cohort profile: The Triple B Pregnancy Cohort Study: A longitudinal study of the relationship between alcohol, tobacco and other substance use during pregnancy and the health and wellbeing of Australian children and families. *International Journal of Epidemiology,* accepted manuscript.

IBM Corp. (2014). *IBM SPSS Statistics for Windows* (*Version 23.0*). Armonk, NY: IBM Corp.

Jarjour, I. T. (2015). Neurodevelopmental outcome after extreme prematurity: a review of the literature. *Pediatric Neurology*, *52*(2), 143-152. doi: http://dx.doi.org/10.1016/j.pediatrneurol.2014.10.027.

Johnson, S., Moore, T., & Marlow, N. (2014). Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? *Pediatric Research, 75*(5), 670-674. doi:10.1038/pr.2014.10.

Kerstjens, J. M., de Winter, A. F., Bocca-Tjeertes, I. F., ten Vergert, E. M., Reijneveld, S. A., & Bos, A. F. (2011). Developmental delay in moderately preterm-born children at school entry. *The Journal of Pediatrics, 159*(1), 92-98. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1016/j.jpeds.2010.12.041

Lain, S. J., Roberts, C. L., Bowen, J. R., & Nassar, N. (2014). Trends in New South Wales infant hospital readmission rates in the first year of life: a population-based study. *The Medical Journal of Australia, 201*(1), 40-43.

Landers, R. N. (2015). Computing intraclass correlations (ICC) as estimates of inter rater reliability in SPSS. *The Winnowner 2*. doi: 10.15200/winn.143518.81744.

Logan, B. A., Brown, M. S., & Hayes, M. J. (2013). Neonatal abstinence syndrome: treatment and pediatric outcomes. *Clinical Obstetrics and Gynecology, 56*(1), 186. doi: 10.1097/GRF.0b013e31827feea4.

Luttikhuizen dos Santos, E. S. L., de Kieviet, J. F., Königs, M., van Elburg, R. M., & Oosterlaan, J. (2013). Predictive value of the Bayley scales of infant development on development of very preterm/very low birth weight children: a meta-analysis. *Early Human Development, 89*(7), 487-496. dio:10.1016/j.earlhumdev.2013.03.008.

McGowan, J. E., Alderdice, F. A., Holmes, V. A., & Johnston, L. (2011). Early childhood development of late-preterm infants: a systematic review. *Pediatrics*, peds-2010. doi: 10.1542/peds.2010-2257.

McPhie, S., Skouteris, H., Mattick, R. P., Wilson, J., Honan, I., Allsop, S., et al. (2017). Weight in the First Year of Life: Associations with Maternal Pre-pregnancy Body Mass

Index and Gestational Weight Gain—Findings from a Longitudinal Pregnancy Cohort. *American Journal of Perinatology*, *34*(08), 774-779. doi: 10.1055/s-0036-1597992

Odd, D. E., Rasmussen, F., Gunnell, D., Lewis, G., & Whitelaw, A. (2008). A cohort study of low Apgar scores and cognitive outcomes. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *93*(2), F115-F120. doi: 10.1136/adc.2007.123745.

Olusanya, B. O. (2011). Perinatal outcomes of multiple births in southwest Nigeria. *Journal of Health, Population and Nutrition*, 29(6), 639-647. PMC3259727.

Patra, K., Wilson-Costello, D., Taylor, H. G., Mercuri-Minich, N., & Hack, M. (2006). Grades I-II intraventricular hemorrhage in extremely low birth weight infants: effects on neurodevelopment. *The Journal of Pediatrics*, *149*(2), 169-173. doi: http://dx.doi.org/10.1016/j.jpeds.2006.04.002.

Peterson, J., Taylor, H. G., Minich, N., Klein, N., & Hack, M. (2006). Subnormal head circumference in very low birth weight children: neonatal correlates and school-age consequences. *Early Human Development*, *82*(5), 325-334. doi: https://doi.org/10.1016/j.earlhumdev.2005.09.014

Rossen, L., Hutchinson, D., Wilson, J., Burns, L., Allsop, S., Elliott, E. J., et al. (2017). Maternal Bonding through Pregnancy and Postnatal: Findings from an Australian Longitudinal Study. *American Journal of Perinatology*, 34(08), 808-817.

Rossen, L., Hutchinson, D., Wilson, J., Burns, L., Olsson, C. A., Allsop, S., et al. (2016). Predictors of postnatal mother-infant bonding: the role of antenatal bonding, maternal substance use and mental health. *Archives of Women's Mental Health*, *19*(4), 609-622. doi: 10.1007/s00737-016-0602-z.

Shenkin, S. D., Starr, J. M., & Deary, I. J. (2004). Birth weight and cognitive ability in childhood: a systematic review. *Psychological Bulletin, 130*(6), 989. doi: http://dx.doi.org/10.1037/0033-2909.130.6.989.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420. doi: http://dx.doi.org/10.1037/0033-2909.86.2.420.

Skiöld, B., Vollmer, B., Böhm, B., Hallberg, B., Horsch, S., Mosskin, M., et al. (2012). Neonatal magnetic resonance imaging and outcome at age 30 months in extremely preterm infants. *The Journal of Pediatrics, 160*(4), 559-566. doi: http://dx.doi.org/10.1016/j.jpeds.2011.09.053.

Soria-Pastor, S., Padilla, N., Zubiaurre-Elorza, L., Ibarretxe-Bilbao, N., Botet, F., Costas-Moragas, C., et al. (2009). Decreased regional brain volume and cognitive impairment in preterm children at low risk. *Pediatrics, 124*(6), e1161-e1170. doi: 10.1542/peds.2009-0244.

SPSS Web Books Regression With SPSS Chapter 2 – Regression Diagnostics. UCLA: Statistical Consulting Group. Retrieved from: http://stats.idre.ucla.edu/spss/webbooks/reg/chapter2/spss-webbooksregressionwith-spsschapter-2-regression-diagnostics/#.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics (6th Edition)*. Boston; Pearson.

Tay, R. Y., Wilson, J., McCormack, C., Allsop, S., Najman, J. M., Burns, L., et al. (2017). Alcohol consumption by breastfeeding mothers: Frequency, correlates and infant outcomes. Drug and Alcohol Review. doi: 10.1111/dar.12473

Van Handel, M., Swaab, H., De Vries, L. S., & Jongmans, M. J. (2007). Long-term cognitive and behavioral consequences of neonatal encephalopathy following perinatal asphyxia: a review. *European Journal of Pediatrics*, *166*(7), 645-654. doi: 10.1007/s00431-007-0437-8.

van Noort-van der Spek, I. L. Franken, M. C. J., & Weisglas-Kuperus, N. (2012). Language functions in preterm-born children: a systematic review and meta- analysis. *Pediatrics*, *129*(4), 745-754. doi: 10.1542/peds.2011-1728

Velikos, K., Soubasi, V., Michalettou, I., Sarafidis, K., Nakas, C., Papadopoulou, V., et al. (2015). Bayley-III scales at 12 months of corrected age in preterm infants: Patterns of developmental performance and correlations to environmental and biological influences. *Research in Developmental Disabilities*, *45*, 110-119. doi: http://dx.doi.org/10.1016/j.ridd.2015.07.014.

Wadhawan, R., Oh, W., Vohr, B. R., Wrage, L., Das, A., Bell, E. F., ... & Higgins, R. D. (2011). Neurodevelopmental outcomes of triplets or higher-order extremely low birth weight infants. *Pediatrics*, *127*(3), e654-e660. doi: 10.1542/peds.2010-2646.

Walker, K., Badawi, N., Halliday, R., & Laing, S. (2010). Brief Report: Performance of Australian children at one year of age on the Bayley scales of infant and toddler development (Version III). *The Australian Educational and Developmental Psychologist*, *27*(01), 54-58. doi: https://doi.org/10.1375/aedp.27.1.54.

Walker, K., Holland, A. J., Halliday, R., & Badawi, N. (2012). Which high-risk infants should we follow-up and how should we do it?. *Journal of Paediatrics and Child Health*, *48*(9), 789-793. doi: 10.1111/j.1440-1754.2012.02540.x

Westrupp, E. M., Mensah, F. K., Giallo, R., Cooklin, A., & Nicholson, J. M. (2012). Mental health in low-to-moderate risk preterm, low birth weight, and small for gestational age

children at 4 to 5 years: the role of early maternal parenting. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(3), 313-323. doi: http://dx.doi.org/10.1016/j.jaac.2011.12.004.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79-94. Retrieved from: http://www.tqmp.org/Content/vol09-2/p079/p079.pdf.

3.7 Appendix

Appendix A: Full linear regression model results for each BSID-III domain

Regression models	df	F	*p*	η_p^2
COGNITIVE				
Indirect risk status	1	11.53	.001	.012
Employment status	1	0.08	.77	.000
Linguistic background	1	11.65	.001	.012
SEIFA percentile	1	3.90	.05	.004
Maternal age	1	2.13	.15	.002
Education	2	2.17	.11	.005
Infant sex	1	1.45	.23	.002
Infant hospitalised	1	3.77	.05	.004
Only child	1	2.27	.13	.002
Full model				.043
RECEPTIVE LANGUAGE				
Indirect risk status	1	1.88	.17	.002
Employment status	1	0.30	.58	.000
Linguistic background	1	6.29	.01	.007
SEIFA percentile	1	3.47	.06	.004
Maternal age	1	0.08	.77	.000
Education	2	1.33	.26	.003
Infant sex	1	36.83	<.001	.038
Infant hospitalised	1	0.86	.36	.001
Only child	1	11.01	.001	.012
Full model				.075
EXPRESSIVE LANGUAGE				
Indirect risk status	1	1.64	.20	.002
Employment status	1	0.29	.59	.000
Linguistic background	1	11.66	.001	.012
SEIFA percentile	1	3.83	.05	.004
Maternal age	1	0.57	.45	.001
Education	2	2.52	.08	.005
Infant sex	1	16.71	<.001	.018
Infant hospitalised	1	0.67	.41	.001
Only child	1	0.92	.34	.001
Full model				.044

Regression models	df	F	*p*	η_P^2
	FINE MOTOR			
Indirect risk status	1	1.35	.25	.001
Employment status	1	0.11	.74	.000
Linguistic background	1	6.51	.01	.007
SEIFA percentile	1	0.94	.33	.001
Maternal age	1	1.16	.28	.001
Education	2	3.29	.04	.007
Infant sex	1	8.39	.004	.009
Infant hospitalised	1	5.78	.02	.006
Only child	1	12.14	.001	.013
Full model				.045
	GROSS MOTOR			
Indirect risk status	1	5.02	.03	.005
Employment status	1	0.14	.71	.000
Linguistic background	1	4.97	.03	.005
SEIFA percentile	1	0.18	.67	.000
Maternal age	1	4.88	.03	.005
Education	2	0.37	.67	.001
Infant sex	1	0.32	.57	.000
Infant hospitalised	1	7.62	.006	.008
Only child	1	0.01	.91	.000
Full model				.025

Appendix B: Linear regression results comparing at risk and low risk infant scaled scores on the BSID-III, controlling for covariates.

BSID-III Domains	***n***	**F**	**df**	***p***	η_p^2	η^2
Cognitive	933	11.53	1, 922	.001	.012	.047
Receptive language	933	2.40	1, 922	.12	.003	.075
Expressive language	922	1.09	1, 911	.30	.001	.055
Fine motor*	930	1.49	1, 919	.22	.002	.050
Gross motor	927	7.83	1, 916	.005	.008	.033

Note. Outliers removed; *Square root transformed data

Appendix C: Linear regression results comparing at risk and low risk infant scaled scores on the BSID-III, controlling for covariates

BSID-III Domains	**F**	***p***	η_p^2	η^2
Cognitive	9.27	.002	.010	.040
Receptive language	2.23	.14	.002	.078
Expressive language	3.38	.07	.004	.048
Fine motor	3.58	.06	.004	.047
Gross motor	7.13	.008	.008	.027

Note. Risk status recoded for Apgar <7 at five minutes post birth only. Total number of infants included in analyses was 930: with 'at risk' infants n= 225; 'low risk' infants n=705. Risk status was undeterminable following recoding of Agpar scores for 5 infants due to missing data on one or more risk factor (see 'participants' section for further details). df = 1, 919.

Study 3

A comparison between Australian and United States normative data at 3-years of age on the Bayley Scales of Infant and Toddler Development-3rd Edition (BSID-III).

Ingrid Honan*, BPsych (Hons), Delyse Hutchinson, PhD, Melanie Porter, PhD, Steve Allsop, PhD, Jake Najman, PhD, Elizabeth J. Elliott, MD, Lucinda Burns, PhD, Craig A. Olsson, PhD, Richard P. Mattick, PhD

Paper has been prepared in accordance with authorship guidelines for:

Australia and New Zealand Journal of Psychiatry

Paper submitted for publication (July 2017) – See Appendix F

4.1 Abstract

Objective: In the toddlerhood to preschool period behaviour becomes more purposeful, individual aptitude begins to be expressed and developmental delays can often first be formally diagnosed. As such, it is a common clinical time-point for assessment and diagnosis. The Bayley Scales of Infant and Toddler Development-3rd Edition (BSID-III) is one of the most commonly used assessment scales in both Australia and internationally for children aged 0- to 42-months with suspected developmental difficulties. Yet no Australian normative data are available. Clinicians currently rely on US normative data, which may be inaccurate in an Australian context. The aim of the current study was to compare a representative cohort of Australian toddler performance at 3-years of age to the US BSID-III normative sample and investigate potential causes for detected differences.

Method: 119 mothers and their term, singleton offspring were recruited through antenatal clinics in New South Wales (NSW), Australia. Women completed structured interviews during pregnancy, at 8-weeks, 1-year and 3-years postpartum. The BSID-III was completed by children at 3-years.

Results: Demographic data comparing the current cohort, Australian population data and US population data are presented. As the current cohort was somewhat advantaged in socio-economic terms, raked weights were applied based on Australian population data to improve cohort representativeness. Once adjusted, Australian toddlers scored significantly higher than the US BSID-III normative sample on the receptive language (effect size, 0.30 [M_{diff} 95% CI, 0.39 – 1.21]), expressive language (effect size, 0.38 [M_{diff} 95% CI, 0.59 – 1.41]), fine motor (effect size, 0.60 [M_{diff} 95% CI, 1.18 – 1.97]) and gross motor domains (effect size, 0.28 [M_{diff} 95% CI, 0.33 – 1.21]), and significantly lower than the US normative sample on the cognitive domain (effect size, 0.31 [M_{diff} 95% CI, -0.97 – -0.49]). Moderate to large

effect sizes were obtained. Australian and US populations differed by ethnicity and educational attainment, possibly accounting for at least some of the detected differences. **Conclusions:** Australian performance in the current cohort significantly differed from the US normative sample at 3-years of age. Our results suggest that utilisation of US BSID-III normative data in an Australian population may be sub-optimal and contribute to diagnostic inaccuracy. The development of full age range, representative Australian normative data for the BSID-III is warranted.

4.2 Introduction

As children develop in late toddlerhood and the early preschool years, individual aptitude becomes more clearly expressed and measurable (Bee et al., 1982; Brownell and Kopp, 2010; Charman et al., 2005). More specifically, behaviour has been shown to become more purposeful and individual variability in aptitude becomes more stable. These aptitudes are considered precursors to intelligence, a construct which remains relatively stable across the lifespan (Deary et al., 2004). Consistent with this view, associations have been reported between language acquisition (Marchman and Fernald, 2008; Silva et al., 1987), motor abilities (Piek et al., 2008), executive functions, attention, working memory and the ability to process information (Bull et al., 2008; Campbell, 1995; Friedman et al., 2011; Rose and Feldman, 1995) in the toddlerhood/preschooler years, and later cognition, intellectual abilities and outcomes, such as academic achievement in early and middle childhood (Duncan et al., 2007; La Paro and Pianta, 2000).

As behaviour becomes more purposeful and individual aptitude stabilises, delays in developmental milestone attainment are often more readily detected. During the toddlerhood/preschool phase, neurodevelopmental disorders such as Autism Spectrum Disorders (ASD) (Chakrabarti and Fombonne, 2001; Charman et al., 2005; Glasson et al., 2004; Kleinman et al., 2008), communication/language disorders (Silva et al., 1987), motor delays (Lloyd et al., 2013) and global developmental delay (Riou et al., 2009; Shevell et al., 2005) are often first formally diagnosed. Accurate identification of delays at this age is fundamental for diagnosis and to gain access to appropriate early intervention and funding support (Campbell et al., 2001).

Standardised neuropsychological assessment tools are developed to assist clinicians to objectively and accurately identify deviations from typical development. Normative data, that is, cohorts of individuals representative of the broader population, are established to provide

clinicians with a reference group with which to compare individual performance, and to determine whether a child is performing below, consistent with, or above, their same aged peers. The Bayley Scales of Infants and Toddler Development III (BSID-III) is one such neuropsychological assessment tool (Bayley, 2006). The BSID-III objectively assesses cognitive, language and motor abilities of children aged 0- to 42-months of age (Bayley, 2006a). The BSID-III is widely used in both clinical practice and research in Australia and internationally (Albers and Grieve, 2007; Walker et al., 2010). However, there are currently no Australian normative data available for the BSID-III in toddlerhood. Australian clinicians currently compare child performance to the US normative sample.

Use of US normative data for an Australian population may not be appropriate (Chinta et al., 2014; Walker et al., 2010; *Study 1* of thesis). Normative data by nature are designed to reflect the population from which it is drawn. The BSID-III normative data are reportedly representative of the US population with regard to child sex, parental education and race/ethnicity, and the developers also endeavoured to obtain a representative geographical spread (Bayley, 2006b). Although many similarities exist between Australia and the US, significant population differences are evident such as: ethnicity; education systems; social systems; access to healthcare; and, linguistic background. These have all been demonstrated to be associated with development and/or achievement (Anderson et al., 2003; Hess et al., 1984; Greene et al., 2013; Sameroff et al., 1987; Sameroff et al., 1993).

A growing body of evidence suggests Australian infant performance on the BSID-III differs from the US normative sample. For example, Anderson et al. (2010), demonstrated that the BSID-III under-estimated developmental delay in a sample of 211 2-year-old children of extremely low birth weight/extreme prematurity. Moreover, the control sample, on average, performed significantly higher than the US normative sample on all domains (Anderson, et al., 2010). Furthermore, Walker et al. (2010) in a study of 211 Australian 1-

year old infants reported higher performance on the cognitive and receptive language domains, and lower performance on the gross motor and expressive language domains when compared to the US normative sample. Similarly, a study of 156 healthy, full-term, 3-year-old Australian children reported significantly higher performance on all objectively assessed domains of the BSID-III when compared to US normative data, with the exception of the gross motor domain, where no difference was detected (Chinta et al., 2014). These studies provide strong support for the need for further investigation into the appropriateness of using US normative data in Australia.

A further criticism of the extant literature is that there is a lack of attention to sampling related cohort effects, which may in part explain some of the differences in BSID-III scores among studies. It is noted that whilst most studies do note population differences such as ethnicity as important, many only provide demographic data on child sex and child health related outcomes, with few studies providing a comprehensive summary of maternal demographic characteristics (Chinta et al., 2014). A recent study by the current authors (Study 1 of this thesis) examining 998 1-year-old Australian infants is the only study to the author's knowledge that describes and matches detailed maternal socio-demographic data to Australian population data from the Australia Bureau of Statistics (ABS) in an attempt to improve cohort representativeness. It is only through collecting maternal demographic data, comparing cohort characteristics to population demographics, and taking steps to improve representativeness, that it is possible to make inferences that detected differences in BSID-III scores obtained, reflect true differences in child performance across countries.

Taken together, the aims of the current study were to: (1) describe the demographic characteristics of a cohort of 3-year-old Australian children (n=119; mean age=36 months) with complete BSID-III data and draw comparisons to Australian and US population data; (2) derive a cohort that is closer in representativeness to the Australian population on key

demographic characteristics (that is: maternal age at birth; maternal level of education; maternal employment status; number of biological children; maternal region of birth; Aboriginal or Torres Strait Islander (ATSI) status; and, child sex); and, (3) compare Australian child performance on the BSID-III at 3-years of age to the US normative sample. By improving sample representativeness to the Australian population, it will be possible to draw inferences that detected differences reflect true population differences.

4.3 Method

4.3.1 Participants

Data were drawn from the Triple B Pregnancy Cohort Study; a prospective longitudinal study of pregnant women, partners and their infant offspring (post-birth). Full cohort eligibility criteria, retention rates and cohort characteristics are described elsewhere (Hutchinson et al., 2017). In addition to full cohort selection criteria, eligibility criteria for the current 3-year subsample included: singleton children; gestational age of ≥36 weeks at birth; recruitment through general public antenatal clinics or birth centres in NSW; BSID-III completion age between 33- and 39-months; completion of all previous follow-up time points; and, no significant health conditions/concerns (such as Autism Spectrum Disorder) indicating atypical development. Due to feasibility, a subsample of 125 participating mothers and children were targeted. Sample size estimates were based on power calculations from the US BSID-III normative sample of 100 children per age range, plus an additional 25 percent to ensure adequate power after accounting for incomplete questionnaires, attrition on interview/assessment rescheduling and post data collection exclusion.

Of the initial 1,305 families in NSW, 166 mothers and children meeting eligibility criteria were invited to participate in the 3-year subsample follow-up. Of these, 125 agreed to participate. Six children were excluded following completion of the 3-year follow-up,

resulting in a final cohort of 119 for the current study (see Figure 1). This sample size was selected as it is commensurate with the US BSID-III normative sample size of 100 children per age band (Bayley, 2006b). Note: some multiple birth children; preterm children; and, children born to mothers recruited through substance use/high-risk clinics also completed the 3-year assessment as part of the broader Triple B study, but were not reported here as they were not eligible for the current study. Informed consent was obtained from mothers. Ethics approval was granted by University and NSW Health Human Research Ethics Committees (lead approval numbers: HREC/11/RPAH/153 and Protocol No X11-0111).

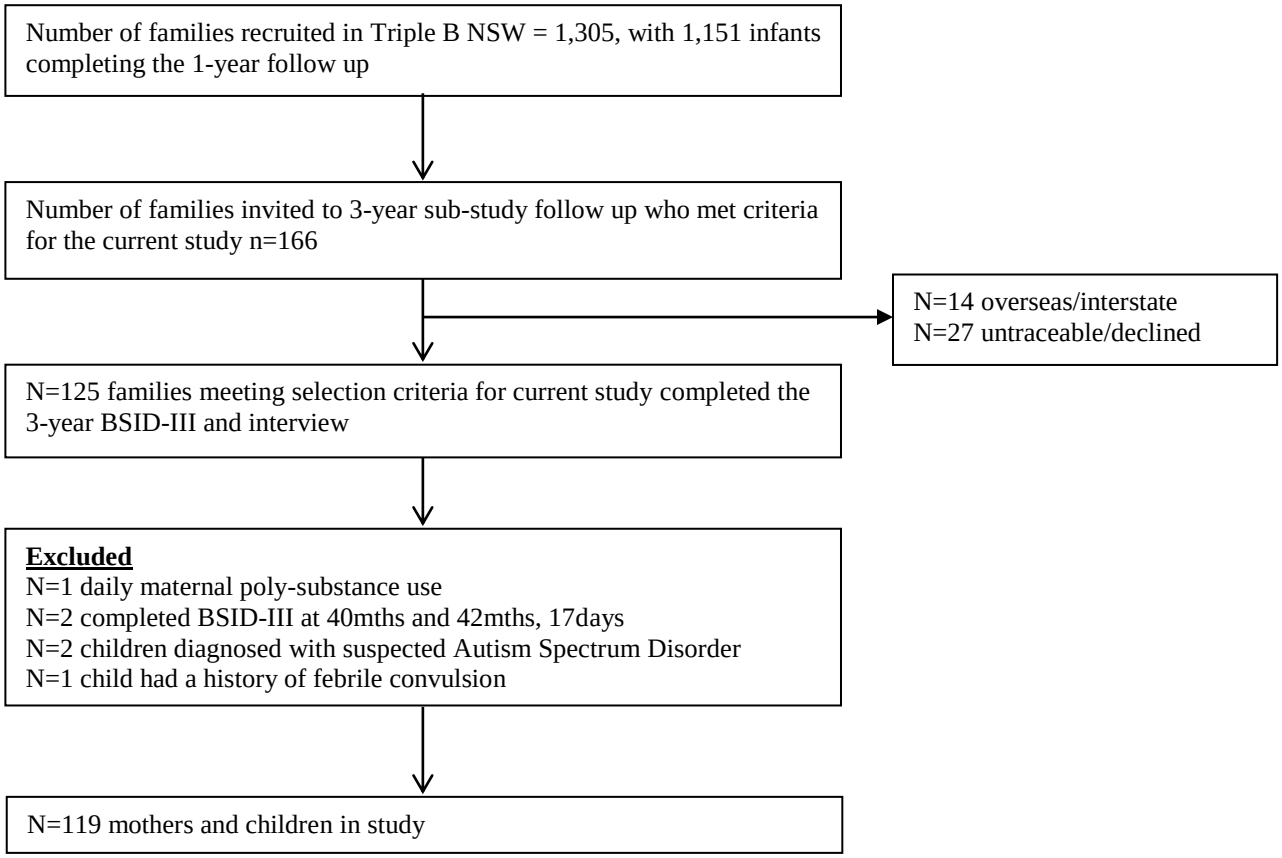


Figure 1: Recruitment and inclusion flow-chart.

4.3.2 Measures

Demographic Data

Life-time demographic information was obtained via maternal interview in Trimester 3 and included: country of birth; chronological age; level of education; and, identification as Aboriginal or Torres Strait Islander. Time-dependent demographic information obtained from maternal interview at the 3-year follow-up included: relationship status; living arrangements; employment status; and, number of biological children. SEIFA percentile rankings (ABS, 2013) were also obtained for each family, based on postcode, as a measure of socio-economic status.

Information pertaining to the child was obtained from maternal interview and from child health record books at 8-weeks, 1-year and 3-years. This included: child sex; gestational age; birth weight, height and head circumference; Apgar scores at one and five minutes post-birth; special care/intensive care admissions; birth order; handedness; medical history; and, child height, weight and head circumference at 3-years.

Development Assessment

The BSID-III was used to assess child development at 1-year and 3-years of age (Bayley, 2006a). The BSID-III comprises five objectively rated domains of development: cognition, receptive language, expressive language, fine motor and gross motor, as well as a caregiver rated questionnaire assessing social-emotional development and adaptive behaviour. The BSID-III reports good psychometric properties with moderate to strong reliability and validity (Albers and Grieve, 2007; Bayley, 2006b). Raw scores obtained on each objectively rated BSID-III domain were converted to standardised scaled scores (M=10, SD=3) based on the US normative sample using Pearson Psychcorp scoring software; all scores were manually cross-checked for accuracy.

Maternal Intellectual Functioning

The Test of Premorbid Functioning (TOPF) (Wechsler, 2011) was completed by mothers at either the 1-year or 3-year interview as an estimate of intellectual functioning. The TOPF is a brief standardised, norm referenced, irregular word reading task designed to predict an individual's full-scale Intelligence Quotient (IQ) (Wechsler, 2011). Due to the reliance on English skills, the measure is not recommended for people with sub-fluent English abilities.

4.3.3 Procedures

Eligible participants were contacted prior to their child's third birthday. Consenting parents schedule the BSID-III assessment as close to the child's third birthday as possible. The TOPF and BSID-III were administered following standardised administration procedures (Bayley, 2006a; Wechsler, 2011). Assessments were either conducted in a quiet space in the child's home or in an interview room at The University of New South Wales. As recommended, all 3-year BSID-III assessments were carried out with two trained assessors present to allow for interrater reliability to be undertaken and to ensure the pace of the assessment remained engaging. All assessors had completed a university undergraduate degree in social sciences and were enrolled in a post-graduate degree, had completed a BSID-III accreditation workshop and were experienced BSID-III assessors. Two trained assessors independently scored 28.8 percent (n=36) of all 3-year BSID-III assessments for the purposes of inter-rater reliability.

Obtaining Raked Weights

Maternal demographic data pertaining to SEIFA scores, age at birth, level of education, employment status, number of biological children, region of birth, Aboriginal or Torres Strait Islander status, and child sex were categorised according to ABS population data, as outlined in Table 2. Region of birth percentages were based on the following (ABS, 2015a) categories for women aged 30- to 34-years: Australia; United Kingdom; New Zealand; Southern, Eastern, Northern and Western Europe-"Europe other"; South-East, North-East, Southern and Central Asia-"Asia"; Sub-Saharan Africa-"Africa", and "Other". ABS data for women aged 30- to 34-years were also used to obtain Australian percentages of employment status (ABS, 2017) and level of education (ABS, 2015b), in line with the median age of the current cohort. Similarly, population percentages of biological children were calculated based on ABS data for women aged ≤ 34 years, in line with the median age of the current cohort (ABS, 2016b). Australian population percentages in Table 2 for the variables outlined above, were then used to obtain raked weights for the current cohort. As no mothers in the current cohort were ≤19 years of age, or ≥ 44 years of age at birth, a combined population percentage of 15.5 percent for women ≤ 24 years, and 4.3 percent for women ≥ 40 years was used for raking.

Raked weights did not take into consideration maternal relationship status, as available population data consistent with characteristics of the current cohort (mothers within the first 3 years of giving birth) were not available, and thus any attempt to match by this variable may have resulted in bias.

4.3.4 Statistical Analysis

Statistical analyses were conducted using the software IBM SPSS Statistics for Windows, version 22 (IBM, 2013), Stata version 8 SE (StataCorp, 2005), and GPower 3.0.10

(Faul et al., 2009) to compute Cohen's *'d'* effect sizes (Cohen, 1988). Demographic variables were summarised using frequency and descriptive statistics.

The primary outcome variable was child performance on the BSID-III at 3-years of age. Aim two was achieved in Stata by obtaining raked weights for participants, with a fixed maximum value of 5. Inter-rater reliability of the BSID-III was assessed via the one-way random, absolute, single-measure intra class correlation coefficients for each of the five objectively rated domains of the BSID-III. One sample t-tests were conducted to compare the mean scaled scores of the current cohort on each domain, to the US normative sample (aim three). Bonferroni adjusted significance level was set at 0.01 for multiple a priori analyses (Westfall et al., 1997). Approximate normality of BSID-III domains was assessed via examination of histograms and Q-Q plots, and skewness and kurtosis statistical acceptance levels were set at (+/-1.5) (Tabachnick and Fidell, 2013). As all scores fell within the possible ranges, no scores were considered true outliers. However, where assumptions were violated, statistical outliers greater than 3-times the inter-quartile range (IQR) were removed. Analyses were run including and excluding outliers (Appendix A) and results remained unchanged, as such, full cohort results were reported.

4.4 Results

4.4.1 Participant characteristics

Table 1 describes child characteristics from birth to the 3-year follow-up. Mean gestational age at birth, height, weight and head circumference were all consistent with national averages for a cohort of singleton, full-term children. Approximately 10 percent of children in the current cohort were transferred to a special care or neonatal intensive care unit; a rate that is lower than the national population average of 15.4 percent (Hilder et al., 2012) yet not unexpected given the inclusion criteria noted earlier.

At 3-years of age the cohort continued to have relatively minor health-related needs: 13 percent of children were using prescription medication; 10 percent of which was for a condition that was expected to last at least 12-months (such as asthma). Approximately 12 percent of children had been hospitalised for at least one night, with the primary reasons being asthma or viral related symptoms. Seven children underwent minor surgeries, predominantly for adenoidectomy, tonsillectomy or grommet insertion. Mean age at BSID-III assessment was 36-months.

Table 1. Child demographic characteristics

Child characteristics	**N**	**Total cohort**		
Male, N (%)	119	59 (49.6%)		
Weeks' gestation, M [SD]	119	39.7 [1.2]		
Birth weight (kg), M [SD]	119	3.5 [0.5]		
Birth length (cm), M [SD]	117	51.0 [2.7]		
Birth head circumference (cm), M [SD]	111	34.8 [1.5]		
Apgar score 1 minute, M [SD]	112	8.5 [1.3]		
Apgar score 5 minutes, M [SD]	112	9.0 [0.4]		
NICU or special care nursery at birth, N (%)	119	11 (9.2%)		
Required oxygen at birth, N (%)	118	10 (8.5%)		
3yr weight (kg), M [SD]	112	15.2 [1.8]		
3yr length (cm), M [SD]	112	96.9 [4.1]		
3yr head circumference (cm), M [SD]	111	50.5 [1.8]		
Current prescription medication use, N (%)	119	15 (12.6%)		
Injury requiring medical attention, N (%)	119	32 (26.8%)		
Hospitalized since birth, N (%)	119	14 (11.8%)		
Underwent surgery, N (%)	119	7 (5.9%)		
Handedness, N (%)		Right	Left	Ambidextrous/unclear
	119	110 (92.4%)	4 (3.4%)	5 (4.2%)

4.4.2 Maternal Characteristics and Raked Weights

As outlined in Table 2, the median age of mothers in the current cohort at the time of birth was 30- to 34-years. Consistent with Australian population data, just over half of the cohort was born in Australia (ABS, 2015a). The majority of families comprised of two children, with 79 percent of mothers engaged in some form of paid employment.

Approximately 87 percent of the cohort had obtained an additional, post school qualification, with 71 percent having undertaken undergraduate/postgraduate university studies. Comparatively, Australian population data show that 65 to 71 percent of women aged 25 to 44 have completed a tertiary qualification (ABS, 2015b), indicating that the study cohort was somewhat more highly educated than the Australian population. However, the mean cohort IQ estimate (TOPF Standard Score) was consistent with population expectations.

Raked weights improved the representativeness of the cohort relative to the Australian population data. Notably, maternal age at birth, estimated IQ, education, employment status, child gender, number of biological children and region of birth, were consistent with population data following raking. However, relative to the Australian population, the lowest SEIFA quartile remained somewhat underrepresented, and people of Aboriginal and Torres Strait Islander origins were overrepresented. Participants in the current cohort were also more likely to be in a married or defacto relationship, however Australian population data are not available for relationship status of parents of young children. Parents of young children are more likely to be in a relationship than the general population and parents of older children. Given population statistics may represent an under-estimation of married/defacto populations, they were not included in raking.

Table 2. Maternal characteristics

Characteristic	Raw Cohort Data, N (%) or M [SD]	Australian Population % or M [SD]	Weighted Cohort Data, % or M [SD]	US Population/Normative Data
SEIFA Percentile	1st = 4 (3.4%) 2nd = 17 (14.3%) 3rd = 31 (26.1%) 4th = 67 (56.3%)	1st = 24.2% 2nd =24.6% 3rd =25.1% 4th =26.1% (ABS, 2013)	1st = 13.9% 2nd =28.0% 3rd =28.5% 4th =29.7%	NA
TOPF Standard Score[$,%]	103.1 [13.2]	100 [15] (Wechsler, 2011)	98.5 [13.6]	Mean of 100, SD of 15 (Wechsler, 2011)
Age at birth (yrs)	≤19 = 0 (0%) 20-24 = 4 (3.4%) 25-29 = 23 (19.3%) 30-34 = 47 (39.5%) 35-39 = 38 (31.9%) 40-44 = 7 (5.9%) 45+ = 0 (0%)	≤19 = 2.8% 20-24 = 12.7% 25-29 = 27.2% 30-34 = 35.0% 35-39 = 18.0% 40-44 = 4.0% ≥45 = 0.3% (ABS, 2016a)	≤24 = 14.0% 25-29 = 27.7% 30-34 = 35.6% 35-39 = 18.3% ≥40 = 4.4%	28-years was the average age of giving birth in 2014 (Mathews and Hamilton, 2016)
ATSI	3 (2.5%)	2.5% (ABS, 2012a).	3.5%	NA

Characteristic	Raw Cohort Data, N (%) or M [SD]	Australian Population % or M [SD]	Weighted Cohort Data, % or M [SD]	US Population/Normative Data
Education	School completion or less = 16 (13.4%) Post school certificate, diploma or trade = 18 (15.1%) Bachelor degree or higher = 85 (71.4%)	School completion or less = 26.3% Post school certificate, diploma or trade = 31.1% Bachelor degree or higher = 42.6% (ABS, 2015b)	School completion or less = 25.8% Post school certificate, diploma or trade = 31.3% Bachelor degree or higher = 43.0%	Approximately 60% of parents in the US normative sample had completed formal education beyond schooling. 26% of 36-month old children's parents had completed a university degree or higher (Bayley, 2006b).
Married/ Defacto[%]	111 (93.3%)	64% were in a registered marriage or de facto relationship in 2012-2013, and 81% of families with children under 17 years were multiple parent families, (ABS, 2015c).	91.8%	71% Women aged 15-44 in the US were married or living in defacto like relationships (Bumpass and Lu, 2000)
Employed	94 (79.0%)	69.7 % women aged 30-34 years in July, 2015 (ABS, 2017).	70.4%	68.3-74.3% of women aged 25-44years (US Department of Labor, 2015).
Child gender	Female = 60 (50.4%) Male = 59 (49.6%)	Female = 48.6% Male = 51.4% (ABS, 2016c)	Female = 48.6% Male = 51.4%	50% of children in the US normative sample were male (Bayley, 2006b).

Characteristic	Raw Cohort Data, N (%) or M [SD]	Australian Population % or M [SD]	Weighted Cohort Data, % or M [SD]	US Population/Normative Data	
Biological children	1 = 29 (24.4%) 2 = 71 (59.7%) 3+ = 19 (16.0%)	1= 25.81%; 2 = 43.02%; 3+ = 31.17%. In 2011, women 34 years and younger (ABS, 2016b)	1 = 25.7% 2 = 43.0 % 3+ = 31.2%	Non-Hispanic 1 = 41.6% 2 = 32.5% 3+ = 15.8%. In 2014, women aged 15-44 years (Mathews and Hamilton, 2016; Hamilton et al., 2015).	Hispanic 1 = 34.0% 2 = 30.0% 3+ = 20.0%
Birth region	Australia = 77 (64.7%) Asia = 9 (7.6%) Europe other = 9 (7.6%) UK = 7 (5.9%) NZ = 3 (2.5%) Africa = 4 (3.4%) Other = 10 (8.4%)	Australia = 66.4% Asia = 16.5% Europe other = 3.7% UK = 3.4% NZ = 3.1% Africa= 1.7% Other = 5.3 Women 30-34yrs in 2011 (ABS, 2015a)	Australia = 64.8% Asia = 16.3% Europe other = 3.6% UK = 3.4% NZ = 3.1% Africa= 2.5% Other = 6.4%	Approximately 60% of the US population is white, followed by 19% Hispanic, 15% African American and 4% Asian (Bayley, 2006b).	

Mean and SD are underlined

[$]N=105: 14 parents chose not to participate/did not have sufficient English abilities to complete the assessment

[%] Cohort was not weighted according to these characteristics. 'Weighted cohort data' reflect demographic frequency and quantity results following application of raked weights described in the procedures section of this paper.

4.4.3 Inter-rater Reliability

Inter-rater reliability of 36 BSID-III administrations was assessed via one-way random, absolute, single-measure intraclass correlation coefficients for each of the five objectively rated sub-domains of the BSID-III. The resulting ICC's fell in the excellent range; Cognitive ICC = 0.995, Receptive Language ICC = 0.999; Expressive Language ICC = 0.987; Fine Motor ICC = 0.996; Gross Motor = 0.939, indicating a high degree of agreement between coders (Landers, 2015; Shrout & Fleiss, 1979).

4.4.4 BSID-III Score Comparisons: Current Cohort and Weighted Cohort with US Normative Sample

Mean scaled scores for the five BSID-III domains, in both the unweighted and weighted cohort significantly differed from the US normative sample at the 0.05 significance level, as demonstrated in Table 3. Differences remained significant at the 0.01 Bonferonni adjusted significance level across the domains except for the unweighted gross motor domain. Interestingly, cognitive scaled scores in the current cohort were significantly lower than the US normative sample, whereas the current cohort performed significantly higher than the US normative sample across all other domains where results were significant. Weighted effect sizes across all domains ranged from medium to large (i.e., *≤0.2=small; 0.21 to 0.5=medium; 0.51 to 0.8=large*) (Cohen, 1988). This, coupled with the magnitude of the mean differences in scaled scores obtained, suggests that these differences are likely to be of clinical significance.

Table 3. BSID-III scaled score comparison between current cohort (weighted and unweighted) and the US normative sample

Unweighted Data					
BSID-III Domain	**M [SD]**	**p**	***d***	**M $_{diff}$**	**M $_{diff}$ 95% CI**
Cognition	9.53 [1.53]	.001	0.20	-0.47	-0.75 – -0.19
Receptive language	11.02 [2.29]	<.001	0.38	1.02	0.60 – 1.43
Expressive language	11.20 [2.27]	<.001	0.45	1.20	0.79 – 1.61
Fine motor	11.58 [2.46]	<.001	0.58	1.58	1.13 – 2.03
Gross motor	10.45 [2.23]	.03	0.17	0.45	0.05 – 0.86
Weighted Data					
BSID-III Domain	**M [SD]**	**p**	***d***	**M $_{diff}$**	**M $_{diff}$ 95% CI**
Cognition	9.27 [1.34]	<.001	0.31	-0.73	-0.97 – -0.49
Receptive language	10.80 [2.25]	<.001	0.30	0.80	0.39 – 1.21
Expressive language	11.00 [2.26]	<.001	0.38	1.00	0.59 – 1.41
Fine motor	11.58 [2.19]	<.001	0.60	1.18	1.18 – 1.97
Gross motor	10.77 [2.42]	.001	0.28	0.77	0.33 – 1.21

Note: US scaled scores Mean [SD]=10 [3]; 'M_{diff}' = Mean difference between the current cohort mean scaled score and the US BSID-III normative sample mean scaled score of 10.
Bonferroni significance correction for multiple (5) a priori analyses= 0.01

4.5 Discussion

As aptitude stabilises, delay takes on greater clinical significance into toddlerhood and the preschool years. Three years for age represents a common diagnostic and paediatric assessment window. The BSID-III is widely used in Australia to assess the child development at this age. However, no Australian normative data are available. This study used a longitudinal, prospective cohort of 119 3-year-old children, with statistical weighting, to investigate whether 3-year-old Australian child performance on the BSID-III differed from the US normative sample. Critically, it provided an examination of detailed maternal demographic characteristics, took steps improve cohort representativeness, and drew

comparisons to both Australian and US population data in order to better understand whether detected differences reflect true population differences. Through this, it was possible to obtain means and standard deviations that may be used clinically to obtain z-scores, and provide a normative comparison group for term born, singleton 3-year old Australian children on the BSID-III.

Our Australian 3-year-old children scored significantly *higher* than the US BSID-III normative sample on the receptive language, expressive language and fine motor domains, a difference that remained significant following statistical adjustment to ensure cohort representativeness (Table 3). The result is consistent with results reported by Chinta et al (2014). However, contrary to previous research, children in the current cohort scored significantly *lower* than the US normative sample on the cognitive domain both pre- and post-statistical adjustment, and significantly *higher* than the US normative sample on the gross motor domain following adjustment (Table 3). Weighted effect sizes were medium to large (Cohen, 1988), and mean scaled score differences ranged from 0.7 to 1.2. Considering three scaled score points make up a standard deviation in US normative data, and smaller standard deviations were obtained in the current study, results suggest both statistical and clinical significance.

An understanding of cohort demographic characteristics is essential in order to assess whether detected differences likely reflect true population differences. Prior to the application of raked weights, the cohort was relatively advantaged when compared to the Australian population, with higher socio-economic status, higher rates of employment and higher educational attainment. Considering the known association between advantage/disadvantage and ability, this may have impacted unweighted BSID-III scores. Application of raked weights however, statistically adjusted the cohort in order to improve representativeness,

although individuals in the lowest quartile of the socio-economic status indicator remained underrepresented.

Although raked weights were applied to improve the representativeness of the cohort to Australian maternal population characteristics, important differences between the US normative sampling approach and the current sampling approach need to be considered. The inclusion/exclusion criteria of the current study were devised to match the US normative sampling approach as closely as possible. However, following sampling, individuals with specific conditions such as cerebral palsy, trisomy 21, and pervasive developmental disorder were selectively re-introduced into the US BSID-III normative sample, making up approximately 10 percent of the sample. The cohort in the current study is healthy, term, singleton children. While this was unavoidable as the proportion of children with expected difficulties in each age group of the US BSID-III normative sample was not clearly reported (Bayley, 2006b), and therefore not replicable, this may account for the higher scores obtained by the current cohort in the language and motor domains, when compared to the US normative sample (Bayley, 2006b). However, this does not account for the lower cognitive scores obtained by the current cohort when compared to the US normative sample.

To further understand possible explanations for detected differences between Australian and US normative data, an examination of differences between the US and Australian populations is necessary. Ethnicity may play a key role in the detected differences (Greene et al, 2013). As demonstrated, the ethnic make-up of the US population differs from the Australian population, with people of Hispanic and African American origins more prevalent in the US population, and people of Asian origins more prevalent in the Australian population. Furthermore, numbers of overseas born citizens and those with a home language other than English differ across countries, with 12.9 percent of Americans reportedly born overseas (United States Census Bureau, 2010) and 20.7 percent speaking a language other

than English at home (United States Census Bureau, 2015), compared to 28 percent of Australians born overseas (ABS, 2016d), and 23.2 percent speaking a language other than English at home (ABS, 2012b). Although difficult to quantify due to the diversity and complexity of 'ethnicity' as a construct, linguistic diversity and ethnicity have long been associated with development, ability and outcomes, with research suggesting both environmental and biological influences (Greene et al., 2012). As such, ethnic and linguistic diversity between Australia and the US may provide one explanation for detected differences in BSID-III scores obtained by the current cohort when compared to the US normative sample.

Parental education is another population factor for which the US normative sample is representative of the US population, and in which Australia and the US differ. The US BSID-III normative sample reported that 26 percent of parents of 3-year-old children had completed education at university degree level of higher (Bayley, 2006b), compared to 42.6 percent in the Australian population (ABS, 2015b). Considering the demonstrated association between parental educational attainment and offspring ability (Bee et al., 1982; Sameroff et al., 1987) and the known association between language and intellect (Wechsler, 2011), this educational attainment may account for the higher scores detected in the current adjusted cohort on the receptive language, expressive language, fine motor and gross motor domains. This however does not account for the lower scores obtained on the cognitive domain.

To understand potential reasons for contradictory results obtained by the current study in the cognition and gross motor domains, when compared to the study by Chinta, et al. (2014), sampling approaches should be considered. Chinta et al. (2014) reported no significant difference in gross motor abilities between the cohort and US normative data. Similarly, in the current study, after Bonferroni adjustment for multiple analyses, no significant difference was detected in unweighted gross motor scores. However, following

application of raked weights to improve cohort representativeness, significant differences were detected. While it is important to note that prior to application of raked weights, results in the current study already approached significance (having exceeded 0.05), whereas no trend at all was evident in the study by Chinta et al. (2014), the current results, with detected differences increasing once representativeness was improved, reinforce the importance of cohort representativeness.

Similarly, the significantly lower scores on the cognitive domain, when compared to the US population was unexpected considering Chinta et al. (2014) found that Australian 3-year old children obtained significantly higher scores on the cognitive domain when compared to the US normative sample. However, again this study did not account for maternal demographic characteristics (Chinta et al., 2014). Conversely, a Dutch study by Steenis et al. (2015), which took steps to ensure cohort representativeness, also reported lower scores obtained by 3-year-old children on the cognitive domain when compared to US normative data. As differences detected in the current cohort increased in significance following the application of weights, this may again may reinforce the importance of cohort representativeness.

However, the small spread of cognitive scores obtained in the current study should also be considered as another possible explanation for the detected difference in cognition. One standard deviation in the US normative sample is three scaled score points on average. Similarly, one standard deviation in the current cohort on the weighted and unweighted language and motor domains was between two and three scaled score points. Conversely, one standard deviation on the cognitive domain in the current weighted cohort was 1.34 scaled score points. Two possible explanations may account for this small spread on the cognitive domain: a possible sampling issue in the current study or a greater issue with the overall item structure of the cognitive domain of the BSID-III. A lack of spread in cognitive scores

reported in both Australian and international samples provides some support for the latter. Chinta et al. (2014) reported a cognitive standard deviation of 1.89 scaled score points and Steenis et al. (2015) reported a cognitive standard deviation of 1.23. This suggests many children are able to complete cognitive items up to a certain point, but very few children can complete items beyond that point, resulting in little spread of obtained scores. Small spread may suggest that the cognitive items are not discriminative enough to detect subtle differences in ability, which could point to a greater issue with the item structure of the cognitive domain of the BSID-III at 3-years for Australian and international populations.

4.5.1 Strengths and Limitations

Whilst the current study demonstrated many strengths in providing the first statically adjusted representative BSID-III data for 3-year-old children, some limitations should be noted. First, although raked weights were obtained based on a range of variables and applied to the data, the proportion of participants in some categories was small, and the lowest quartile of SEIFA scores was unable to be fully adjusted. Application of weights to a relatively small cohort meant that some participants in underrepresented categories were weighted approximately equivalent to five participants. It is possible that this may have inadvertently compounded sampling bias of variables not taken into account in the weighting process. However, considering scores remained relatively unchanged pre-and post-application of weights, this is unlikely to have significantly impacted results.

Second, as previously discussed, differences between US normative sampling approach and the current sampling approach may have impacted results, and may therefore account for higher scores obtained by the current cohort on the language and motor domains (Bayley, 2006b).

Third, all participants included in the current study received a brief summary report on their child's development at 1-year of age. Delays may have thus been detected earlier in the current cohort than expected in the general population, and early intervention sought sooner than would otherwise be expected, ameliorating outcomes.

Nevertheless, results of the current study have important implications for the utility of the BSID-III in an Australian population. Higher scores obtained on language (receptive and expressive) and motor (fine and gross) domains by the current cohort suggest that utilisation of US BSID-III normative data in Australian 3-year old children will lead to the under-detection and diagnosis of language and motor delays; and thus prohibit access to early interventions. Furthermore, results suggest that utilisation of cognitive US BSID-III normative data for Australian 3-year old children may lead some children to be unnecessarily referred for early cognitive intervention. Means and standard deviations provided may be used clinically to obtain z-score and provide a normative comparison group for term born, singleton 3-year old Australian children on the BSID-III. Research providing full age range, representative, Australian normative data, is warranted.

4.6 References

Albers CA and Grieve AJ (2007) Test review: Bayley, N. (2006). Bayley scales of infant and toddler development–third edition. San Antonio, TX: Harcourt assessment. *Journal of Psychoeducational Assessment* 25(2): 180-190.

Anderson LM, Shinn C, Fullilove MT, Scrimshaw SC, Fielding JE, Normand J et al. (2003) The effectiveness of early childhood development programs: A systematic review. *American Journal of Preventive Medicine*, *24*(3), 32-46.

Anderson PJ, De Luca CR, Hutchinson E, Roberts G and Doyle LW (2010) Underestimation of developmental delay by the new Bayley-III Scale. *Archives of Pediatrics and Adolescent Medicine* 164(4): 352-356.

Australian Bureau of Statistics (2012a) Census of Population and Housing - Counts of Aboriginal and Torres Strait Islander Australians, *2011.* Available at: http://www.abs.gov.au/ausstats/abs@.nsf/lookup/2075.0main+features32011 (accessed 17 November 2016.

Australian Bureau of Statistics (2012b) 2011 Census shows Asian languages on the rise in Australian households. Available at: http://www.abs.gov.au/websitedbs/censushome.nsf/home/CO-60 (accessed 4 April 2017).

Australian Bureau of Statistics (2013) 2033.0.55.001 - Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia.

Australian Bureau of Statistics (2015a) 34120DO006_201415 Migration, Australia, 2014-15. Stat dataset: Estimated Resident Population by Country of Birth, 30 June 1992 to 2015.

Australian Bureau of Statistics (2015b) 6227.0 - Education and Work, Australia, May 2015.

Australian Bureau of Statistics (2015c) 4442.0 - Family Characteristics and Transitions, Australia, 2012-13.

Australian Bureau of Statistics (2016a) 3301.0- Births, by Nuptiality, by Age of Mother.

Australian Bureau of Statistics (2016b) 3301.0- 24b Number of Children Ever Born by Age and Sex of Parent.

Australian Bureau of Statistics (2016c) 3101.0 - Australian Demographic Statistics, Jun 2016: Table 59- Estimated Resident Population By Single Year Of Age, Australia.

Australian Bureau of Statistics (2016d) Overseas born Aussies highest in over a century. Available at: http://www.abs.gov.au/ausstats/abs@.nsf/lookup/3412.0Media%20Release12014-15 (accessed 4 April 2017).

Australian Bureau of Statistics (2017) 6291.0.55.001- Labour force, Australia, Table 01. Labour force status by Age, Social marital status, and Sex.

Bayley N (2006a) *The Bayley Scales of Infant and Toddler Development, 3rd Edition.* San Antonio, Texas: Pearson Psychcorp.

Bayley N (2006b) *Bayley Scales of Infant and Toddler Development, 3rd Edition, Technical Manual.* San Antonio, TX: Pearson PsychCorp.

Bee HL, Barnard KE, Eyres SJ, Gray CA, Hammond MA, Spietz AL et al. (1982) Prediction of IQ and language skill from perinatal status, child performance, family characteristics, and mother-infant interaction. *Child Development*, 1134-1156.

Bull R, Espy KA and Wiebe SA (2008) Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology* 33(3): 205-228.

Bumpass L and Lu HH (2000) Trends in cohabitation and implications for children's family contexts in the United States. *Population Studies* 54(1): 29-41.

Brownell CA, and Kopp CB, editors (2010) *Socioemotional development in the toddler years: Transitions and transformations*. Guilford Press; New York.

Campbell SB (1995) Behavior problems in preschool children: A review of recent research. *Journal of child Psychology and Psychiatry* 36(1): 113-149.)

Campbell FA, Pungello EP, Miller-Johnson S, Burchinal M and Ramey CT (2001) The development of cognitive and academic abilities: growth curves from an early childhood educational experiment. *Developmental Psychology*, *37*(2), 231.

Chakrabarti S and Fombonne E (2001) Pervasive developmental disorders in preschool children. *The Journal of the American Medical Association* 285(24):3093-3099.

Charman T, Taylor E, Drew A, Cockerill H, Brown JA and Baird G (2005) Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry* 46(5): 500-513.

Chinta S, Walker K, Halliday R, Loughran-Fowlds A and Badawi N (2014) A comparison of the performance of healthy Australian 3-year-olds with the standardised norms of the Bayley Scales of Infant and Toddler Development (version-III). *Archives of Disease in Childhood* 99(7): 621-624.

Cohen J (1988) *Statistical Power Analysis for the Behavioural Sciences* (2nd ed.). New York, NY: Psychology Press.

Deary IJ, Whiteman MC, Starr JM, Whalley LJ and Fox HC (2004) The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology* 86(1): 130.

Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P et al. (2007) School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428.

Faul F, Erdfelder E, Buchner A, and Lang AG (2009) Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41: 1149-1160.

Friedman NP, Miyake A, Robinson JL and Hewitt JK (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: a behavioral genetic analysis. *Developmental Psychology* 47(5): 1410.

Glasson EJ, Bower C, Petterson B, de Klerk N, Chaney G and Hallmayer JF (2004) Perinatal factors and the development of autism: a population study. *Archives of general Psychiatry* 61(6): 618-627.

Greene, M. M., Patra, K., Nelson, M. N., & Silvestri, J. M. (2012). Evaluating preterm infants with the Bayley-III: patterns and correlates of development. *Research in Developmental Disabilities*, *33*(6), 1948-1956. doi: http://dx.doi.org/10.1016/j.ridd.2012.05.024.

Greene, M. M., Patra, K., Silvestri, J. M., & Nelson, M. N. (2013). Re-evaluating preterm infants with the Bayley-III: Patterns and predictors of change. *Research in Developmental Disabilities*, *34*(7), 2107-2117. doi: https://doi.org/10.1016/j.ridd.2013.04.001

Hamilton BE, Martin JA sterman MJK, Curtin SC, Mathews TJ (2015) Births: Final Data for 2014. *National Vital Statistics Reports* 64(12): 1-63.

Hess RD, Holloway SD, Dickson WP and Price GG (1984) Maternal variables as predictors of children's school readiness and later achievement in vocabulary and mathematics in sixth grade. *Child Development*, 1902-1912.

Hilder L, Zhichao Z, Parker M, Jahan S, Chambers G (2014) *Australia's Mothers and Babies 2012*. Perinatal statistics series 30: Cat. PER 69. Canberra: AIHW.

Hutchinson D, Wilson J, Allsop S, Elliott E, Najman J, Burns L, et al. (2017) Cohort profile: The Triple B Pregnancy Cohort Study: A longitudinal study of the relationship between alcohol, tobacco and other substance use during pregnancy and the health and wellbeing of Australian children and families. *International Journal of Epidemiology*, accepted manuscript.

IBM Corp. (2013) *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: IBM Corp.

Kleinman JM, Ventola PE, Pandey J, Verbalis AD, Barton M, Hodgson S, et al. (2008) Diagnostic stability in very young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 38(4): 606-615.

La Paro KM and Pianta RC (2000) Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research*, *70*(4), 443-484.

Landers RN (2015) Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *The Winnower 2*:e143518.81744. DOI: 10.15200/winn.143518.81744

Lloyd M, MacDonald M and Lord C (2013) Motor skills of toddlers with autism spectrum disorders. *Autism* 17(2): 133-146.

Marchman VA and Fernald A (2008) Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science* 11(3): F9-F16.

Mathews T and Hamilton BE (2016) Mean age of mothers is on the rise: United States, 2000-2014. *NCHS data brief* 232*:* 1-8.

Piek JP, Dawson L, Smith LM and Gasson N (2008) The role of early fine and gross motor development on later motor and cognitive ability. *Human Movement Science* 27(5): 668-681.

Intraclass Correlations: Uses in Assessing Rater Reliability in Shrout PE and Fleiss JL (1979) Division of Biostatistics Columbia University, School of Public. Psychological Bulletin 86(2): 420-428.

Riou EM, Ghosh S, Francoeur E and Shevell MI (2009) Global developmental delay and its relationship to cognitive skills. *Developmental Medicine & Child Neurology*, *51*(8), pp.600-606.

Rose SA and Feldman JF (1995) Prediction of IQ and specific cognitive abilities at 11 years from infancy measures. *Developmental Psychology* 31(4): 685.

Sameroff AJ, Seifer R, Barocas R, Zax M and Greenspan S (1987) Intelligence quotient scores of 4-year-old children: Social-environmental risk factors. *Pediatrics*, *79*(3), 343-350.

Sameroff AJ, Seifer R, Baldwin A, and Baldwin C (1993) Stability of intelligence from preschool to adolescence: The influence of social and family risk factors. *Child Development*, *64*(1), 80-97.

Shevell M, Majnemer A, Platt RW, Webster R and Birnbaum R (2005). Developmental and functional outcomes at school age of preschool children with global developmental delay. *Journal of Child Neurology* 20(8): 648-654.

Shrout PE and Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420.

Silva PA, Williams S and McGee R (1987) A longitudinal study of children with developmental language delay at age three: later intelligence, reading and behaviour problems. *Developmental Medicine & Child Neurology* 29(5): 630-640.

StataCorp. 2005. *Stata Statistical Software: Release 9*. College Station, TX: StataCorp LP.

Steenis LJ, Verhoeven M, Hessen DJ and Van Baar AL (2015) Performance of Dutch children on the Bayley III: a comparison study of US and Dutch norms. *PloS one* 10(8): p.e0132871.

Tabachnick BG and Fidell LS (2013) *Using Multivariate Statistics (6th Edition)*. Boston; Pearson.

United States Department of Labor (2015) Labor Force Participation rates. Labor force participation rate by age and sex, 2015 annual averages and 2024 projections. Available at: https://www.dol.gov/wb/stats/Laborforce_par_rate_age_sex_2015_txt.htm (accessed 17 November 2016).

United States Census Bureau (2010) The Foreign-Born Population in the United States. Data derived from: American Community Survey. Available at: https://www.census.gov/newsroom/pdf/cspan_fb_slides.pdf (accessed 17 November 2016).

United States Census Bureau (2015) Detailed Languages Spoken at Home and Ability to Speak English for the Population 5 Years and Over: 2009-2013. Available at: https://www.census.gov/data/tables/2013/demo/2009-2013-lang-tables.html (accessed 17 November 2016).

Walker K, Badawi N, Halliday R and Laing S (2010) Brief Report: Performance of Australian children at one year of age on the Bayley scales of infant and toddler development (Version III). *The Australian Educational and Developmental Psychologist* 27(01): 54-58.

Wechsler D (2011) Test of premorbid functioning. UK version (TOPF UK). *UK: Pearson Corporation*.

Westfall PH, Johnson WO and Utts JM (1997) A Bayesian perspective on the Bonferroni adjustment. *Biometrika* 84(2): 419-427.

4.7 Appendix

Appendix A: One sample t-test results comparing current cohort to the US normative sample with outliers removed

Unweighted Data					
BSID-III Domain	**N**	**M [SD]**	**p**	**M $_{diff}$ 95% CI**	***d***
Cognition	117	9.42 [1.28]	<.001	-0.82 – -0.35	0.25
Receptive language	117	11.15 [2.05]	<.001	0.78 – 1.53	0.45
Expressive language	118	11.28 [2.11]	<.001	0.89 – 1.66	0.49
Fine motor	119	11.58 [2.46]	<.001	1.13 – 2.03	0.58
Gross motor	118	10.39 [2.12]	.05	0.00 – 0.78	0.15
Weighted Data					
BSID-III Domain	**N**	**M [SD]**	**p**	**M $_{diff}$ 95% CI**	***d***
Cognition	117	9.20 [1.15]	<.001	-1.01 – -0.59	0.35
Receptive language	117	10.80 [2.25]	<.001	0.39 – 1.21	0.30
Expressive language	118	11.02 [2.21]	<.001	0.62 – 1.42	0.46
Fine motor	119	11.58 [2.19]	<.001	1.18 – 1.97	0.60
Gross motor	118	10.71 [2.34]	.001	0.28 – 1.14	0.26

Note: 'M_{diff}' = Mean difference between the current cohort mean scaled score and the US BSID-III normative sample mean scaled score of 10.

Study 4

Predictive utility of the Bayley Scales of Infant and Toddler Development- 3rd Edition (BSID-III) from 1-year to 3-years of age.

Ingrid Honan*, BPsych (Hons), Delyse Hutchinson, PhD, Melanie Porter, PhD, Steve Allsop, PhD, Jake Najman, PhD, Elizabeth J. Elliott, MD, Lucinda Burns, PhD, Craig A. Olsson, PhD, Richard P. Mattick, PhD

Paper has been prepared in accordance with authorship guidelines for:

Journal of Child Neuropsychology

Paper submitted for publication (August 2017) – See Appendix F

5.1 Abstract

The Bayley Scales of Infant and Toddler Development- 3rd Edition (BSID-III) is a standardised psychometric assessment tool used internationally to objectively assess infant development. However, research examining the predictive utility of the BSID-III is inconsistent. Moreover, despite research suggesting developmental trajectories differ by sex, the predictive utility of the BSID-III stratified by sex, has not been examined. The aim of this study was to examine the predictive utility of the BSID-III from age 1-year to 3-years, and to investigate whether predictive utility differed by sex. A cohort of 122 mothers and their offspring were drawn from a longitudinal pregnancy cohort study. Mothers completed structured interviews and children were assessed using the BSID-III at both 1-year and 3-years of age.

Results indicated that performance at 1-year was significantly and positively correlated with performance at 3-year, with fair associations (r= 0.26 to 0.44). Results remained relatively similar after controlling for: maternal age at birth; linguistic background; education; socio-economic status; parity; and, infant hospitalisations since birth. The receptive language domain of the BSID-III was most stable over time, with 1-year scores explaining 18 percent of variation in 3-year scores, compared to a range of four to seven percent on other BSID-III domains. When broken down by sex, performance on receptive language, fine motor and gross motor domains held significant predictive utility for females, and performance on receptive language and expressive language domains held significant predictive utility for males, with explained variances ranging from 11 to 19 percent. Knowledge of the predictive utility of BSID-III domains will assist clinicians to make prognostic decisions and inform appropriateness for early intervention.

5.2 Introduction

Standardised neuropsychological assessment tools have been developed to help clinicians objectively and reliably assess whether a child is performing below, consistent with, or above their same aged peers (Aylward, 2002). If a child is performing below their same aged peers at a given point in time, this information is commonly used clinically as an indicator of potential future delay, and thus is often a precursor for early intervention. Yet not all delays are equally predictive of future ability. Different behaviours and skills at different ages are thought to hold varying degrees of predictive utility. For example, fine motor abilities assessed on the Ages and Stages Questionnaire between 4- and 48-months of age were not found to predict school aged cognitive function, but gross motor abilities were (Piek, Dawson, Smith, & Gasson, 2008). Understanding an assessment tool's predictive utility (i.e. the degree to which scores at one timepoint predict performance at on a measure assessing the same or similarly ability at a future timepoint) is of clinical import.

The Bayley Scales of Infant and Toddler Development-3rd Edition (BSID-III) (Bayley, 2006a) are one of the most widely used standardised developmental assessment tools in Australia for children aged 0- to 42-months (Aylward, 2002). Previous versions of the tool came under criticism due to poor predictive utility (Aylward, 2002; Hack et al., 2005; Crowe, Deitz, & Bennett,1987; Janssen et al., 2011). Moreover, whilst some research reported better predictive utility in at risk/delayed populations compared to low risk/typically developing populations (Harris, Megens, Backman, & Hayes, 2005; Aylward, 2002), results were conflicting, with some research reporting poor stability in scores from 1- to 2-years of age for specific risk groups (such as trisomy 21 versus medically fragile) (Niccols & Latchman, 2002). A recent meta-analysis of 16 studies of the BSID, BSID-II and one study of the BSID-III, in very preterm/low birth weight children, reported stronger predictive power of the Mental Development Index (MDI) of the BSID, assessing functions of cognition and

language, than the Psychomotor Development Index (PDI), assessing motor functions (Luttikhuizen dos Santos, de Kieviet, Königs, van Elburg, & Oosterlaan, 2013). However, language and motor domain specific delays (i.e. gross motor delays with age appropriate fine motor skills), may be confounding these results. Other studies have reported low to moderate predictive utility for previous versions of the BSID (Aylward 2002; Crowe, Deitz, & Bennett,1987; Hack et al., 2005; Potharst et al., 2012; Roberts Anderson, Doyle, & Victorian Infant Collaborative Study Group, 2010).

Both developmental research and studies on earlier versions of the BSID suggest that developmental trajectories may vary based on a range of factors, including child sex, and therefore, that the predictive utility of developmental assessment tools may vary for males and females and at different ages (Hyde & Linn 1988; Lung, Shu, Chiang, Chen, & Lin, 2009; Reznick, Corley, Robinson, & Matheny, 1997; Sajaniemi, Hakamies-Blomqvist, Katainen, & von Wendt, 2001). Research examining the underlying causes for differing developmental trajectories based on sex, point to a range of factors including both: environmental factors, such as sex-typed behaviour (i.e. boys encouraged to participate in rough and tumble play, and girls encouraged to fine motor play tasks such as craft) (Golombok et al., 2008); and biological factors, such as differences in brain volume and region development (Baron-Cohen, Knickmeyer, & Belmonte, 2005). When examining the BSID specifically, a longitudinal study of 70 children from 6- to 36-months on the previous version of the BSID, reported male performance to be more stable across the study period than female performance. However, this appeared to be due to a large growth in female language ability between 6- to 18-months, whereas performance at 18-months was predictive of performance at 36-months (Lung et al., 2009). Here, results highlight sex as a potential moderator variable in predictive utility research, and thus emphasise the importance of stratification by sex in order to appropriately interpret the predictive utility of the BSID-III.

Revision of the BSID-II to the BSID-III resulted in changes to approximately 50 percent of the items, and the development of five domains: cognition; receptive language; expressive language; fine motor and gross motor (Albers & Grieve, 2007). Although test re-test validity studies conducted during the re-norming process reported correlation coefficients of 0.72 to 0.81 overall (with increased correlation coefficients in older age groups), predictive utility was not assessed during test development (Bayley, 2006b). Stability and predictive validity studies have since aimed to address this gap, however the majority of research has been conducted in preterm samples with somewhat inconsistent findings. For example, Spittle et al. (2013) reviewed the predictive utility of the motor scale of the BSID-III at 2-years of age on motor outcomes at 4-years of age in a sample of 96 very preterm children. Although specificity was reportedly high, sensitivity was low, with many children with motor delay at 4-years on the Movement Assessment Battery for Children- 2nd Edition (MABC-2) going undetected on the BSID-III motor scale at 2-years of age (Spittle et al., 2013). Moreover, Spencer-Smith, Spittle, Lee, Doyle and Anderson (2015) investigated the utility of the cognitive and language domains of the BSID-III at 2-years of age for predicting outcomes on the Differential Ability Scale-II at 4-years of age in 105 preterm children. Results again suggested poor predictive utility of the BSID-III. Similarly, Lobo et al (2014) examined 24 low risk and 30 preterm infants across seven time-points between 3- and 24-months of age and reported poor stability in BSID-III scores over time.

Conversely, Bode, D'Eugenio, Mettelman and Gross (2014) compared 156 preterm and 155 term children's performance on the cognitive and language BSID-III domains at 2-years of age, with outcomes on the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) at 4-years of age, and reported strong predictive utility for preterm children (r=0.73 to r=0.84) and moderate predictive utility for term children (r=0.63 and r=0.67) (i.e., *<0.25= weak; 0.25 to 0.5=fair; 0.5 to 0.75=moderate; >0.75=strong* (Colton, 1974)).

Moreover, a study of 131 preterm children examining the stability of BSID-III from 8- to 20-months of age reported fair to moderate correlation coefficients (0.37 to 0.56) (Greene, Patra, Silvestri, & Nelson, 2013).

Inconsistency of the extant literature highlights the need for further investigation of the predictive utility of the BSID-III. To date, no studies have examined the predictive utility of all five objectively rated domains of the BSID-III in a general population cohort of infants. Moreover, the predictive utility of the BSID-III by sex remains unknown.

The aims of the current study were to: (1) investigate whether 1-year performance on the BSID-III predicted 3-year performance on the BSID-III (both statistically and clinically), after adjustment for key background covariates (maternal age at birth, linguistic background, education, socio-economic status, parity and infant hospitalisations since birth) and (2) examine whether predictive utility of the BSID-III differs by child sex.

Criteria for clinical significance was set apriori, and required: (1) a Pearson correlation co-efficient of greater than 0.3 (indicating a linear relationship within the 'fair' or higher range (Colton, 1974); and, (2) the proportion of explained variance to be greater than 10 percent, following covariate adjustment. This is consistent with literature stating that correlations in the fair range reflect a linear relationship (Colton, 1974), and ensures that the variance in 3-year scores explained by 1-year scores is greater than that explained by demographic factors alone, while allowing for the high degree of individual variability in aptitude and the 'catch-up' of delay known to occur across this period (Anderson & Burnett, 2017).

5.3 Method

5.3.1 Participants

Data were drawn from the Triple B Pregnancy Cohort Study; a prospective longitudinal study of 1,305 pregnant women, partners and their infant offspring (post-birth). Women were recruited during pregnancy from three major public hospital general antenatal clinics, birth centres and high-risk/substance use antenatal clinics across NSW, Australia (Hutchinson et al., 2017; McPhie et al., 2017). Full cohort eligibility criteria, retention rates and characteristics are described elsewhere (Hutchinson et al., 2017).

Recruitment of a 3-year subsample was undertaken for the purposes of this thesis. Eligibility criteria for the current study included: being pregnant; ≥16 years of age; adequate mental ability and English literacy; intention to reside in Australia until the child's first birthday; no previous siblings enrolled in the study; intention to remain one of the primary caregivers for the child; no known major medical complications for the mother or fetus during pregnancy (e.g., chromosomal abnormalities); completion of previous follow-up time points (including 1-year BSID-III); completion of 3-year BSID-III between 33 and 39-months; and being a singleton child. While selection criteria for the current cohort was devised in an effort to closely match US normative sampling techniques, the proportion of children with developmental difficulties reintroduced into each age group of the US normative sample was not clearly reported (Bayley, 2006b), and therefore not replicable.

Of the 166 eligible mothers and children invited to participate in the 3-year follow-up, the final available subsample included in the current study was 122 (see Figure 1). Informed consent was obtained from all mothers, and from legal guardians for child participants. Ethics approval was granted by relevant University and NSW Health Human Research Ethics Committees (lead approval numbers: HREC/11/RPAH/153; HREC/08/RPAH/218; X11-0111; X12-0232).

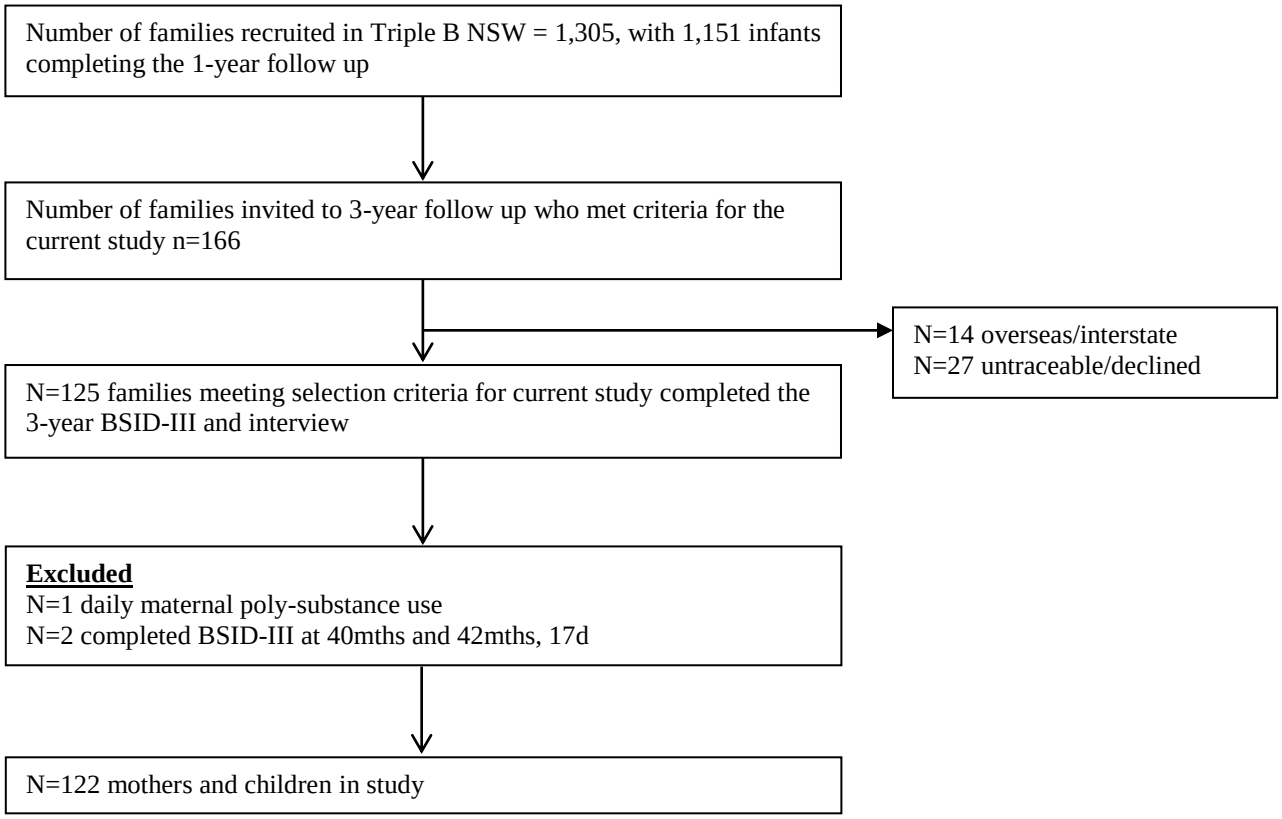


Figure 1: Recruitment and participant inclusion flow-chart

5.3.2 Measures

Demographic variables

Structured interviews in Trimester 3 of pregnancy were conducted to obtain the following life-time maternal demographic data: country of birth; chronological age; level of education and identification as Aboriginal or Torres Strait Islander. Country of birth data were subsequently categorised into a dichotomous variable based on the official language of the mother's country of birth (Non-English Speaking Background country of birth or NESB). SEIFA percentile rankings were also obtained based on postcode, as a measure of socio-economic status (Australian Bureau of Statistics (ABS), 2013).

Maternal relationship status, employment status and number of biological children data were obtained via structured interview at the 3-year follow-up. Employment status was coded as "employed" or "unemployed", with employment including full-time, part-time and

causal employment. Mothers who reported “home duties”, “student”, “unemployed”, “disability pension” or “other”, for example, as their primary form of employment were considered unemployed. Number of biological children was dichotomised into an “only child” variable with possible yes/no responses.

Mothers were also asked to complete the Test of Premorbid Functioning (TOPF) as an estimate of intellectual functioning at either the 1-year or 3-year follow-up (Wechsler, 2011). As the TOPF is a reading task comprised of 70 irregular words, it is not recommended for people with sub-fluent English abilities (14 mothers in the current cohort were not administered it for this reason or declined to complete the TOPF). Raw scores were converted to standard scores (M=100; SD=15) using normative data.

Child demographic data were obtained via structured maternal interview at 8-weeks, 1-year and 3-years, and from child health record books at each timepoint. Information obtained included child: sex; gestational age; birth weight; height; head circumference; Apgar scores at one and five minutes’ post-birth; special care/intensive care admissions; oxygen use after birth; handedness; medical history (including developmental concerns); and height, weight and head circumference at 1- and 3-years.

BSID-III Development Assessment

The Bayley Scales of Infant and Toddler Development - 3rd Edition (BSID-III) was used to assess child development at 1-year and 3-years of age (Bayley, 2006a). The BSID-III comprises five objectively rated domains of development: cognition, receptive language, expressive language, fine motor and gross motor, as well as a caregiver rated questionnaire assessing social-emotional development and adaptive behaviour. The BSID-III reports generally good psychometric properties with moderate to strong reliability and validity (Albers & Grieve, 2007; Bayley, 2006b). Raw scores obtained on each objectively rated

BSID-III domain were converted to standardised scaled scores (M=10, SD=3) based on the US normative sample using Pearson Psychcorp scoring software. Scores were manually cross-checked for accuracy.

5.3.3 Procedures

Maternal interviews at Trimester 3, 8-weeks and 1-year postnatal were conducted either in person or over the phone (based on participant preference). In order to schedule development assessments, participants were contacted immediately prior to the child's first and third birthday to arrange an assessment time. Structured maternal interviews at 3-years were conducted on the same day as the developmental assessment, where possible.

The TOPF and BSID-III were administered following standardised instructions and administration procedures (Bayley, 2006a; Wechsler, 2011). Assessments were either conducted in a quiet space in the child's home or in an interview room at The University of New South Wales (UNSW). All 3-year BSID-III assessments were conducted with two trained assessors to ensure the pace of the assessment remained engaging (Bayley, 2006a). All assessors had completed a University undergraduate degree in social sciences and were enrolled in a post-graduate degree; had completed a BSID-III accreditation workshop and were experienced BSID-III assessors. To achieve reliability, staff administered at least ten assessments under the supervision of a trained assessor and were required to achieve 100 percent scoring agreement on two consecutive assessments prior to independent administration. Two trained assessors independently scored 28.8 percent (n=36) of all 3-year BSID-III assessments for the purposes of inter-rater reliability.

5.3.4 Statistical Analyses

Statistical analyses were conducted using the software IBM SPSS Statistics for Windows, version 22 (IBM 2013). Demographic variables were summarised using frequency and descriptive statistics. Inter-rater reliability was assessed via the one-way random, absolute, single-measure intra class correlation coefficients for each of the five objectively rated domains of the BSID-III. The primary outcome variables were infant performance on the five objectively measured domains of the BSID-III at 1- and 3-years.

Dependent t-tests and Pearson correlation coefficients were obtained for the full cohort, as well as for each sex, to assess the relationship between BSID-III performance at 1-year and 3-years. As scaled scores are standardised, no significant differences on dependent t-test analyses were expected. This would suggest that 1-year performance has some predictive power, as performance relative to peers did not vary over time. Assumptions were assessed via examination of histograms and Q-Q plots of differences. Although no extreme outliers were detected (that is data points greater than 3 box-lengths away from box ends), one moderate outlier was detected on the receptive language domain and three on the gross motor domain and were, therefore, excluded from analyses.

Multiple regression models were then fitted to determine whether 1-year performance significantly predicted 3-year performance on the cognitive, receptive language, expressive language, fine motor and gross motor domains of the BSID-III, controlling for: only child/sibling; hospitalisation since birth; maternal age at birth; maternal linguistic background; maternal education and socio-economic status. Maternal education was re-coded into a dichotomous yes/no variable in response to “mother had completed a university/college degree”. Analyses were conducted for the full cohort, as well as by child sex. Linearity was assessed via scatterplots. Assumptions of homoscedasticity, normality of errors and multicollinearity were assessed by examination of Q-Q plots, skewness and kurtosis statistics

(+/-1.5) (Tabachnick, Fidell, & Osterlind, 2001), histogram and scatterplots of standardised residuals to fitted values and by examining tolerance (>0.10) and variance inflation factors (<10) (SPSS Web Books Regression with SPSS Chapter 2 – Regression Diagnostics). Examination of residuals in full model analyses revealed two outliers on the cognitive domain, two on the receptive language domain, one on the expressive language domain and one outlier on the fine motor domain. Log10 transformation was conducted on 3-year gross motor scaled scores to improve skewness and spread of residuals. Examination of performance by sex revealed: two outliers on the cognitive domain and two outliers on both the receptive and expressive language domains for boys; and two outliers on the cognitive domain and one on each expressive language, fine motor and gross motor domains for girls. Although the *statistical* significance of the results did not differ based on the inclusion or exclusion of outliers, outliers appeared to impact the interpretation of the *clinical* significance of the results by impacting the proportion of variance explained at 3-years by 1-year BSID-III results. As such, outliers were excluded from analyses.

5.4 Results

5.4.1 Cohort Characteristics

Table 1 shows the maternal and child demographic characteristics. High frequency of maternal employment, high proportion of mothers having completed a bachelor or college degrees and a high average SEIFA percentile, suggest the cohort is relatively advantaged. Given the association between socio-economic factors and intelligence, the demographic characteristics of the cohort may have implications for BSID-III results (Forns et al., 2012). However, it should be noted that maternal TOPF scores for the current cohort were consistent with the TOPF normative cohort and, therefore, do not suggest elevated maternal intelligence.

Child demographic data show that approximately 10 percent of the cohort were transferred to neonatal intensive or special care units, an indicator for increased risk of developmental difficulties. Furthermore, just over 10 percent of children had a chronic illness requiring medication. Eleven percent of children had been readmitted to hospital and stayed overnight since birth, with 6.6 percent requiring surgery. Surgeries were generally low-risk surgeries for procedures such as tonsillectomies, adenoidectomies and grommet insertion, which are generally not expected to impact development. Eight mothers raised concerns indicting they felt their child's development was delayed, with three children undergoing investigations with suspected diagnoses of an Autism Spectrum Disorder, and a further five parents reporting concerns over language development.

Table 1. Cohort demographic characteristics

Maternal Characteristics			
Demographic characteristic	**N**	**M [SD] or N (%)**	
SEIFA percentile, M [SD]	122	72.1	[24.1]
Age at birth, M [SD]	122	32.9	[4.4]
Married/ defacto, N (%)	122	114	(93.4)
Employed, N (%)	122	95	(77.9)
English speaking birth country, N (%)	122	95	(77.9)
Level of Education, N (%)			
School completion or below	122	17	(13.9)
TAFE or technical college	122	19	(15.6)
University/College	122	86	(70.5)
TOPF Standard Score, M [SD]	108	103.3	[13.3]
Aboriginal /Torres Strait Islander, N (%)	122	3	(2.5)
Child Characteristics			
Demographic characteristic	**N**	**M [SD] or N (%)**	
Male, N (%)	122	61	(50.0)
Weeks' gestation, M [SD]	122	39.6	[1.4]
Birth weight (kg), M [SD]	122	3.5	[0.5]
Birth length (cm), M [SD]	120	50.9	[2.8]
Birth head circumference (cm), M [SD]	114	34.8	[1.5]
Apgar score 1 minute, M [SD]	115	8.49	[1.3]
Apgar score 5 minutes, M [SD]	115	9.0	[0.4]
NICU or special care nursery at birth, N (%)	122	12	(9.8)
Required oxygen at birth, N (%)	121	11	(9.1)
3yr weight (kg), M [SD]	115	15.2	[1.8]
3yr length (cm), M [SD]	115	96.9	[4.1]
3yr head circumference (cm), M [SD]	114	50.6	[1.8]
Chronic illness requiring medication, N (%)	122	13	(10.7)
Hospitalized overnight since birth, N (%)	122	16	(11.1)
Underwent surgery, N (%)	122	8	(6.6)
Developmental concern, N (%)	122	8	(6.6)
Only child, N (%)	122	29	(23.8)
Handedness, N (%)		Right	Left / Ambidextrous/unclear
	122	113 (92.6)	4 (3.3) / 5 (4.1)

5.4.2 Inter-rater Reliability

Inter-rater reliability of 36 BSID-III administrations was assessed via one-way random, absolute, single-measure intraclass correlation coefficients for each of the five objectively rated sub-domains of the BSID-III. The resulting ICC's fell in the excellent range; Cognitive ICC = 0.995, Receptive Language ICC = 0.999; Expressive Language ICC = 0.987; Fine Motor ICC = 0.996; Gross Motor = 0.939, indicating a high degree of agreement between coders (Landers, 2015; Shrout & Fleiss, 1979).

5.4.3 BSID-III Scores at 1-year and 3-years

The average age of BSID-III completion was 12-months and 36-months, respectively. As seen in Table 2, 1-year cognitive scores for the full cohort were, on average, three scaled score points higher than cognitive scores obtained at 3-years. The mean of the US normative sample at both 1-year and 3-years is 10 with a standard deviation of three. As such, our results indicate that the cognitive performance of infants in the current cohort differs significantly from US infants across the ages of 1-year and 3-years. As expected, female receptive and expressive language scores were significantly higher than males at both 1-year and 3-years of age.

Table 2. BSID-III scaled score means and standard deviations at 1-year and 3-years

	1-year			3-years		
BSID-III Domain	**Total, M [SD]**	**Girls, M [SD]**	**Boys, M [SD]**	**Total, M [SD]**	**Girls, M [SD]**	**Boys, M [SD]**
Cognitive	12.5 [2.3]	12.8 [2.4]	12.2 [2.3]	9.5 [1.6]	9.8 [1.7]	9.2 [1.4]
Receptive Language	10.9 [2.4]	11.4 [2.2]	10.5 [2.6]	10.9 [2.4]	11.3 [2.2]	10.6 [2.6]
Expressive Language	11.3 [2.1]	11.5 [2.2]	11.0 [2.0]	11.2 [2.3]	11.4 [1.7]	11.0 [2.7]
Fine Motor	11.3 [2.4]	11.5 [2.4]	11.0 [2.4]	11.5 [2.5]	12.1 [2.4]	11.0 [2.5]
Gross Motor	9.9 [3.4]	10.0 [3.7]	9.8 [3.0]	10.4 [2.2]	10.6 [2.1]	10.2 [2.3]

5.4.4 Predictive Utility of 1-year BSID-III Scores for 3-year BSID-III Outcomes

T-test results in Table 3 demonstrate that, with the exception of cognitive scores, mean scaled scores obtained by the full cohort as well as each sex, did not significantly differ across timepoints from 1-year to 3-years, suggesting consistency in scores across time. As previously discussed, cognitive scores did significantly differ between 1-year and 3-years, with scores obtained at 3-years significantly lower, on average, than scores obtained at 1-year.

Pearson correlation coefficients were used to assess the strength of the association between BSID-III scores at 1-year and 3-years across domains. Full cohort data revealed fair correlations between 1-year and 3-years on all domains of the BSID-III (Colton, 1974). Once broken down by sex, female performance at 1-year was correlated with more BSID-III domains at 3-years than males, with significant, fair strength, correlations detected across all but the expressive language domain (Colton, 1974). Male 1-year performance was only significantly correlated with 3-year performance on the cognitive and language domains, though correlations on the language domains were approaching moderate strength.

Table 3. Paired sample t-tests, correlations and effect size for the association between 1-year and 3-year BSID-III results

	Full Cohort (N=122)			**Female (N=61)**			**Male (N=61)**		
BSID-III domain	**t (*p*)**[1]	**M diff 95% CI**	**r (*p*)**	**t (*p*)**[1]	**M diff 95% CI**	**r (*p*)**	**t (*p*)**[1]	**M diff 95% CI**	**r (*p*)**
Cognition	-14.19 (<.*001*)	-3.45 – -2.60	0.32 (<.*001*)	-9.94 (<.*001)*	-3.62 – -2.41	0.34 (.*008*)	-10.04 (<.*001*)	-3.64 – -2.43	0.27 (.*036*)
Receptive language	0.44 (.*664*)[2]	-0.35 – 0.55[2]	0.44 (<.*001*)[2]	-0.39 (.*700*)	-0.81 – 0.48	0.43 (.*001*)	0.92 (.*363*)[3]	-0.387 – 1.01[4]	0.42 (.*001*)[4]
Expressive language	-0.44 (.*663*)	0.35 – -0.44	0.37 (<.*001*)	-0.37 (.*717*)	-0.74 – 0.52	0.23 (.*069*)	-0.25 (.*801*)	-0.73 – 0.57	0.47 (<.*001*)
Fine motor	1.06 (.*289*)	-0.25 – 0.82	0.26 (.*004*)	1.68 (.*098*)	-0.12 – 1.33	0.31 (.*015*)	-0.08 (.*935*)	-0.83 – 0.77	0.18 (.*160*)
Gross motor	0.89 (.*374*)[3]	-0.34 – 0.89[3]	0.27 (.*003*)[3]	1.18 (.*241*)[4]	-0.36 – 1.39[4]	0.39 (.*002*)[4]	0.08 (.*939*)[5]	-0.85 – 0.92[5]	0.12 (.*377*)[5]

[1]Note: paired t-test results are based on 3-year BSID-III scaled score minus 1-year BSID-III scaled scores; [2]N= 122 (1 outlier removed); [3]N=119 (3 outliers removed); [4]N= 60 (1 outlier removed); [5] N= 59 (2 outliers removed).
Note: 'M_{diff}' = Mean difference between the current cohort mean scaled score and the US BSID-III normative sample mean scaled score of 10.

Table 4 shows the regression results for 1-year BSID-III scores predicting 3-year BSID-III scores, after covariate adjustment. Performance across all domains of the BSID-III at 1-year was (statistically) significantly predictive of performance at 3-years. In terms of clinical significance, the receptive language domain had the greatest predictive power, with 1-year scores explaining 17.7 percent of variance in 3-year scores, once controlling for covariates. While other domains were statistically significant predictors of 3-year outcomes, the proportion of variance explained by 1-year scores for each domain ranged from 4.0 percent to 6.7 percent, suggesting limited clinical utility.

Following stratification by sex, male performance on the language domains at 1-year was predictive of performance at 3-years, after controlling for covariates. The language domains at 1-year appear to have predictive clinical utility for males, each predicting 18.3 percent of the variance in 3-year language scores. Similarly, female 1-year performance on receptive language, fine motor and gross motor domains was again predictive of performance at 3-years, after controlling for covariates. Scores obtained on these domains at 1-year appear to hold clinical value, explaining 11.3 percent to 17.4 percent of the variance in 3-year scores. In contrast, cognitive scores at 1-year were not clinically predictive of cognitive scores at 3-years for females nor males after covariate adjustment.

Table 4. Linear regression results comparing the 1-year and 3-year performance on the BSID-III, controlling for covariates.

	Full Cohort (N=122)[1]				**Females (N=61)[1]**				**Male (N=61)[1]**			
BSID-III Domain	**F**	**df**	***p***	**η_p^2**	**F**	**df**	***p***	**η_p^2**	**F**	**df**	***p***	**η_p^2**
Cognitive	6.42	1,112	*.013*	.054	2.88	1,51	*.096*	.054	1.64	1,51	*.206*	.031
Receptive language	24.14	1,112	*<.001*	.177	11.14	1,53	*.002*	.174	11.40	1,51	*.001*	.183
Expressive language	5.26	1,113	*.024*	.044	1.41	1,52	*.241*	.026	11.41	1,51	*.001*	.183
Fine motor	8.14	1,113	*.005*	.067	6.64	1,52	*.013*	.113	0.74	1,53	*.394*	.014
Gross motor[2]	4.76	1,114	*.031*	.040	10.10	1,52	*.003*	.163	0.02	1,53	*.882*	<.001

Note: [1]N included in individual analyses varies dependent on outliers removed. Please see 'statistical analyses' section for further details; [2]Log10 transformed data presented

5.5 Discussion

The BSID-III is widely used in clinical settings to inform the likelihood of a child experiencing future difficulties. Despite this, no research to date is available examining the predictive utility of all domains of the BSID-III in a general population cohort of infants. Moreover, the unique predictive utility of the BSID-III for male and female children is unknown. This study used a longitudinal, prospective cohort of study of 122 children, who completed the BSID-III at 1-year and 3-years, to examine the statistical and clinical predictive significance of the BSID-III in a general population cohort of children, and stratified by sex.

Full cohort results (unadjusted and adjusted) indicated that 1-year BSID-III scores were statistically predictive of 3-year BSID-III scores across all five domains. Clinically, the receptive language domain appeared to hold the greatest utility with a fair correlation co-efficient, and 1-year receptive language scores explained approximately 18 percent of the variance in 3-year receptive language scores after adjustment. Considering the variability in developmental milestone attainment at 1-year (Anderson & Burnett, 2017), this result provides support for the clinical use of the BSID-III receptive language domain at 1-year to inform predictive prognoses. In contrast, although full cohort correlation coefficients ranged from 0.26 to 0.37 on cognitive, expressive language, fine motor and gross motor domains, 1-year performance only predicted approximately four to seven percent of variation in performance at 3-years after adjustment for: maternal age at birth; linguistic background; education; socio-economic status; parity and infant hospitalisations since birth. Given the high resource demand required to administer and interpret a BSID-III assessment clinically and that the proportion of additionally explained variance is little greater than that explained by solely considering socio-demographic factors, the current full

cohort results do not support the use of the BSID-III at 1-year as a predictive assessment tool for these domains.

Examination of the predictive utility of the BSID-III by sex showed that after control for covariates, performance on the 1-year receptive language, fine motor and gross motor domains was both statistically and clinically predictive of 3-year female performance, with fair correlation coefficients, and explained variance after covariate adjustment of 11 to 17 percent. Comparatively, male 1-year performance was statistically and clinically predictive of 3-year performance on the receptive and expressive language domains, explaining 18 percent of variation in 3-year performance. The fair correlation coefficients detected in the current study once stratified by sex are reasonably consistent with predictive utility data on other developmental assessment tools once stratified by sex, such as the Griffiths Scale of Infant Development (Smith, Bidder, Gardner and Gray, 1980).

Interestingly, after adjustment for covariates, 1-year cognitive performance of the current cohort was neither statistically nor clinically predictive of performance at 3-years for either sex. While this result appears somewhat inconsistent with Greene et al. (2013) and Bode et al. (2014), which reported moderate strength full cohort correlation coefficients on the cognitive domain, studies were predominantly conducted with at risk children. Predictive utility of the BSID-II was reportedly stronger predictive utility for at risk children than typically developing children, thus a similar phenomenon may be true for the BSID-III. Further, prior to stratification by sex the current full cohort 1-year cognitive scores were also statistically significantly predictive of 3-year scores. However, the proportion of explained variance was low suggesting limited clinical predictive utility after adjusting for socio-demographic factors. Results highlight the need for

future research adjusting for covariates and reporting the subsequent explained variance of each BSID-III domain in order to accurately assess the clinical predictive utility of the tool.

Our results also show that expressive language scores at 1-year for females and gross motor scores at 1-year for males did *not* predict performance at 3-years in the same domain. One possible explanation for the poor predictive power of the BSID-III on these domains might be the extent of growth during this period for male and female children (Lung et al 2009; Galsworthy et al 2000; Malina 2004). Whether as a result of genetic, biological or societal influences, females of toddler to pre-schooler age are reported to have stronger language abilities, and males of this same age are reported to have stronger gross motor abilities (Lung et al., 2009; Galsworthy, Dionne, Dale, & Plomin, 2000; Malina, 2004). The differing levels of growth between 1-year and 3-years in these areas for each sex may explain the poor predictive power of the BSID-III during this period. This result was partially consistent with a previous study examining the BSID-II by sex, which reported female performance at 6-months on the Mental Developmental Index (MDI) of the BSID-II was not predictive of performance at 36-months due to rapid language development over this period (Lung et al. 2009). Although this study reported male development to be more stable and, therefore, more predictive from 6- to 36-months, the BSID-II only included MDI and Psychomotor Development Index (PDI) domains, rather than the five domains included in the BSID-III, and therefore gross motor performance was not assessed separately.

5.5.1 Strengths and Limitations

This study uniquely examined the predictive utility of the BSID-III in a general population cohort of infants stratified by sex. However, some important limitations should be

noted. First, the cohort is of relatively high socio-economic status, as evidenced by high proportion of mothers working, having completed a bachelor or college degree, and SEIFA scores. Although this may have impacted results given the known association between maternal ability, socio-economic factors and offspring ability, TOPF scores suggested intellectual ability was consistent with population expectations (Wechsler, 2011). Moreover, we adjusted for maternal education and socio-economic status in regression models with subsequent results remaining relatively unchanged.

Second, whilst the cohort size of the current study is commensurate with research in the field, and is greater than any one of the BSID-III US normative groups, a larger cohort size would have enhanced power. Whilst the current study included covariates commonly associated with development, due to power constraints, not all potential covariates could be included; for example, maternal mental health (Rossen et al., 2016) and infant birth data, such as admission to NICU or special care (see *Study 2*). The current results may therefore over-estimate the associations between 1-year and 3-year performance.

Third, sensitivity and specificity could not be assessed due to the small proportion of children in the current cohort with suspected developmental difficulties. As such, the results cannot be used to inform the predictive utility of the BSID-III in specific clinical sub-populations. Future research on specific sub-populations with different developmental trajectories is therefore recommended, such as those with chromosomal conditions.

Fourth, as part of the Triple B Pregnancy Cohort Study, parents were provided a developmental report following the 1-year assessment. As such, children performing below expectations may have been detected early, increasing the likelihood of intervention. Given the potential for early intervention to ameliorate delays, the stability in scores may have been

reduced somewhat between 1-year and 3-years of age. Nevertheless, most parents did indicate that their child's performance was commensurate with their expectations. As such, it is possible that parents who sought further assessment or intervention, would have done so irrespective of the study report.

Fifth, we did not examine development in later childhood, where ability may stabilise. Further longitudinal follow-up to school age is recommended to determine the-long term predictive utility of the BSID-III.

Sixth, a correlation co-efficient of >0.3 and a minimum of 10 percent of explained variance following adjustment, was selected apriori as a requirement to indicate clinical predictive significance. While this is expected to be of clinical significance in some settings, and is thought be a realistic expectation for a developmental assessment tool at 1-year of age given the variability in development in this phase (Anderson & Burnett, 2017), some clinical settings may require a higher percentage of explained variance. Thus, clinicians should use judgment to interpret the results of the study with respect to their individual setting.

5.5.2 Conclusions

The results of the current study suggest that the BSID-III has some predictive utility across all domains, with fair correlation coefficients detected. Notably, the receptive language domain had the greatest predictive utility; 1-year performance predicted up to 18 percent of variation in 3-year receptive language. However, after control for covariates, the predictive utility of the BSID-III was shown to improve when stratified by sex. Receptive language, fine motor and gross motor domains for females, and the language domains for males held clinical and statistical predictive utility. Results indicate that infants experiencing delays on these

domains at 1-year of age are more likely to continue to experience delays at 3-years of age, and thus warrant consideration for early intervention. Furthermore, results highlight the importance of sex as a moderator of infant development and thus the importance of stratification by sex in future developmental predictive research. Knowledge of the different predictive utility of each BSID-III domain, and the predictive utility by sex, will assist clinicians in determining its value as a tool for the provision of early intervention. Further research is required examining the predictive utility of the BSID-III stratified by sex in at risk cohorts.

5.6 References

Albers, C. A., & Grieve, A. J. (2007). Test review: Bayley, N.(2006). Bayley scales of infant and toddler development–third edition. San Antonio, TX: Harcourt assessment. *Journal of Psychoeducational Assessment*, *25*(2), 180-190.

Anderson, P. J., & Burnett, A. (2017). Assessing developmental delay in early childhood—concerns with the Bayley-III scales. *The Clinical Neuropsychologist*, *31*(2), 371-381. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1080/13854046.2016.1216518

Australian Bureau of Statistics (2013). 2033.0.55.001 - Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia.

Aylward, G. P. (2002). Cognitive and neuropsychological outcomes: more than IQ scores. *Mental retardation and developmental disabilities research reviews*, *8*(4), 234-240.

Baron-Cohen, S., Knickmeyer, R. C., & Belmonte, M. K. (2005). Sex differences in the brain: implications for explaining autism. *Science*, *310*(5749), 819-823.

Bayley, N. (2006a) *The Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III).* 3rd ed. San Antonio, Texas: Pearson Psychcorp.

Bayley, N. (2006b). *Bayley Scales of Infant and Toddler Development, Third Edition Technical Manual.* San Antonio, TX: Pearson PsychCorp.

Bode, M. M., D'Eugenio, D. B., Mettelman, B. B., & Gross, S. J. (2014). Predictive validity of the Bayley, at 2 years for intelligence quotient at 4 years in preterm infants. *Journal of Developmental & Behavioral Pediatrics*, *35*(9), 570-575.

Colton, T. (1974). *Statistics in Medicine*. Boston: Little, Brown and Company

Crowe, T. K., Deitz, J. C., & Bennett, F. C. (1987). The relationship between the Bayley Scales of Infant Development and preschool gross motor and cognitive performance. *American Journal of occupational therapy, 41*(6), 374-378.

Forns, J., Julvez, J., García-Esteban, R., Guxens, M., Ferrer, M., Grellier, J., et al. (2012). Maternal intelligence-mental health and child neuropsychological development at age 14 months. *Gaceta Sanitaria, 26*(5), 397-404.

Galsworthy, M., Dionne, G., Dale, P. & Plomin, R. (2000) Sex differences in early verbal and non-verbal cognitive development. *Developmental Science,* 3, 206–215.

Golombok, S., Rust, J., Zervoulis, K., Croudace, T., Golding, J., & Hines, M. (2008). Developmental Trajectories of Sex-Typed Behavior in Boys and Girls: A Longitudinal General Population Study of Children Aged 2.5–8 Years. *Child development, 79*(5), 1583-1593.

Greene, M. M., Patra, K., Silvestri, J. M., & Nelson, M. N. (2013). Re-evaluating preterm infants with the Bayley-III: Patterns and predictors of change. *Research in Developmental Disabilities, 34*(7), 2107-2117. doi: https://doi.org/10.1016/j.ridd.2013.04.001

Hack, M., Taylor, H. G., Drotar, D., Schluchter, M., Cartar, L., Wilson-Costello, D., et al. (2005). Poor predictive validity of the Bayley Scales of Infant Development for cognitive function of extremely low birth weight children at school age. *Pediatrics, 116*(2), 333-341.

Harris, S. R., Megens, A. M., Backman, C. L., & Hayes, V. E. (2005). Stability of the Bayley II Scales of Infant Development in a sample of low-risk and high-risk infants. *Developmental Medicine & Child Neurology, 47*(12), 820-823.

Hutchinson, D., Wilson, J., Allsop, S., Elliott, E., Najman, J., Burns, L., et al. (*in press*). Cohort profile: The Triple B Pregnancy Cohort Study: A longitudinal study of the relationship between alcohol, tobacco and other substance use during pregnancy and the health and wellbeing of Australian children and families. *International Journal of Epidemiology*.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. American Psychological Association.

IBM Corp. (2013) *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: IBM Corp.

Janssen, A. J., Akkermans, R. P., Steiner, K., de Haes, O. A., Oostendorp, R. A., Kollée, L. A., et al. (2011). Unstable longitudinal motor performance in preterm infants from 6 to 24 months on the Bayley Scales of Infant Development—Second edition. *Research in developmental disabilities, 32*(5), 1902-1909.

Landers, R. N. (2015). Computing intraclass correlations (ICC) as estimates of interraterreliability in SPSS. *The Winnowner 2*. DOI: 10.15200/winn.143518.81744.

Lobo, M. A., Paul, D. A., Mackley, A., Maher, J., & Galloway, J. C. (2014). Instability of delay classification and determination of early intervention eligibility in the first two years of life. *Research in developmental disabilities, 35*(1), 117-126. doi: https://doi.org/10.1016/j.ridd.2013.10.017

Lung, F. W., Shu, B. C., Chiang, T. L., Chen, P. F., & Lin, L. L. (2009). Predictive validity of Bayley scale in language development of children at 6–36 months. *Pediatrics International, 51*(5), 666-669.

Luttikhuizen dos Santos, E. S. L., de Kieviet, J. F., Königs, M., van Elburg, R. M., & Oosterlaan, J. (2013). Predictive value of the Bayley scales of infant development on development of very preterm/very low birth weight children: a meta-analysis. *Early human development, 89*(7), 487-496.

Malina, R. M. (2004). Motor development during infancy and early childhood: Overview and suggested directions for research. *International Journal of Sport and Health Science*, *2*, 50-66.

McPhie, S., Skouteris, H., Mattick, R. P., Wilson, J., Honan, I., Allsop, S., et al. (2017). Weight in the First Year of Life: Associations with Maternal Prepregnancy Body Mass Index and Gestational Weight Gain—Findings from a Longitudinal Pregnancy Cohort. *American Journal of Perinatology*.

Niccols, A., & Latchman, A. (2002). Stability of the Bayley mental scale of infant development with high risk infants. *The British Journal of Development Disabilities, 48*(94), 3-13.

Piek, J. P., Dawson, L., Smith, L. M., & Gasson, N. (2008). The role of early fine and gross motor development on later motor and cognitive ability. *Human movement science*, *27*(5), 668-681.

Potharst, E. S., Houtzager, B. A., Van Sonderen, L., Tamminga, P., Kok, J. H., Last, B. F., et al. (2012). Prediction of cognitive abilities at the age of 5 years using developmental follow-up assessments at the age of 2 and 3 years in very preterm children. *Developmental Medicine & Child Neurology, 54*(3), 240-246.

Reznick, J. S., Corley, R., Robinson, J., & Matheny Jr, A. P. (1997). A longitudinal twin study of intelligence in the second year. *Monographs of the society for research in child development*, i-160.

Roberts, G., Anderson, P. J., Doyle, L. W., & Victorian Infant Collaborative Study Group. (2010). The stability of the diagnosis of developmental disability between ages 2 and 8 in a geographic cohort of very preterm children born in 1997. *Archives of Disease in Childhood*, *95*(10), 786-790.

Rossen, L., Hutchinson, D., Wilson, J., Burns, L., Olsson, C. A., Allsop, S., et al. (2016). Predictors of postnatal mother-infant bonding: the role of antenatal bonding, maternal substance use and mental health. *Archives of women's mental health*, *19*(4), 609-622.

Sajaniemi, N., Hakamies-Blomqvist, L., Katainen, S., & von Wendt, L. (2001). Early cognitive and behavioral predictors of later performance: a follow-up study of ELBW children from ages 2 to 4. *Early Childhood Research Quarterly, 16*(3), 343-361.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420.

Smith, J. A., Bidder, R. T., Gardner, S. M., & Gray, O. P. (1980). Griffiths Scales of Mental Development and different users. *Child: Care, Health and Development, 6*(1), 11-16.

Spencer-Smith, M. M., Spittle, A. J., Lee, K. J., Doyle, L. W., & Anderson, P. J. (2015). Bayley-III cognitive and language scales in preterm children. *Pediatrics*, *135*(5), e1258-e1265. doi: 10.1542/peds.2014-3039

Spittle, A. J., Spencer-Smith, M. M., Eeles, A. L., Lee, K. J., Lorefice, L. E., Anderson, P. J., et al. (2013). Does the Bayley-III Motor Scale at 2 years predict motor outcome at 4 years in very preterm children? *Developmental Medicine & Child Neurology, 55*(5), 448-452.

SPSS Web Books Regression With SPSS Chapter 2 – Regression Diagnostics. UCLA: Statistical Consulting Group. http://stats.idre.ucla.edu/spss/webbooks/reg/chapter2/spss-webbooksregressionwith-spsschapter-2-regression-diagnostics/#. Accessed 12th January 2017.

Tabachnick, B. G., Fidell, L. S., & Osterlind, S. J. (2001). Using multivariate statistics.

Wechsler, D. (2011). Test of premorbid functioning. UK version (TOPF UK). *UK: Pearson Corporation*.

6. General Discussion.

The four studies presented in this thesis explored the clinical and predictive utility of the BSID-III for Australian 1-year-old and 3-year-old children. Study 1 compared the performance of 1-year-old Australian infants on the BSID-III to the US normative sample. Study 2 investigated the utility of the BSID-III to detect differences in infant performance based on indirect perinatal risk factors for neurodevelopmental delay. Study 3 compared the performance of 3-year-old Australian children on the BSID-III to the US normative sample; and Study 4 examined the predictive utility of the BSID-III from 1-year to 3-years of age, stratified by sex.

Taken together, the results of this thesis converge to suggest that current methods of interpreting child performance in Australia, using US BSID-III normative data, may be sub-optimal and result in both under-identification and over-identification of developmental delays on domains of the BSID-III at both 1- and 3-years. Specifically, the results suggest that application of the US BSID-III normative data for Australian children may result in under-detection of both cognitive delays in 1-year-old Australian infants, and motor and language delays in 3-year old Australian children; reducing potential for the delivery of early intervention. Moreover, motor delays in 1-year-old Australian infants and cognitive delays in 3-year-old Australian children may be over-detected, unnecessarily increasing demands on early intervention services.

Below, the key findings of each study are discussed in turn, with a particular emphasis on how findings relate to and extend on the existing literature, as well as the implications for clinical practice. Limitations and directions for future research are also discussed.

6.1 Study 1: A Comparison Between Australian Infant Performance and US Normative Data at 1-year on the BSID-III.

This study compared the performance of 1-year-old Australian infants on the BSID-III to the US normative sample in a cohort of 998 term born, singleton infants. The study also examined maternal socio-demographics, with comparisons drawn between the current cohort under study and Australian and US population data. A key strength of the study was the use of statistical adjustment to improve cohort representativeness to the Australian population. Application of raking improved the overall representativeness so that the study cohort more closely matched the Australian population on a range of socio-demographics (child sex, maternal age at birth, SEIFA quartiles, maternal region of birth, maternal education, maternal employment status and Aboriginal and Torres Strait Islander identification), improving our confidence that the results reflect true population differences (rather than being artefacts of the cohort).

Study 1 found that Australian infants scored significantly *higher* than the US normative sample on the cognitive domain of the BSID-III, and significantly *lower* than the US normative sample on the gross motor domain, across both the 12- and 13-month age groups. The weighted study cohort also scored significantly higher than the US normative sample on the expressive language and fine motor domains at 12-months, and significantly lower than the US BSID-III normative sample on the receptive language domain at 13-months. However, results on the receptive language, expressive language and fine motor domains were not consistent across both the 12- and 13-month age groups and effect sizes were small, suggesting that the statistically significant results on these outcomes domains may have been a bi-product of the large cohort size, rather than holding clinical significance.

The differences found on the cognitive and gross motor domains in Study 1 are consistent with the limited extant research suggesting that application of US BSID-III normative data internationally may be inappropriate and lead to misclassifications (Anderson, De Luca, Hutchinson, Roberts, & Doyle, 2010; Chinta, Walker, Halliday, Loughran-Fowlds, & Badawi, 2014; Krogh, Væver, Harder, & Køppe, 2012; Steenis, Verhoeven, Hessen, & Van Baar, 2015; Walker, Badawi, Halliday, & Laing, 2010; Westera, Houtzager, Overdiek, & Van Wassenaer, 2008). Specifically, the results of this thesis indicated that children with cognitive delay may be more likely to be classified in the non-delayed category and children with non-delayed gross motor abilities may be more likely to be classified in the delayed category. This finding is consistent with Walker et al. (2010) who reported significantly higher scores on the cognitive domain and significantly lower scores on the gross motor domain for Australian 1-year-old infants when compared to the US normative sample. However, unlike Walker et al. (2010) our results did not find consistently significant differences between US and Australian infant scores at 1-year of age on the receptive and expressive language domains.

Given the use of raking to improve the representativeness of the current cohort relative to the Australian population, it is likely that population differences between Australia and the US account for the differences detected in Study 1. Examination of demographic data in the current study suggested that Australian mothers were more likely to have completed formal education post-school than US mothers, and that the regions of birth of mothers differed between Australia and the US (i.e., higher proportions of people of Asian ethnicities, and lower proportions of people of Hispanic and African American ethnicities in Australia). Past research suggests that parental education is associated with offspring cognitive ability (Forns et al., 2012), and that infant/child developmental trajectories differ based on culture/ethnicity (Greene, Patra, Nelson,

& Silvestri, 2012; Greene, Patra, Silvestri, & Nelson, 2013; Halle et al., 2009). As such, population differences between Australia and the US with regards to maternal education and ethnicity may explain the detected differences, and highlight the need for country specific normative data.

Although mothers from the lowest socio-economic regions in Australia were somewhat under-represented in this study, the inclusion of detailed maternal demographic data is unique. Moreover, the steps taken to improve cohort representativeness suggest that the differences detected are unlikely to result from sampling biases. Rather, they reflect true differences in 1-year-old infant ability between Australian and US populations. As such, the means and standard deviations obtained in this study may be used to calculate z-scores, providing normative data for singleton, term-born 12- and 13-month-old Australian infants on the BSID-III.

6.2 Study 2: Utility of the BSID-III to Distinguish 1-year Infants at Perinatal Risk of Neurodevelopmental Delay.

Study 2 examined the clinical utility of the BSID-III to detect differences in infant performance between those with indirect perinatal risk factors associated with NDD and those without in a cohort of 935 1-year-old infants. The majority of research conducted to date has addressed children older than 18-months of age, in singular risk factor sub-populations (e.g., extremely low birth weight, prematurity etc) (De Jesus et al., 2013; Jarjour, 2015; Skiöld et al., 2012), which limits generalisability of results to real world populations where risk factors are co-occurring. Study 2 addressed these limitations by examining indirect perinatal risk of NDD in an at-risk group derived from the general population which comprised of: prematurity (Bos & Roze, 2011; Greene et al., 2012), low birth weight (de Moura et al., 2010), small head circumference

(Peterson, Taylor, Minich, Klein, & Hack, 2006), admission to neonatal intensive care unit (NICU) and/or special care units (SCU) post-birth (Walker, Holland, Halliday, & Badawi, 2012), low Apgar scores (Odd, Rasmussen, Gunnell, Lewis, & Whitelaw, 2008), maternal substance use (Bandstra, Morrow, Mansoor, & Accornero, 2010; Huizink, 2014), and multiple birth infants (Wadhawan et al., 2011).

A further contribution of Study 2 was the inclusion of a cohort where the youngest mean age was 18-months; this is lower than the mean age examined in the majority of previous research in Australia. Study 2 also accounted for the role of environmental, biological and medical factors in the analyses, improving on the methodology of some previous studies, which typically did not account for environmental factors known to be important correlates of infant developmental outcomes.

Study 2 results suggested that the BSID-III holds limited clinical utility in detecting differences in performance between infants at indirect perinatal risk and those at low-risk, after accounting for: infant sex; hospitalisations since birth; maternal age at birth; linguistic background; employment status; educational attainment; socio-economic status and parity. Specifically, whilst infants at indirect perinatal risk performed lower than infants at low-risk on the cognitive and gross motor domains, after accounting for covariates, group scaled score differences were small: 0.57 and 0.44 respectively, and risk status explained only 1.2 percent and 0.5 percent of variance in performance, respectively. Considering a standard deviation based on the US normative sample is 3 scaled score points (Bayley, 2006b), half a scaled score point difference between groups is unlikely to have clinical utility.

The lack of significant differences detected on the expressive language and fine motor domains in Study 2 is also of import. Specifically, the results suggest that the BSID-III is of

limited value for the detection of delays on these domains based on indirect perinatal risk status at 1-year of age. This may reflect a lack of sensitivity of the BSID-III to subtle delay, leading to potential under-detection of delay. Conversely, it is also possible that delays associated with indirect risk factors only impact on higher order skills, developed later in childhood. A meta-analysis examining preterm infants found that language delays become more pronounced with increasing age in childhood. Moreover, a study comparing 95 late preterm infants with term infants found no significant difference in motor dexterity at 3-years of age (Baron et al., 2009b), whereas a study by the same author found that preterm/extremely low birth weight infants without intraventricular haemorrhage at 6-years of age performed poorer on motor dexterity tasks than non-preterm children (Baron et al., 2009a). Inconsistency in the literature suggests that determining whether indirect perinatal risk is associated with early expressive language and fine motor delays, and the timing of the onset of those delays, are important areas for continued research.

6.3 Study 3: A comparison between Australian and US normative data at 3-years of age on the BSID-III

Study 3 investigated whether Australian BSID-III performance differed from the US normative sample in a cohort of 119 3-year-old children. Study 3 builds on the extant literature by examining the role of socio-demographic factors in the detected differences between Australian and US child performance on the BSID-III at 3-years of age. A strength of Study 3 was the inclusion of detailed maternal socio-demographics, including an estimate of maternal intelligence (i.e., TOPF), and the application of statistical methods to improve overall cohort representativeness. This study represents an important contribution to the Australian literature, as

3-years of age is a standard follow-up age for many developmental follow-up clinics in Australia. Furthermore, developmental assessments commonly occur around the age of 3-years for the diagnosis of conditions such as Autistic Spectrum Disorder and language delays. Thus, knowledge of the clinical utility of the BSID-III at 3-years of age is especially important for clinical assessment and intervention.

Following the application of raked weights, Australian 3-year-old children performed significantly higher on the language (receptive and expressive) and motor (fine and gross) domains, and significantly lower on the cognitive domain when compared to the US normative sample. The findings of higher performance of the language domains (receptive and expressive) and on the fine motor domain when compared to the US BSID-III normative data are consistent with Chinta et al. (2014). In contrast, however, 3-year-old children in the current study also scored higher than the US normative sample on the gross motor domain. Further, unlike Chinta et al. (2014) which reported higher scores obtained by Australian children on the cognitive domain, the current study found significantly lower scores obtained by Australian children on the cognitive domain when compared to 3-year US BSID-III normative data. Importantly, cognitive scores reported in Chinta et al. (2014) would not have met significance criteria had Bonferroni adjustment for multiple analyses been applied. Moreover, gross motor scores in the current cohort did not meet Bonferroni adjusted significance levels prior to the application of raked weights. This contrast in results further highlights the important contribution of the current study in taking steps to adjust for cohort representativeness.

The results of Study 3 have important clinical implications with respect to the application of US normative data to the Australian context. Namely, that 3-year-old children with mild language and motor delays may go undetected and not receive intervention and 3-year-old

children with cognitive delays are likely to be erroneously over-diagnosed and unnecessary intervention recommended.

Taken together, results from Study 3 further support findings from Study 1, as well as findings in the literature suggesting that utilisation of US normative data internationally may result in inaccurate interpretation of child performance, and that the development of Australian BSID-III normative data is warranted.

6.4 Study 4: Predictive Utility of the BSID-III from 1-year to 3-years of Age.

The fourth and final study of this thesis aimed to examine the predictive utility of the BSID-III from 1- to 3-years of age in a longitudinal cohort study of 122 children after adjustment for covariates, and stratification by sex. Although variability in developmental milestone attainment in infancy is well documented, with large gains achievable over short timeframes (Brownell & Kopp, 2010; Charman et al., 2005), predictive utility is an important psychometric property of developmental assessment tools. Developmental assessment are typically used for the purposes informing early intervention decision making, aimed at reducing the likelihood of future delay. Despite this, predictive utility of the BSID-III was not assessed during the development of the third edition. Moreover, prior to this thesis, no study to the author's knowledge had examined the predictive utility of the five domains of the BSID-III in a general population cohort. Rather, research examining score stability and predictive validity has predominantly been conducted on specific BSID-III domains and predominantly in preterm cohorts. For example, Spittle et al. (2013) examined the predictive validity of the motor domain at 2-years of age on motor outcomes at 4-years of age in very preterm children, and Bode et al. (2014) examined the predictive validity of the cognitive and language domains at 2-years of age

on WPPSI-III outcomes at 4-years of age in preterm and healthy control children (Bode, D'Eugenio, Mettelman, & Gross, 2014). As such, Study 4 aimed to address this gap. Key background covariates including: maternal age at birth; maternal linguistic background; maternal education; socio-economic status; parity and infant hospitalisations since birth, an additional strength of Study 4.

Another unique contribution of Study 4 was the stratification of predictive utility by sex. Earlier research suggests that developmental trajectories differ based on sex in infancy, toddlerhood and the pre-school years, with language typically developing earlier for females, and gross motor typically developing earlier for males. As such, sex may act as a moderator and dilute full cohort estimates of predictive utility. Previous versions of the BSID have demonstrated differing predictive utility of the tool by sex (Lung, Shu, Chiang, Chen, & Lin, 2009), however this has not been examined in the most recent version of the BSID. Therefore, Study 4 aimed to address this gap in order to provide a more accurate understanding of the predictive utility of the tool across these age groups.

Study 4 found that overall BSID-III performance at 1-year of age significantly predicted BSID-III performance at 3-years of age on all outcome domains. However, cognitive, expressive language and motor domains (fine and gross) held little clinical utility, with scores at 1-year on these domains only explaining four to seven percent of variance in scores at 3-years. Receptive language, conversely, appeared to hold greatest clinical utility with 1-year scores explaining 17.7 percent of variance in 3-year scores once controlling for covariates. This highlights the importance of referring children with poor receptive language skills on the BSID-III at 1-year of age for intervention.

When stratified by sex, and after covariate adjustment, female performance on receptive language, fine motor and gross motor domains at 1-year were predictive of performance at 3-years. Moreover, male 1-year performance on the receptive and expressive language domains was predictive of 3-year performance. Once stratified by sex, 1-year scores on these domains explained between 11 and 18 percent of variance in 3-year scores. This highlights the importance of considering sex as a moderator variable in predictive research. Previous research examining the predictive utility of specific domains of the BSID-III, without stratification by sex, may therefore under-estimate the predictive utility of the BSID-III. Moreover, results highlight the clinical importance of referring females with receptive language, fine motor or gross motor delays and males with receptive or expressive language delays on the BSID-III at 1-year of age, to early intervention.

6.5 Limitations

The results of this thesis address significant gaps in knowledge of the clinical and predictive utility of the BSID-III in an Australian population. Although there are many strengths of this thesis, some limitations are noted. First, the cohort (at both ages) included a higher proportion of socio-economically advantaged mothers than observed in the general Australian population. Considering the known association between socio-economic factors and offspring ability, this may have impacted results such that the current cohort may have scored higher on the BSID-III cognitive and language domains than the general population. Nevertheless, the cohort did include participants from a wide range of socio-economic backgrounds, as evident by the spread of SEIFA scores. Moreover, steps were taken in each study to account for this sampling bias, through the application of raked weights and covariate adjustment. It was also

notable that despite the sampling bias, maternal TOPF (verbal IQ) scores (Wechsler, 2011) assessed in Studies 3 and 4 do not suggest elevated global levels of intellectual functioning of mothers in the cohort at 3-years of age.

Another limitation was that the Triple B Pregnancy Study recruited women through antenatal clinics at public hospitals in NSW. While women with substance use history, multiple birth infants and children born prematurely were included in the broader study, women known to be having children with likely developmental difficulties (e.g., children with chromosomal conditions) did not meet study inclusion criteria. Furthermore, in Studies 1 and 3, preterm children, those recruited through specialist antenatal clinics dedicated to people with complex substance use/mental health needs, and multiple birth children were excluded, consistent with the normative sampling criteria used in the BSID-III (Bayley, 2006b). Yet, unlike the US BSID-III normative sample, 10 percent of children with known developmental conditions or risk factors for delay were not reintroduced into the cohort in the current studies. This decision was driven by two main factors: (1) 10 percent re-inclusion appears excessive considering population estimates for disability in children aged 0 to 4 years in Australia are approximately 3 to 4 percent; (2) reintroduction methods did not appear consistent across all age groups of the BSID-III US normative data, and the proportion reintroduced in the age bands relevant to this thesis are not deducible based on the BSID-III administration nor technical manual (Bayley, 2006b). Nevertheless, some caution should be used when interpreting results where the current cohort performed higher than the US normative sample. Similarly, while Studies 2 and 4 were more inclusive of: preterm children; those with antenatal substance exposure; and, multiple birth infants (Study 2 only), it is recommended that future research includes more diverse cohorts of children with a range of developmental conditions.

Thirdly, a subsample was recruited for Studies 3 and 4 as it was not feasible to undertake a 3-year follow-up with all families from the larger, longitudinal study within the candidature time-period. Although the cohort size was commensurate with the US BSID-III normative data sample size of 100 children per age band (Bayley, 2006b), replication in a larger cohort of 3-year-old children would be of value.

Finally, consistent with ethical requirements, all families were offered a developmental assessment report with information relating to child outcomes on the BSID-III at 1-year of age. This may have impacted the results of Studies 3 and 4, as children who were performing below expectations may have been detected earlier than would be expected in the general population. Considering the known benefits of early intervention, this may have impacted the developmental trajectory of some infants (i.e., Study 4 may provide a conservative estimate of the BSID-III's predictive utility). Nevertheless, it should be noted that, anecdotally, child results on the BSID-III were commensurate with parent expectations, and, therefore, parents who did seek intervention may have done so regardless.

6.6 Clinical Implications and Future Directions

The results of the current thesis have important implications for clinical practice and future research. These are discussed in turn.

6.6.1 Importance of Australian Normative Data

Studies 1 and 2 examined the BSID-III in 1-year old infants. The cognitive and gross motor domains were identified in both these studies as important to furthering our understanding of the clinical utility of the BSID-III in Australia. Australian children performed significantly

differently to the US BSID-III normative data on these domains, most likely due to differences in maternal educational attainment and ethnicity across Australian and US populations (Study 1). These domains also appeared the most sensitive to detecting differences in infant performance between infants *at risk* and infants at *low risk* based on indirect perinatal risk factors (albeit only statistically rather than clinically) (Study 2). As such, the cognitive and motor domains appear to be the most clinically useful at this age, although application of US normative data remains problematic. To improve clinical utility of the BSID-III in this age group, the development of Australian normative data is essential. Without Australian normative data, 1-year old Australian children with cognitive delays may go undetected and, therefore, may not receive early intervention with known preventative and ameliorative potential. Furthermore, current use of US BSID-III normative data for 1-year old Australian children may result in over-detection of motor delays. Children may therefore be unnecessarily referred for early motor intervention, using limited resources and causing unnecessary stress for families.

Studies 3 and 4 provide important data relating to the clinical utility of the BSID-III in Australian 3-year-old children. Results from Study 3 again suggest that the clinical utility of the BSID-III in Australia would be enhanced with the development of Australian normative data. Currently, results suggest that use of US BSID-III normative data in 3-year-old Australian children will result in under-identification of language and motor delays. Given 3-years of age is an important window for early diagnosis, the results of this thesis suggest that some children may not be appropriately diagnosed, limiting potential access to funding, planning for school and access to early intervention. Moreover, cognitive delays may be being over-identified in children who, as a result, are unnecessarily referred for early intervention.

The change in cognitive and gross motor performance from 1- to 3-years in the current cohort relative to the US normative sample was unexpected (Studies 1 and 3). Australian children performed higher than the US normative sample on the cognitive domain at 1-year and lower than the US BSID-III normative sample at 3-years. Conversely, Australian children performed lower than the US normative sample on the gross motor domain at 1-year and higher than the US normative sample at 3-years. One possible explanation for this may be that the BSID-III over estimates ability. International research, such as that conducted by Anderson et al. (2010) examining 211 premature and/or low birth weight and 202 full-term, normal birth 2-year-old Australian children, suggests that the BSID-III overestimates ability (Anderson et al., 2010). While this may provide an explanation for the high scores obtained on the cognitive domain at 1-year and on the motor and language domains at 3-years, it does do not account for the low scores obtained on the gross motor domain at 1 year, nor for the change in scores from 1- to 3-years. Instead, two possible explanations may be considered to account for this. First, test item order may not be appropriate in an Australian culture. Furthermore, anecdotally, at 1-year of age many children in the current cohort who did not yet pull to standing or did not bounce were able to cruise around furniture once placed in a standing position. However, bouncing and cruising was required to meet basal criteria at 1-year for the gross motor domain. Children were, therefore, regressed to the previous section if they were unable to complete these items (Bayley, 2006a). In the previous section, four items required crawling skills. If the child ‘bum shuffled’ instead of moving to hands and knees, they could not receive credit for these items. These children therefore lost a considerable number of credit points despite being able to age appropriately cruise furniture, suggesting potentially inappropriate test item order for Australian children. Second, developmental trajectories may wax and wane across ages and countries based on a

range of country specific factors such as day care frequency and early learning programs. US child performance may, for example, match Australian child performance at 15-months on the gross motor domain, and Australian children may again match US children on the cognitive domain at 40-months. However, this is not possible to deduce without full age range Australian normative data.

6.6.2 Application of the Results to Assessment in an Australian Context

The means and standard deviations provided in Studies 1 and 3 are expected to improve the clinical and research utility of the BSID-III in Australian populations. These results may be used to calculate z-scores, as Australian normative data for 12- and 13-month, term born, singleton infants, and 3-year-old term born, singleton children. Results of this thesis underscore the need for future research to provide full age range Australian normative data for the BSID-III. Furthermore, a rigorous sampling frame is needed to ensure the normative sample is representative of socio-demographic factors relevant to the Australian population.

6.6.2 Sensitivity of the BSID-III to Detect and Predict Delays in Australia and Internationally

Taken together, results from Studies 1, 2 and 4 provide insight into the clinical utility of the BSID-III in Australia and internationally. Gold standard developmental assessment tools should possess two important characteristics: sensitivity to the detection of delay and the capacity to accurately predict future ability.

Study 1 and 2 raise questions about the sensitivity of the BSID-III to detect subtle delays, particularly in the language and fine motor domains. No clinically significant differences

in infant performance were detected in Study 2 between infants at indirect risk of NDD and infants at low risk, based on a range of perinatal risk factors. Furthermore, Study 1 demonstrated no clinically significant difference between Australian and US child performance on the language and fine motor domains at 1-year of age.

Conversely, Study 4 demonstrates that the BSID-III at 1-year does have some capacity to predict child outcomes at 3-years, particularly when stratified by sex. One-year receptive language, fine motor and gross motor scores for females and 1-year receptive language and expressive language scores for males, explained clinically significant proportions of variance in scores at 3-years (Study 4). This is consistent with developmental theory and research which suggests that developmental trajectories differ by sex. Results of this thesis not only bring into question the accuracy of previous predictive research where results were not stratified by sex, but also highlight the need for future research to consider the importance of sex as a moderator when analysing the predictive power of assessment instruments.

6.6.4 Future Directions for Assessing the Clinical Utility of the BSID-III

Future research examining the sensitivity and specificity of the BSID-III to detect and predict delays in infancy and early childhood, stratified by sex, is necessary to further inform the clinical utility of the BSID-III. Examination of the clinical utility of the BSID-III in children with conditions known to be associated with mild delays is of importance. Yet the clinical utility of the BSID-III must also be further examined in general population cohorts to determine its accuracy in identifying children with delay without clear biological markers. Re-assessment of the 1-year cohort in this thesis at school age, once developmental delays are likely to have been identified, on a measure such as the Wechsler Intelligence Scale for Children-5th Edition (WISC-

V) (Wechsler, 2014), and/or correlations drawn with school results such as National Assessment Program- Literacy and Numeracy (NAPLAN), would be of considerable value. This would allow between group analyses of BSID-III performance at 1-year, comparing those with known developmental delays at school age to those without. While it is noted that these measures assess slightly different constructs: the BSID-III is as a multi-domain measure of developmental delay (Bayley, 2006a); the WISC-V is a measure of intellect; and NAPLAN results reflect school performance, comparisons would be expected to be of value in order to determine the sensitivity of the BSID-III to detect those who go on to experience delay/excel, and in informing the overall clinical utility of the tool.

As the BSID-IV is currently in development, future directions for test development are important to consider. The results of this thesis provide compelling evidence for the need to develop Australian normative data, representative of the Australian population particularly with regard to parent ethnicity and education. Research examining the sensitivity and specificity of the BSID-IV to detect delay, as well as the predictive utility of the BSID-IV by sex would also increase its clinical and research utility. Moreover, results of this thesis suggest that examination of test item order and relevance, for an Australian population, is also necessary to consider. Gross motor items at the 12-month start point and cognitive items administered at 3-years of age warrant review, as the results of this thesis may be interpreted to suggest issues with item order/appropriateness in the BSID-III at these timepoints.

6.7 Conclusions

This thesis set out to provide a quantitative examination of the BSID-III in an Australian sample of 1-year-old and 3-year-old children to inform the clinical and predictive utility of the

tool in Australia. Major findings from this thesis included: (1) Statistical and clinically relevant differences between Australian and US children, both at 1- and 3-years of age, across two or more BSID-III domains; (2) Detected differences were likely to reflect population disparities between Australia and the US, such as ethnicity and maternal educational attainment; (3) The BSID-III is statically (but not clinically) sensitive to the detection of children at risk of NDD based on a range of indirect perinatal risk factors, compared to those at low risk; and, (4) the BSID-III holds some predictive utility from 1- to 3-years of age, with the predictive utility across BSID-III domains improving once stratified by sex.

Taken together, results improve knowledge of the clinical utility of the BSID-III for Australian children. This thesis provides important data that may be used clinically to interpret term born, singleton Australian child performance on the BSID-III at 12-months, 13-months and 3-years of age. It also provides the first examination of the predictive utility of all domains of the BSID-III in a general population cohort, and the first examination of the predictive utility of the BSID-III by sex. Future research providing full age range Australian normative data for the BSID-III is necessary. Furthermore, longitudinal studies examining the sensitivity and predictive utility of the BSID-III to detect delays in those children who go on to experience delays in later childhood are needed.

Considering the BSID-III is one of the most widely used developmental assessment tools in Australia, and 1- and 3-years form pivotal diagnostic and developmental follow-up ages in Australia, the results of this thesis provide much needed quantitative data pertaining to the utility of BSID-III in Australia.

6.8 References

Anderson, P. J., De Luca, C. R., Hutchinson, E., Roberts, G., & Doyle, L. W. (2010). Underestimation of developmental delay by the new Bayley-III Scale. *Archives of Pediatric and Adolescent Medicine, 164(4)*: 352-356. doi:10.1001/archpediatrics.2010.20.

Bandstra, E. S., Morrow, C. E., Mansoor, E., & Accornero, V. H. (2010). Prenatal drug exposure: infant and toddler outcomes. *Journal of Addictive Diseases*, *29*(2), 245-258. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1080/10550881003684871.

Baron, I. S., Ahronovich, M. D., Erickson, K., Larson, J. C. G., & Litman, F. R. (2009a). Age-appropriate early school age neurobehavioral outcomes of extremely preterm birth without severe intraventricular hemorrhage: A single center experience. *Early Human Development*, *85*(3), 191-196. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1016/j.earlhumdev.2008.09.411.

Baron, I. S., Erickson, K., Ahronovich, M. D., Coulehan, K., Baker, R., & Litman, F. R. (2009b). Visuospatial and verbal fluency relative deficits in 'complicated' late-preterm preschool children. *Early Human Development*, *85*(12), 751-754. doi: http://dx.doi.org.simsrad.net.ocs.mq.edu.au/10.1016/j.earlhumdev.2009.10.002.

Bayley, N. (2006a) *The Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III).* 3rd ed. San Antonio, Texas: Pearson Psychcorp.

Bayley, N. (2006b). *Bayley Scales of Infant and Toddler Development, Third Edition Technical Manual.* San Antonio, TX: Pearson PsychCorp.

Bode, M. M., D'Eugenio, D. B., Mettelman, B. B., & Gross, S. J. (2014). Predictive validity of the Bayley, at 2 years for intelligence quotient at 4 years in preterm infants. *Journal of Developmental & Behavioral Pediatrics, 35*(9), 570-575. doi: 10.1097/DBP.0000000000000110.

Bos, A. F. & Roze, E. (2011). Neurodevelopmental outcome in preterm infants. *Developmental Medicine and Child Neurology; 53*(s4), 35-39. doi: 10.1111/j.1469-8749.2011.04062.x.

Brownell CA, and Kopp CB, editors (2010) *Socioemotional development in the toddler years: Transitions and transformations*. Guilford Press; New York.

Charman T, Taylor E, Drew A, Cockerill H, Brown JA and Baird G (2005) Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry* 46(5): 500-513.

Chinta, S., Walker, K., Halliday, R., Loughran-Fowlds, A., & Badawi, N. (2014). A comparison of the performance of healthy Australian 3-year-olds with the standardised norms of the Bayley Scales of Infant and Toddler Development (version-III). *Archives of disease in childhood, 99*(7), 621-624. doi: 10.1136/archdischild-2013-304834

De Jesus, L. C., Pappas, A., Shankaran, S., Li, L., Das, A., Bell, E. F., et al. (2013). Outcomes of small for gestational age infants born at< 27 weeks' gestation. *The Journal of Pediatrics, 163*(1), 55-60. doi: http://dx.doi.org/10.1016/j.jpeds.2012.12.097

de Moura, D. R., Costa, J. C., Santos, I. S., Barros, A. J., Matijasevich, A., Halpern, R., et al. (2010). Risk factors for suspected developmental delay at age 2 years in a Brazilian birth

cohort. *Paediatric and Perinatal Epidemiology*, *24*(3), 211-221. doi: 10.1111/j.1365-3016.2010.01115.x.

Forns, J., Julvez, J., García-Esteban, R., Guxens, M., Ferrer, M., Grellier, J., et al. (2012). Maternal intelligence-mental health and child neuropsychological development at age 14 months. *Gaceta Sanitaria*, *26*(5), 397-404.

Greene, M. M., Patra, K., Nelson, M. N., & Silvestri, J. M. (2012). Evaluating preterm infants with the Bayley-III: patterns and correlates of development. *Research in Developmental Disabilities*, *33*(6), 1948-1956. doi: http://dx.doi.org/10.1016/j.ridd.2012.05.024.

Greene, M. M., Patra, K., Silvestri, J. M., & Nelson, M. N. (2013). Re-evaluating preterm infants with the Bayley-III: Patterns and predictors of change. *Research in Developmental Disabilities*, *34*(7), 2107-2117. doi: https://doi.org/10.1016/j.ridd.2013.04.001

Halle, T., Forry, N., Hair, E., Perper, K., Wandner, L., Wessel, J., & Vick, J. (2009). Disparities in early learning and development: lessons from the Early Childhood Longitudinal Study–Birth Cohort (ECLS-B). *Washington, DC: Child Trends*, 1-7.

Jarjour, I. T. (2015). Neurodevelopmental outcome after extreme prematurity: a review of the literature. *Pediatric Neurology*, *52*(2), 143-152. doi: http://dx.doi.org/10.1016/j.pediatrneurol.2014.10.027.

Krogh, M. T., Væver, M. S., Harder, S., & Køppe, S. (2012). Cultural differences in infant development during the first year: A study of Danish infants assessed by the Bayley-III and compared to the American norms. *European Journal of Developmental Psychology*, *9*(6), 730-736. doi: 10.1080/17405629.2012.688101.

Lung, F. W., Shu, B. C., Chiang, T. L., Chen, P. F., & Lin, L. L. (2009). Predictive validity of Bayley scale in language development of children at 6–36 months. *Pediatrics International, 51*(5), 666-669. doi: 10.1111/j.1442-200X.2009.02844.x.

Odd, D. E., Rasmussen, F., Gunnell, D., Lewis, G., & Whitelaw, A. (2008). A cohort study of low Apgar scores and cognitive outcomes. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *93*(2), F115-F120. doi: 10.1136/adc.2007.123745.

Peterson, J., Taylor, H. G., Minich, N., Klein, N., & Hack, M. (2006). Subnormal head circumference in very low birth weight children: neonatal correlates and school-age consequences. *Early Human Development*, *82*(5), 325-334. doi: https://doi.org/10.1016/j.earlhumdev.2005.09.014

Skiöld, B., Vollmer, B., Böhm, B., Hallberg, B., Horsch, S., Mosskin, M., et al. (2012). Neonatal magnetic resonance imaging and outcome at age 30 months in extremely preterm infants. *The Journal of Pediatrics*, *160*(4), 559-566. doi: http://dx.doi.org/10.1016/j.jpeds.2011.09.053.

Spittle, A. J., Spencer-Smith, M. M., Eeles, A. L., Lee, K. J., Lorefice, L. E., Anderson, P. J., et al. (2013). Does the Bayley-III Motor Scale at 2 years predict motor outcome at 4 years in very preterm children? *Developmental Medicine & Child Neurology, 55*(5), 448-452. doi: 10.1111/dmcn.12049.

Steenis, L. J., Verhoeven, M., Hessen, D. J., & Van Baar, A. L. (2015). Performance of Dutch children on the Bayley III: a comparison study of US and Dutch norms. *PloS one*, *10*(8), e0132871. doi:10.1371/journal.pone.0132871.

Wadhawan, R., Oh, W., Vohr, B. R., Wrage, L., Das, A., Bell, E. F., ... & Higgins, R. D. (2011). Neurodevelopmental outcomes of triplets or higher-order extremely low birth weight infants. *Pediatrics*, *127*(3), e654-e660. doi: 10.1542/peds.2010-2646.

Walker, K., Badawi, N., Halliday, R., & Laing, S. (2010). Brief report: performance of Australian children at one year of age on the Bayley scales of infant and toddler development (version III). *The Educational and Developmental Psychologist*, *27*(1), 54-58. doi:10.1375/aedp.27.1.54.

Walker, K., Holland, A. J., Halliday, R., & Badawi, N. (2012). Which high-risk infants should we follow-up and how should we do it? *Journal of Paediatrics and Child Health*, *48*(9), 789-793. doi: 10.1111/j.1440-1754.2012.02540.x.

Wechsler, D. (2011). Test of premorbid functioning. UK version (TOPF UK). *UK: Pearson Corporation*.

Wechsler, D. (2014). Wechsler intelligence scale for children-fifth edition. *Bloomington, MN: Pearson.*

Westera, J. J., Houtzager, B. A., Overdiek, B., & Van Wassenaer, A. G. (2008). Applying Dutch and US versions of the BSID-II in Dutch children born preterm leads to different outcomes. *Developmental Medicine & Child Neurology*, *50*(6), 445-449. doi: 10.1111/j.1469-8749.2008.02067.x.

Appendix

Appendix A – Triple B Protocol Paper

Cohort profile: The Triple B Pregnancy Cohort Study: A longitudinal study of the relationship between alcohol, tobacco and other substance use during pregnancy and the health and wellbeing of Australian children and families.

Authors (*Order to be confirmed*): Delyse Hutchinson[1234*], Judy Wilson[1], Steve Allsop[5], Elizabeth Elliott[7], Jake Najman[6], Lucinda Burns[1], Anne Bartu[8], Sue Jacobs[9], Ingrid Honan[1], Clare McCormack[1], Larissa Rossen[1], Hannah Fiedler[1], Chiara Stone[1], Sarah Khor[2], Joanne Ryan[34], George J Youssef[234], Craig A. Olsson[234], Richard P. Mattick, PhD[1]

And the Triple B Research Consortium*

Affiliations:
[1]National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia

[2]Deakin University, Centre for Social and Early Emotional Development, School of Psychology, Faculty of Health, Melbourne, Australia

[3]Murdoch Children's Research Institute, Centre for Adolescent Health, Royal Children's Hospital, Melbourne, Australia

[4]University of Melbourne, Department of Paediatrics, Royal Children's Hospital, Melbourne, Australia

[5]National Drug Research Institute, Curtin University, Perth, Australia

[6]Queensland Alcohol and Drug Research and Education Centre and Schools of Public Health and Social Science, University of Queensland, Brisbane, Australia

[7]Discipline of Paediatrics and Child Health, The University of Sydney, The Children's Hospital at Westmead, Sydney, Australia

[8]School of Nursing and Midwifery, Curtin University, Perth, Australia

[9]Department of Obstetrics, Royal Prince Alfred Hospital, Sydney, Australia

*This paper represents a collaborative effort on behalf of the Triple B Pregnancy Cohort Study. The contribution from authors to this paper was part of a broader collaborative effort from investigators who are a part of the research consortium. We request that investigators who are part of the consortium be listed at the end of the manuscript. Contributions made by other individuals have been acknowledged in the usual way. We hope you will give this request some consideration.

The Triple B Research Consortium: Including the primary investigators already listed and: Joanne Cassar, Aurora Popescu, Gabrielle Campbell, Lee Taylor, Maria Gomez, Emma Black, Danya Braunstein, Laura Dewberry, Erin Kelly, Alex Aiken, Sarah Brann, Sara Clews, Sharon Dawe, Adrienne Gordon, Paul Haber, Dale Hamilton, Andrew Lewis, Nyanda McBride, Elizabeth Moore, Raewyn Mutch, Julee Oei, George Patton, Ronald Rapee, Tim Slade, Marian Shanahan, Christine Stephens and Meredith Ward.

***Corresponding author:** Dr Delyse Hutchinson, Centre for Social and Early Emotional Development, Faculty of Health, School of Psychology, Deakin University, Burwood, Victoria, 3125. Email: delyse.hutchinson@deakin.edu.au

Short Title: Cohort Profile: The Triple B Pregnancy Cohort Study

Key Words: alcohol, tobacco, substance use, child development, family functioning, longitudinal study

Summary
The Triple B Pregnancy Cohort Study investigates the effects of parental alcohol, tobacco and other substance use on infant development and family functioning. The study (also known as: *Bumps, Babies and Beyond*), recruited two sub-samples: (1) a general antenatal clinic sample of pregnant women and their partners (n=1,534 women; 841 of their partners); and, (2) a smaller sample of pregnant women with diagnosed substance use disorders (SUD; n=89 women only). Participants were recruited through public antenatal clinics attached to major hospitals and area health services in New South Wales (NSW) and Western Australia (WA). Of 4,068 eligible women from the general antenatal clinics, 37.7% participated, with equivalent numbers for the SUD clinics being 198 and 44.9%. There were 1,453 and 65 live births from the two groups respectively, with a total of 1,414 and 65 mothers in the two groups. Data were also collected on 1,264 (86.9%) of 1,455 eligible partners of women recruited through the general antenatal clinics. The study collected repeated measures across pregnancy (Trimester 1, 2 and 3), and at 8-weeks and 12-months postnatally; retention at 12-months was 84.0% and 73.8% for mothers in the general antenatal and specialist SUD clinics, respectively. The data collected include demographic, parental, familial and infant factors, with a focus on parental substance use and mental health, parenting practices, familial functioning and infant development. Following pregnancy awareness, 42% of women consumed alcohol, 12% smoked tobacco, and 4% used illicit drugs at some stage in pregnancy. Comprehensive assessments have been conducted with infants at 12-months to test numerous developmental domains, including cognitive, motor, and language skills, along with measures of social and emotional functioning. Data access enquiries can be made to the Principal Investigator (d.hutchinson@unsw.edu.au).

Why was the cohort set up?
In 2010, the Australian National Medical Health Research Council (NHMRC) funded this longitudinal pregnancy cohort study to address limitations in knowledge of the effects of parental substance use on infants and families. The Triple B Pregnancy Cohort Study (known as: *Bumps, Babies and Beyond*), includes two sub-samples: (a) a public antenatal clinic sample of pregnant women and their partners, and (b) a smaller sample of pregnant women with diagnosed substance use disorders. Participants were recruited though public antenatal clinics at major hospitals and health services in New South Wales (NSW) and Western Australia (WA). The study aims to examine: the effects of parental substance use in pregnancy on infant development (e.g., cognitive, motor, language and socio-emotional development) and family functioning; the extent to which substance use is interrelated among couples; and, the influence substance use may have on parents' respective substance use patterns. The study aims to contribute to improved public health services. The study represents a significant investment in longitudinal research that builds on a history of quality Australasian longitudinal research.

The Developmental Origins of Health and Disease (DOHaD)[1] framework posits that early life environmental exposures induce critical changes in development that have long-term impacts on health and disease risk. Consistent with this framework, clinical studies of high-risk samples of parents diagnosed with substance use disorders suggest that parental substance use has significant adverse impacts on infant and child development[2 3]. For example, substance dependent pregnant women and their infants have an increased risk of obstetric, fetal and neonatal complications. These include miscarriage, prematurity, low birth weight for gestational

age, Fetal Alcohol Spectrum Disorders (FASD), Neonatal Abstinence Syndrome (NAS), and longer-term deficits in children's physical, cognitive, behavioural and emotional development[4-7]. The primary limitation of clinical studies is that the data are often derived from small-scale cross-sectional surveys, so that the effects over time are not known and the findings are not generalisable to the population.

Longitudinal studies provide insights into potential causes of health problems as well as factors capable of preventing or moderating problems. Emerging evidence from general population-based longitudinal studies suggests that the effect of parental substance use can vary considerably as a function of parent gender, pattern and type of substance use, and the presence of associated socioeconomic, physical health and psychosocial risk factors[2 3 8-11]. Parental substance misuse may impact adversely on children and families via direct and indirect pathways. For example, parental substance use can impact negatively on the quality of the marital/intimate partner relationship and the parent-child relationship by reducing family cohesion and increasing conflict and violence[12 13]. However, there is a dearth of data on parental substance use in the prenatal period, particularly defined by timing of exposure during pregnancy (i.e., pre-pregnancy awareness and at each trimester post-awareness), with most research focussing solely on maternal consumption and neonatal outcomes. Furthermore, many studies use poor measures of parental substance use (e.g., retrospective reports of use). Notably, few assess partner substance use and its impacts.

As a result, there are major gaps in current knowledge about the extent to which perinatal (pre- and postnatal) parental substance use impacts on early child development and family functioning, and the mechanisms by which these influences occur, particularly for low/moderate levels of use, which are the most frequent levels of substance use among Australian parents[14].

This cohort protocol aims to: (1) describe the Triple B Cohort samples and study methodology; and, (2) provide key summary data on perinatal substance use patterns, along with findings from three recent publications.

The Triple B Cohort Study is led by the National Drug and Alcohol Research Centre (NDARC) at UNSW Australia, and the National Drug Research Institute (NDRI) at Curtin University, in collaboration with Deakin University, Sydney University, the University of Queensland, the Murdoch Childrens Research Institute and the University of Melbourne. The study has been supported by the NHMRC (2010-2014), Australian Rotary Health (ARH; 2012-2013) and the Foundation for Alcohol Research and Education (FARE; 2010-2011). Additionally, PhD candidates on the project have been funded through ARH; the NDARC Education Trust (NET), Macquarie University; and both the Australian Centre for Perinatal Science (ACPS) and NDARC at UNSW. Ethics approval was granted by relevant university, hospital, and health services Human Research Ethics Committees.

Who is in the cohort?

The study used a prospective cohort design. Pregnant women were recruited through general antenatal clinics (N=1,534: NSW=1,246; WA=288) and specialist drug and alcohol antenatal clinics (N=89; NSW=59; WA=30) in public hospitals and health services in NSW and WA between 2009 and 2013. Participating hospitals in NSW were the Royal Prince Alfred Hospital,

Camperdown; The Royal Women's Hospital, Randwick; and Liverpool Hospital, Liverpool. Participants in Western Australia were recruited through the King Edward Memorial Hospital, Subiaco. Pregnant women were invited to participate in the study by research officers who attended antenatal clinics at each hospital, across all days and months of the year, to represent (proportionally), all clinics operating at each recruitment site. A standardised script was used to describe the study to women. Eligibility criteria included: being pregnant; being over 15 years of age; having no major medical complications (mother or fetus); intention of mother or both parents to be the primary caregiver/s; being mentally able to complete assessments; possessing sufficient literacy in English; and informed consent.

As outlined in Figures 1 and 2, 6,597 pregnant women (255 from the specialist substance use disorder (SUD) cohort) were approached and informed about the study; 4,266 (198 specialist SUD cohort) of these met eligibility criteria and were invited to participate. A total of 1,534 (37.7%) and 89 (44.9%) women from the general and specialist cohorts respectively, provided consent and completed at least one study measure. The participation rates of eligible women for individual hospitals ranged from 22.2% to 55.0%, with an overall participation rate of 44.0% for NSW; and 25.0% for WA.

Participation rates

Figures 1 and 2 present study flow diagrams of mother and infant participation rates, and partners, separated by non-exposed and substance use exposed groups, respectively. Of those who participated, 79 withdrew (6 from the specialist SUD cohort), and a further 65 (18 from the specialist cohort) were lost to follow-up prior to giving birth (attrition rate: 7.8% and 31.5% for general and specialist SUD cohort respectively). The remaining 1,414 mothers in the general cohort gave birth to a total of 1,436 offspring, which included 37 twin pairs, and one set of triplets. In the specialist SUD cohort, 65 singletons were born. Parents of twins and triplets were asked to complete a separate survey about each child. Data were collected on 1,264 (86.87%) of eligible partners from the general cohort, either directly (n=842), or indirectly (n=422) via maternal report.

INSERT FIGURES 1 AND 2 ABOUT HERE

Cohort characteristics

The characteristics of participating mothers and infants are presented in Table 1. Partner characteristics are presented in Table 2. Maternal and partner data were collected in Trimester 3; data on the infant was collected at the 8-week postnatal follow-up and was derived from infant *Blue Books* completed at birth by hospital staff. Compared with the Australian population, the Triple B general antenatal cohort is similar to the Australian population on rates of employment ($z = 1.40$; $p=0.08$), and the proportion of participants of Aboriginal or Torres Strait Islander origin ($z=-0.53$; $p=0.23$). The sex distribution of infants was also similar to Australian population figures ($z= -0.46$; $p=0.32$), although other infant characteristics differed from the general population, with longer gestation at birth ($z=13.8$; $p<0.001$; Somers' d= 0.000; 95% CI [-0.066, 0.068]), higher birthweight ($z=3.82$; $p=0.001$, Somers' d=0.07; 95% CI [-0.01, 0.13]), and a higher proportion of twins/multiple births ($z=4.55$; $p<0.001$; Cohen's h = 0.12; 95% CI [0.042, 0.186]). Mothers in the cohort were also older ($z=12.86$, $p<0.001$, Somers' d=0.11, 95% CI [0.05, 0.17]), more socio-economically advantaged (SEIFA; $t_{1577}=31.56$, $p<0.001$, Cohen's

d=0.47, 95% CI [0.42, 0.52]), and better educated than the general population (University educated, z=18.84; p<0.001; Cohen's h = 0.49; 95% CI [0.44, 0.55]). In addition, binomial tests showed that there were more women born overseas (z=15.03; p<0.001; Cohen's h = 0.36, 95% CI [0.31, 0.41]), a higher proportion of nulliparous women (z=10.70; p<0.001; Cohen's h = 0.28; 95% CI [0.22, 0.33]), and fewer living in single parent households (z=-11.45; p<0.001; Cohen's h=0.36; 95% CI [0.31, 0.41]) compared to population figures.

By contrast, the Triple B specialist SUD cohort reported higher levels of unemployment (z =-9.50, p<0.001; Cohen's h=1.09; 95% CI [0.889, 1.34], higher proportions of Aboriginal or Torres Strait Islander participants (z=7.10; p<0.001; Cohen's h=0.47; 95% CI [0.22, 0.56], higher proportions of Australian-born mothers (z=-0.32; p=0.001; Cohen's h=0.41; 95%CI [0.22, 0.68], and lower birth weights (z=-2.80, p=0.005; Somers' d=-0.41; 95% CI [-0.75, -0.07]) when compared to the Australian population. Mothers were also younger (z=-2.89; p=0.004, Somers' d=0.11; 95% CI [0.05, 0.17], less educated (University educated; z=-0.65; p<0.001; Cohen's h=0.90; 95% CI [0.69, 1.18]), with more living in single parent households (z=8.85; p<0.001; Cohen's h=0.80; 95% CI [0.58,1.03] compared to the Australian population. There were no differences in parity (z=-1.43, p=0.08), the number of multiple births (z=-1.41, p=0.08), infant gestational age (z=-0.26, p=0.80) and infant sex distribution (z=-0.76, p=0.22), between the specialist SUD cohort and the Australian population, although mothers in the specialist cohort were somewhat more socio-economically advantaged (t=4.25; p<0.001; Cohen's d=0.29; 95% CI [0.07, 0.50].

INSERT TABLES 1 AND 2 ABOUT HERE

Demographic data were provided by 823 (97.9%) of the 841 participating partners for mothers in the general antenatal cohort. In addition, mothers in this group reported partner demographic characteristics for 418 (68.1%) of the 614 eligible partners who refused to participate. Comparisons with Australian population data suggest that the proportion of partners of Aboriginal or Torres Strait Islander origin was less than the general population (z=-2.10; p=0.02). However, like mothers, partners in the cohort appear to be slightly older (z=10.55; p<0.001; Somers' d=0.27; 95% CI [0.20, 0.34]), more highly educated (University education; z=24.54; p<0.001; Cohen's h = 0.64; 95% CI [0.58, 0.70]), and more likely to be born overseas compared to the general population (z=15.41; p<0.001; Cohen's h = 0.41; 95% CI [0.35, 0.47]). In addition, the rate of employment was higher among partners in the cohort (z=13.17; p<0.001; Cohen's h = 0.47; 95% CI [0.42, 0.53]), and there was a higher proportion of same sex partners in comparison to the general population (z=2.20; p=0.014; Cohen's h=0.05, 95% CI [0.01, 0.11]).

Comparisons were also conducted on the demographic characteristics of partners from the general cohort as a function of whether data were obtained via self-report, or indirectly via maternal report. These comparisons showed that the two partner groups did not differ on employment status (93.7% versus 95.2% in full- or part-time employment; $\chi^2(1, N=1215)=1.06$; p = 0.30), same sex partner relationships (1.5% versus 0.7% female for partners who self-reported and those who were reported on indirectly, respectively ($\chi^2[1, N=1224]=1.41$; p = 0.24); or Aboriginal or Torres Strait Islander origin (1.5% versus 1.7%; $\chi^2[1, N=1217]=0.05$; p = 0.83). Nevertheless, participating partners were slightly younger than refusers (mean age 34.7 versus

35.5 years; t_{1215}=-2.35, p=0.01, Cohen's d = 0.14, 95% CI [0.02, 0.26]), and reported higher educational attainment (60.4% versus 50.1% completed University/college; $\chi^2[1, N=1245]=11.75$; p = 0.001; φ = 0.09, 95% CI [0.02, 0.19]).

How often have they been followed up?

Five assessment points are shown in Table 3. These include: Trimester 1 (conception to 12-weeks), Trimester 2 (13-weeks to 27-weeks), Trimester 3 (28-weeks to birth) and an 8-week follow-up (8-weeks postnatal). A comprehensive developmental follow-up occurred at infant age 12-months. Mothers were assessed at all time points; partners at Trimester 3, 8-weeks postnatal, and the 12-month follow-up; and infants at the 8-week and 12-month follow-up. Survey response rates for eligible mothers and infants are presented in Figure 1.

INSERT TABLE 3 ABOUT HERE

What is attrition like?

Attrition across the five waves of data collection for the general antenatal cohort has been low (Figure 1). Of the 1,399 mothers remaining in the cohort following delivery, 118 (8.4%) withdrew or were lost to further follow-up, such that the total attrition rate at 12-months from the original cohort of 1,534 was 16.0% (i.e., 84.0% retention). Of the 1,436 infants included in the study, developmental data was collected from 1,310 (91.2%) at the 12-month follow-up. Of the 841 general antenatal cohort partners who participated directly in the study, 57 (6.8%) withdrew or were lost to follow-up at 8-weeks, and a further 74 (8.8%) at 12-months (resulting in a total retention rate of 84.4%). Attrition rates for mothers from the specialist SUD drug and alcohol antenatal clinics were higher than the general antenatal population (Figure 2; 46.1% versus 16.0%; $\chi^2[1, N=1623]=52.5$; p < 0.001; φ= 0.18; 95% CI [0.10, 0.26]). After infant delivery, 61 (68.5%) mothers remained in the study and by 12-months, another 15 (24.6%) participants had withdrawn or were lost to follow-up, leaving a final retention rate of 53.9% from the original cohort.

Mothers in the both cohorts who withdrew or were lost to follow-up by 12-months were younger than those who continued (general cohort mean age 30.8 versus 33.1 years; $t_{289.8}$=-5.47, p<0.001; Cohen's d=0.44; 95% CI [0.30, 0.58]; specialist cohort mean age 28.2 versus 30.2 years; $t_{77.3}$=-1.69, Cohen's d=0.37; 95% CI [-0.06, 0.80]) and had higher rates of unemployment (51.8% vs 30.0%; $\chi^2[1, N=1623]=44.32$; p < 0.001; φ=0.17; 95% CI [0.10, 0.24]). In addition, participants in the general cohort who withdrew or were lost to follow-up at 12-months reported lower socioeconomic backgrounds through their Socio-Economic Indexes for Areas (SEIFA) scores (1027.7 versus 1050.5; $t_{314.4}$=-5.14; p<0.001; Cohen's d = 0.40; 95% CI [0.26, 0.54]), and were more likely to be from WA than from NSW (31.3% versus 12.4%; $\chi^2[1, N=1534]=72.74$; p < 0.001; φ=0.20; 95% CI [0.12, 0.28]).

Comparison of available infant data for both the general and specialist SUD antenatal cohort who did complete the 12-month development assessment with those who did not, showed no differences in sex (general, 55.0% versus 51.8% male; $\chi^2[1, N=1453]=0.49$; p=0.48; specialist, 58.8% versus 56.3% male; $\chi^2[1, N=55]=0.03$; p=0.85), birthweight (general, 3.41kgs versus 3.42kgs; t_{1424}=-1.55, p=0.44; specialist, 2.97kgs versus 3.18kgs; t_{56}=-1.07, p=0.14), or five-minute Apgar scores (general, 8.90 versus 8.93; $t_{86.38}$=-0.38, p=0.35; specialist, 8.70 versus 8.80;

t_{38}=-0.30; p=0.38) and weeks' gestation at birth (general, 38.88 versus 39.23; $t_{140.25}$=-1.59; p=0.06; specialist, 35.13 versus 38.33; t_{58}=-0.27; p=0.40).

What has been measured?
Table 4 provides a detailed overview of the measures included at each wave.

INSERT TABLE 4 ABOUT HERE

What has it found?
Data on alcohol and substance use in pregnancy and the postnatal period are presented in Table 5. Rates of alcohol, tobacco, and illicit substance use during pregnancy were highest in the period prior to pregnancy awareness, and decreased considerably after pregnancy awareness in the general cohort (trimester 1 pre- versus post-awareness $\chi^2(1, N=1302) = 548.48$, p<0.001, OR=0.03, 95% CI OR [0.01, 0.04]; $\chi^2(1, N=1300) = 108.04$, p<0.001, OR=0.01, 95% CI OR [0.00, 0.05]; $\chi^2(1, N=1301) = 36.36$, p<0.001, OR=0.05, 95% CI OR [0.01, 0.18], respectively). For women in the specialist cohort, decreases to alcohol and illicit substance use pregnancy awareness were also highest in trimester 1, $\chi^2(1, N=70) = 20.17$, p<0.001, OR = 0.04, 95% CI OR (0.00, 0.27); $\chi^2(1, N=70) = 14.00$, p<0.001, OR= 0.00, 95% CI OR (0.00, 0.30). Tobacco use, however remained unchanged $\chi^2(1, N=70) = 0.67$, p=0.414 before and after pregnancy awareness in trimester 1.,

INSERT TABLE 5 ABOUT HERE

Overall, following pregnancy awareness and during the course of pregnancy, 36.94% and 30.26% of women consumed any alcohol, 6.02% and 86.84% smoked cigarettes, and 1.37% and 52.63% consumed illicit drugs in the general and specialist groups, respectively. The quantity of alcohol consumption in the sample was generally low, averaging around two standard drinks per occasion in the specialist cohort and less than one in the generalist cohort. Alcohol use[16-18] was comparable with the general population for the two subgroups, whereas tobacco smoking[19 20], and illicit drug use[20] were lower in the general cohort subgroup but higher for the specialist subgroup, compared to the general population during pregnancy.

A number of articles have been published on the Triple B Cohort[21-28]. McCormack et al. [21] examined the patterns and predictors of alcohol consumption by women prior to awareness of pregnancy, and change in alcohol use following pregnancy recognition. Binge and heavy drinking were common in the early weeks of pregnancy, prior to pregnancy recognition (15.5% and 19.3%, respectively). Importantly, the rate of alcohol-exposed pregnancies was shown to be considerably higher than previous estimates when the period prior to pregnancy recognition is taken into account. Factors associated with changes in women's alcohol use following pregnancy recognition included level of alcohol use prior to pregnancy recognition, older maternal age, pregnancy planning, and illicit substance use. Heavy drinkers were more likely to cease drinking than low or moderate drinkers were. Women drinking at low or moderate levels were more likely to continue drinking at the same level than they were to cease completely relative to heavy drinkers. The results have important relevance to health policy and prevention to minimise alcohol-related harms to mothers and their offspring.

In regard to the SU group, there is a dearth of prospective data on women affected by substance use disorders during the perinatal period, often due to challenges with recruitment and retention. Yet understanding the experiences of these women at this time critical to informing perinatal services to promote maternal wellbeing and infant development. A recent, prospective study on the SUD group found that these women experience psychosocial disadvantage, poorer bonding to their developing fetus in-utero and elevated levels of perinatal distress, and postnatal parenting stress[22]. Findings highlight the critical importance of psychological and parenting support for these high-risk pregnant women and their offspring.

What are the main strengths and weaknesses?

The study provides five areas of innovation. It provides the most comprehensive longitudinal assessment of substance use in the perinatal period to date in Australia. Comprehensive assessment during this period will improve knowledge of the impact of substance use on infant development, and help identify critical risk thresholds and periods. Importantly, the study takes into account substance use behaviour both before and after pregnancy awareness; a distinction often overlooked in previous research.

Second, this is the first study comparing pregnant women recruited from a general antenatal clinic and a substance dependence treatment clinic, allowing for substance use to be examined across a wide spectrum from low/moderate, to harmful/dependent use. This will improve understanding of the psychosocial and physical risk factors from varying levels of substance use.

Third, the study is the first to comprehensively assess the impact of the partner's substance use and mental health, both pre- and postnatally, on child health and family function. It also assesses the influence partners have on each other's substance use.

Fourth, collection of buccal cells from infants (at 8-weeks and 12-months) and parents (at infant age 12-months) will provide a basis for epigenetic research into factors conferring individual differences in risk for substance use in parents, and adverse effects of parental substance use on children[23] (although we also note that cord-blood samples were not obtained, limiting the potential of epigenetic studies related to developmental origins and the effect of pregnancy exposure to these substances).

Finally, there is potential for data synthesis with intergenerational cohort studies. For example, major components of the Triple B Cohort assessments have been aligned with the Australian Temperament Project Generation Three Study (ATPG3) and the Victorian Intergenerational Health Cohort Study (VIHCS). This alignment has the potential to develop an integrated network of intergenerational cohorts, each focusing on specific prenatal and preconception periods. Specifically, the Triple B Cohort will provide rare and detailed data on exposures in pregnancy. The ATPG3 and VIHCS have a single antenatal assessment in Trimester 3 but rich preconception data across three and two generations, respectively.

The Triple B Study also captures key patterns of substance use for which there are major public health, prevention or treatment implications. Namely, it captures heavy/dependent substance use, addressed through oversampling pregnant women in treatment for substance use problems; and low to moderate (and binge) alcohol and tobacco use, which are adequately captured in the

antenatal clinic sample (based on power analyses). Given the low prevalence of illicit substance use (other than cannabis) in pregnancy it is unlikely that Triple B, or any other single study, unless very large scale, will be able to examine the impacts of low to moderate stimulant or opioid use on children. We also note that the sample is underpowered for genetic (but not epigenetic) research.

Although the Triple B Study is a multi-site study conducted in two states, it was conducted in public urban hospitals and therefore is not representative of rural areas of Australia, nor of families that utilise private hospitals. The planned cross-comparison and harmonisation of data with other major national and international cohorts may allow for increased pooled data and the potential capacity to examine outcomes of lower prevalence in smaller subgroups.

A further potential limitation of the study relates to the generalisability and validity of inferences drawn from the study, given that the general antenatal cohort differs in a number of ways from the general population, and there was evidence to suggest that attrition was higher among women with less privileged socioeconomic circumstances. Nevertheless, the cohort includes participants from a range of demographic backgrounds and with varying substance use patterns, and overall attrition was low. In addition, as noted above, substance use in the general antenatal cohort during pregnancy was consistent with Australian population data.

A major focus of the study was on comprehensive (prospective) data capture in the antenatal period; as such, there is a more lengthy gap between and the 8-week an 12-month assessments. This limits the capacity to understand how early postnatal exposures may affects growth and development. We do note, however, that information on some key developmental indicators (i.e., breastfeeding and sleep, for example), was assessed in the intervening period via recent retrospective reports.

We also note that response rates for the study ranged from 38-45%. Whilst these rates are consistent with some recent longitudinal cohorts with hospital-based recruitment in Australia[15], the limitation is that risk estimates may reflect underestimates of the true estimates because the extreme end of the distribution is less likely to be captured.

Finally, much of the information collected, including substance use data, was via self-report, and is thus subject to potential biases. In order to address this limitation, 85 participants were randomly selected for urine analysis during their third trimester of pregnancy, to confirm self-reported illicit substance use. Agreement between self-reported substance use and urine analysis was 97%, indicating that the information provided was reliable.

Despite these limitations, the study will improve understanding of the effects of parental substance use on infants and families, which can direct health policy. The results can inform development of public health prevention and early intervention campaigns to allow parents to make informed choices about substance use during the prenatal period. The results will also identify the health and obstetric needs of pregnant women characterised by harmful and/or risky patterns of substance use. Improvements in these areas may subsequently reduce developmental problems in infant and family functioning problems in the community. The results of this study can also inform national guidelines on use of alcohol and other substances pre-conception, in

pregnancy and whilst breastfeeding, which may guide public health education and policy on substance use.

Can I get hold of the data? Where can I find out more?
Further information can be obtained through the National Drug and Alcohol Research Centre, University of New South Wales: (https://ndarc.med.unsw.edu.au/project/triple-b-bumps-babies-and-beyond). Enquires can be directed to Dr Hutchinson (corresponding author). Data access is governed by the investigators. Research proposals must be consistent with ethical approval and participant consent, confidentiality and data management. The study protocol for collaborative research requires ratification by the respective ethics committee affiliated with the research.

Acknowledgements
We gratefully acknowledge the NDARC and NDRI research staff and students who assisted with collection of the data, the hospitals and antenatal clinics for their assistance with recruitment, and the study participants and their families. We wish to acknowledge Rosa Alati, Brandi Baylock, Lauren Bell, Elissa Bowey Annie Bleeker, Apo Demirkol, Genevive Eckstein, David Fergusson, Thea Gumbert, Helen Gunn, Jeannie Minnis, Colleen O'Leary, Vaughan Palmer, Jemma Pope, Jarrod Proudfoot, Candice Rainsford, Joanne Ross, Fiona Shand, Lisa Sin, Matthew Sunderland, Wendy Swift, Scarlet Wilcock and Jesse Young. We also wish to acknowledge the Cannabis Cohorts Research Consortium; NHMRC Project Grants: AAP1009381, AAP1064893).

Funding
The research was funded by an Australian National Health and Medical Research Council (NHMRC) Project Grant #GNT630517 for $2,196,179 to Richard P Mattick, Delyse Hutchinson, Steve Allsop, Jake Najman, Elizabeth Elliott, Lucy Burns, Sue Jacobs, Craig Olsson and Anne Bartu, and was financially supported by the National Drug and Alcohol Research Centre (NDARC), University of New South Wales (UNSW). NDARC and the National Drug Research Institute (NDRI), Curtin University are funded by the Australian Government under the Substance Misuse Prevention and Service Improvements Grants Fund. Richard P Mattick is financially supported by an NHMRC Principal Research Fellowship Award from the NHMRC, and Delyse Hutchinson is financially supported by an Australian Unity Industry Partner Senior Research Fellowship. Craig Olsson is supported by an Australian Research Council Senior Research Fellowship (DORA: DP 130101459). Elizabeth Elliott is supported by an NHMRC Practitioner Fellowship #1021480.

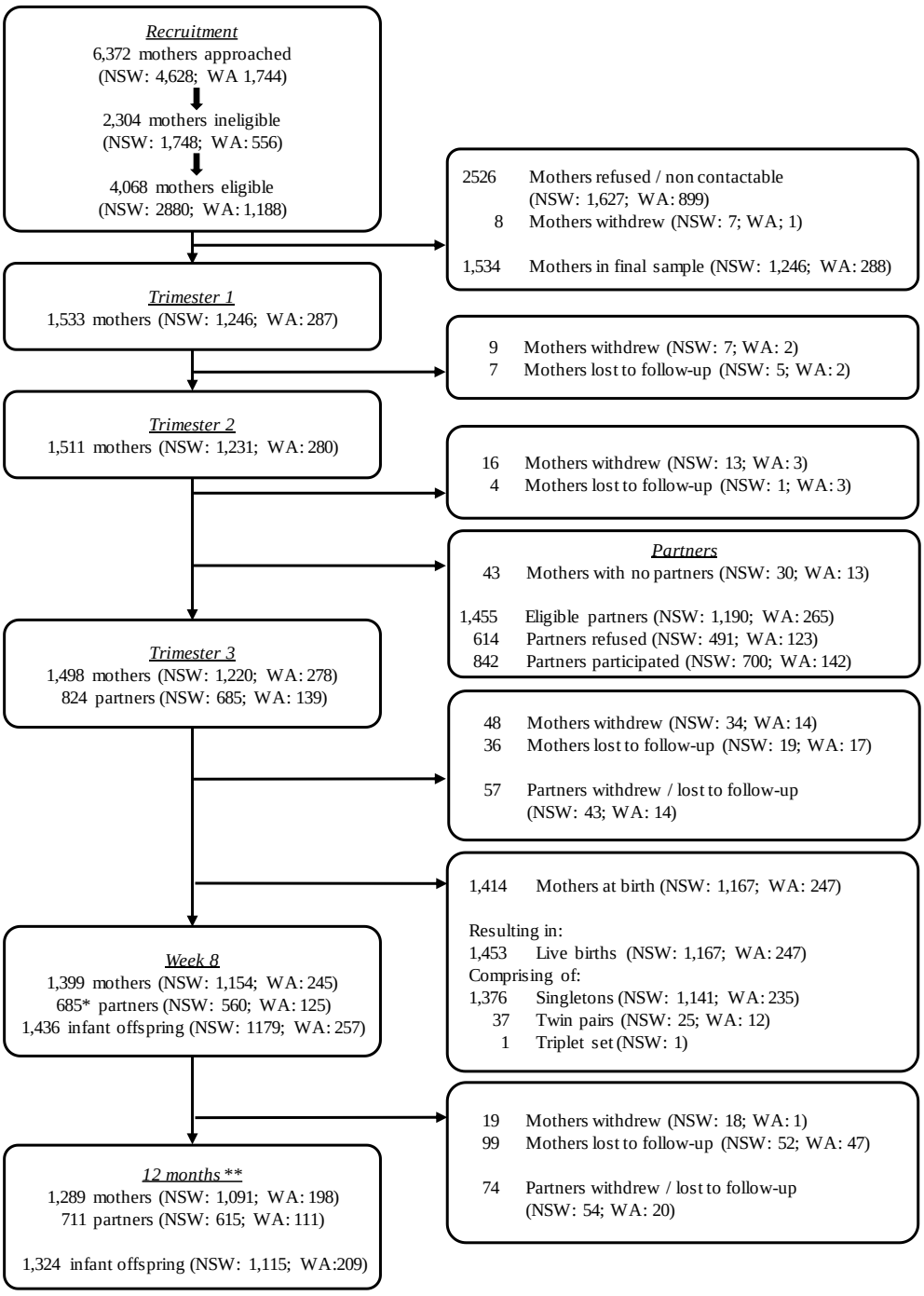
References

1. Gluckman P, Hanson M (eds.). *Developmental origins of health and disease*. Cambridge: Cambridge University Press, 2006.
2. Mitchell P, Spooner, C., Copeland, J., Vimpani, G., Toumbourou, J., Howard, J., et al. *The role of families in the development, identification, prevention and treatment of illicit drug problems. NHMRC Report*. Canberra: Commonwealth of Australia., 2001.
3. Hutchinson D, Mattick RP, Braunstein D, et al. The impact of alcohol use disorders on family life: A review of the empirical literature. National Drug and Alcohol Research Centre, University of New South Wales. (Technical Report), 2014.
4. Elliott E, Payne, J., Haan, E., Bower, C. . Diagnosis of Foetal Alcohol Syndrome and alcohol use in pregnancy: A survey of paediatricians' knowledge, attitudes and practice. . Journal of Paediatrics and Child Health 2006;**42**(11):698-703.
5. Stromland K, Mattson, S., Adnams, C., Autti-Ramo, I., Riley, E., Warren, K. Fetal Alcohol Spectrum Disorders: An international perspective. . Alcoholism: Clinical and Experimental Research 2005;**29**(6):1121-26.
6. Bandstra ES, Morrow, C.E., Mansoor, E., Accomero, VH. . Prenatal drug exposure: Infant and toddler outcomes. . Journal of Addictive Disorders 2010;**29**(2):245-58.
7. National clinical guidelines for the management of drug use during pregnancy, birth and the early development years of the newborn. Canberra: Commonwealth of Australia, 2006.
8. Alati R, Davey Smith G, Lewis SJ, et al. Effect of prenatal alcohol exposure on childhood academic outcomes: Contrasting maternal and paternal associations in the ALSPAC Study. PLoS ONE 2013;**8**(10):e74844.
9. Alati R, Macleod J, Hickman M, et al. Intrauterine exposure to alcohol and tobacco use and childhood IQ: findings from a parental-offspring comparison within the Avon Longitudinal Study of Parents and Children. Pediatric Research 2008;**64**(659-666).
10. O'Callaghan FV, O'Callaghan M, Najman JM, et al. Prenatal alcohol exposure and attention, learning and intellectual ability at 14 years: A prospective longitudinal study. Early Human Development 2007;**83**:115-23.
11. Streissguth A. Offspring effects of prenatal alcohol exposure from birth to 25 years: The Seattle Prospective Longitudinal Study. Journal of Clinical Psychology in Medical Settings 2007;**14**:81-101.
12. Grekin ER, Brennan, P.A., Hammen, C. . Parental alcohol use disorders and child delinquency: the mediating effects of executive functioning and chronic family stress. Journal of Studies on Alcohol 2005;**66**:14-22.
13. Goodwin RD, Fergusson DM, Horwood LJ. Childhood abuse and familial violence and the risk of panic attacks and panic disorder in young adults. Psychologial Medicine 2005;**35**:881-90.
14. Dawe S, Frye S, Best D, et al. Drug use in the family: Impacts and implications for children. In: ANCD C, ed., 2007.
15. Vuillermin P, Saffery R, Allen K, et al. Cohort profile: The Barwon infant study. International Journal of Epideniology 2015:1148-60.
16. Hutchinson D, Moore EA, Breen C, et al. Alcohol use in pregnancy: Prevalence and predictors in the Longitudinal Study of Australian Children. Drug and alcohol review 2013;**32**(5):475-82.

17. O'Keeffe L, Kearney P, McCarthy F, et al. Prevalence and predictors of alcohol use during pregnancy: Findings from international multicentre cohort studies. British Medical Journal 2015;**5**(7):e006323.
18. Australian Institute of Health and Welfare 2014. National Drug Strategy Household Survey detailed report 2013. Drug statistics series no. 28. Cat. no. PHE 183. Canberra: AIHW.
19. Hilder L, Zhichao Z, Parker M, Jahan S, Chambers GM. Australia's mothers and babies 2012. Perinatal statistics series no. 30. Cat. no. PER 69. Canberra: AIHW, 2014.
20. Australian Institute of Health and Welfare. Drugs in Australia 2010: tobacco, alcohol and other drugs. Drug statistics series no. 27. Cat. no. PHE 154. Canberra: AIHW, 2011.
21. McCormack C, Hutchinson D, Burns L. et al. Prenatal alcohol consumption between conception and recognition of pregnancy. Alcoholism: Clinical and Experimental Research 2017;**41**(2):369-378.
22. Hutchinson D, Taylor L, Rapee R. et al. Mental health, maternal bonding and parenting stress in women attending a specialist perinatal drug health service. Drug Dependence: Avid Science Publications. Accepted December 2016.
23. Fransquet P, Hutchinson D, Olsson CA. et al. Perinatal maternal alcohol consumption and methylation of the dopamine receptor DRD4 in the offspring: The Triple B Study. Environmental Epigenetics 2016:1-9. 10.1093/eep/dvw023.
24. Rossen L, Hutchinson D, Wilson J. et al. Maternal bonding through pregnancy and postnatal: Findings from an Australian longitudinal study. American Journal of Perinatology 2017. 10.1055/s-0037-1599052.
25. McPhie S, Skouteris H, Mattick RP. et al. Weight in the first year of life: Associations with maternal prepregnancy body mass index and gestational weight gain: Findings from a longitudinal pregnancy cohort. American Journal of Perinataology 2017. 10.1055/s-0036-1597992
26. Tay R, Wilson J, McCormack C. (In press). Alcohol consumption by breastfeeding mothers: Frequency, correlates and infant outcomes. Drug and Alcohol Review. Accepted July 2016.
27. Rossen L, Hutchinson D, Wilson J. Predictors of postnatal mother-infant bonding: The role of antenatal bonding, maternal substance use and mental health. Archives of Women's Mental Health 2016;**19**(4):609-622.
28. Burns L, Conroy E, Maloney E. et al. Psychosocial characteristics and obstetric health of women attending a specialist substance use antenatal clinical in a large metropolitan hospital. International Journal of Pediatrics 2011:1-7. Article ID 729237, 10.1155/2011/729237.
29. Australian Bureau of Statistics. Births, Australia, 2013. 3301.0. Canberra: Australian Bureau of Statistics, 2014.
30. Pink B. Socio-Economic Indexes for Areas (SEIFA). Technical Paper, 2006. Canberra: Australian Bureau of Statistics, 2008.
31. Australian Bureau of Statistics. Education and Work, Australia, May 2014. 6227.0. Canberra: Australian Bureau of Statistics, 2014.
32. Australian Bureau of Statistics. Gender Indicators, Australia, August 2015. 4125.0. Canberra: Australian Bureau of Statistics, 2015.
33. Australian Bureau of Statistics (ABS). Cultural diversity in Australia - Reflecting a Nation: Stories from the 2011 Census, 2012-2013. 2071.0. Canberra: Australian Bureau of Statistics; 2012.

34. Australian Bureau of Statistics (ABS). (2012). http://www.adelaide.edu.au/phidu/help-info/about-our-data/indicators-notes/sha-aust/demography-ses/nes-countries.html.
35. Australian Bureau of Statistics (ABS). Labour force, Australia: Labour Force Status and Other Characteristics of Families, June 2012. 6224.0.55.001. Canberra: Australian Bureau of Statistics; 2013.
36. Australian Bureau of Statistics (ABS). Census of Population and Housing - Counts of Aboriginal and Torres Strait Islander Australians, 2011. 2075.0. Canberra: Australian Bureau of Statistics, 2012.
37. Li Z, Zeki R, Hilder L & Sullivan EA. Australia's mothers and babies 2011. Perinatal statistics series no. 28. Cat. no. PER 59. Canberra: AIHW National Perinatal Epidemiology and Statistics Unit, 2013.
38. Australian Bureau of Statistics (ABS). Same-sex couple families - Reflecting a Nation: Stories from the 2011 Census. 2071.0. Canberra: Australian Bureau of Statistics, 2012.
39. Australian Bureau of Statistics (ABS). Cultural diversity in Australia - Reflecting a Nation: Stories from the 2011 Census, 2012-2013. 2071.0. Canberra: Australian Bureau of Statistics; 2012.
40. SEIFA: Mean Index of Relative Socio-economic Advantage and Disadvantage (IRSAD) Score (SD).
41. World Health Organisation (1997). Composite International Diagnostic Interview (CIDI) core version 2.1. Geneva: World Health Organisation.
42. Cottler LB, Robins lN, Grant BF, et al. The CICI-core substance abuse and dependence questions: Cross-cultural and nosological issues. The WHO/ADAMHA Field Trial. British Journal of Psychiatry 1991;**159**:653-58.
43. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. Sydney, Australia: Psychology Foundation, 1995.
44. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: Development of the 10-item Edinburgh postnatal depression Scale (EPDS). British Journal of Psychiatry 1987;**150**:782-86.
45. Busby DM, Christensen C, Crane RD, et al. A revision of the dyadic adjustment scale for use with distressed and nondistressed couples: Construct hierarchy and multidimensional scales. Journal of Marital and Family Therapy 1995;**21**(3):289-308.
46. Hudson WW, McIntosh SR. The assessment of spouse abuse: Two quantifiable dimensions. Journal of Marriage and Family 1981;**43**:873-85.
47. Condon JT, Corkindale CJ, Boyce P. Assessment of postnatal paternal-infant attachment: Development of a questionnaire instrument. Journal of Reproductive and Infant Psychology 2008;**26**(195-210).
48. Biringen Z, Robinson, J. L. & Emde, R. N. The Emotional Availability scales (3rd ed) Unpublished Manuscript, University of Colorado Health Sciences Center, Denver, 1999.
49. Abidin, R. A. (2005). Parenting stress index (3rd Ed.): Professional Manual. Florida: PAR Inc.
50. Squires J, Bricker D, Potter L. Revision of a parent-completed developmental screening tool: Ages and Stages Questionnaires. Journal of Pediatric Psychology 1997;**22**(3):313-28.
51. Bayley N. *Bayley Scales of Infant and Toddler Development, Third edition (Bayley-III).* Texas, U.S: Harcourt Assessment, 2005.

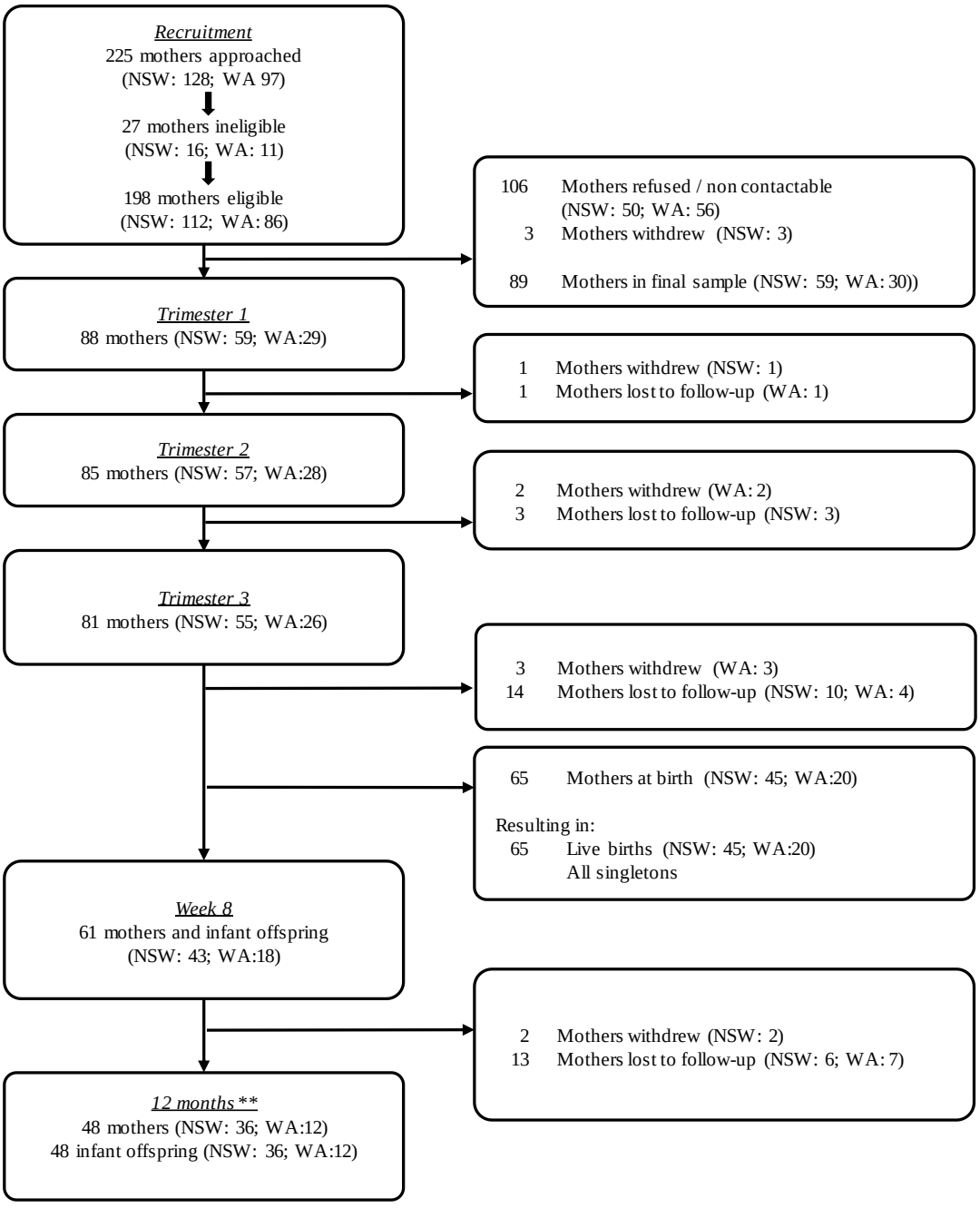
Figure 1: Triple B Pregnancy Cohort Study: Study flow diagram of mother, partner and infant participation rates (N=1,533 families) for the general antenatal sample.



*Note: The 8-week follow-up interview for partners was introduced after the pilot study. As such, 8-week data was unavailable for 60 participating partners as it was not offered.

**In some families, only infants were assessed at 12-months; in total, 1,295 families remained in the cohort.

Figure 2: Triple B Pregnancy Cohort Study: Study flow diagram of mother and infant participation rates (N=88) for the specialist substance use disorder (SUD) antenatal sample.



**In some families, only infants were assessed at 12 months; in total, 50 families remained in the cohort.

Table 1: Mother and infant cohort characteristics and comparison with Australian population data

Mother characteristics	**General Cohort at Trimester 3 (n=1,498)[a]**	**Specialist SUD Cohort at Trimester 3 (n=81)[a]**	**Australian population**
Mean age (years)	32.5 (SD = 5.1) Range: 17-52 Median: 33.0	28.8 (SD = 5.6) Range: 17-42 Median: 28.0	In 2013, the median age of Australian women giving birth was 30.8 years[29].
Mean Index of Relative Socio-economic Advantage and Disadvantage (IRSAD)	1047.2 (SD=57.8) Range: 790-1,164	1028.7 (SD=60.8) Range: 853-1,122	IRSAD is standardised to a distribution with a mean score of 1000, and a standard deviation of 100[30].
Education	**n (%)**	**n (%)**	
Year 10 or below	92 (6.1)	47 (58.0)	In 2014, 63.6% of females aged 20-64 had a post-high school qualification; 31.0% had a university degree. Corresponding figures for 30-34 year-old females show that 74.1% had a post-school qualification; 43.3% had a university degree[31].
Year 12	170 (11.4)	9 (11.1)	
Diploma, trade qualification	223 (14.9)	19 (23.4)	
University/college degree	1,008 (67.3)	6 (7.4)	
Employment			
Employed (full or part time)	1,001 (66.8)	12 (14.8)	In 2012-2013, 65.1% of females aged 20-74 were employed[32].
Country of birth			
Australia	832 (55.5)	72 (88.9)	In 2011, 27% of people living in Australia were born overseas[33], and 15.7% of the population were born in Non main English-speaking countries[34].
Other English-speaking	279 (18.6)	6 (7.4)	
Non main English-speaking	382 (25.5)	3 (3.7)	
Single parent household	90 (6.0)	44 (54.3)	In 2012, 17.2% of families with children under 15 were single mother families[35].
Aboriginal/Torres Strait Islander	34 (2.3)	12 (14.8)	In 2011, 2.5% of Australia's population identified as being of Aboriginal and/or Torres Strait Islander origin[36].
Parity			
0	841 (56.1)	28 (34.6)	In 2012, 42.4% of mothers had no previous pregnancies; 33.2% had one; 14.1% had two; 8.5% had three or more[19].
1	439 (29.3)	29 (35.8)	
2	150 (10.0)	10 (12.4)	
3 or more	68 (4.1)	12 (14.8)	

Infant characteristics	General Cohort at Trimester 3 (n=1,453)[a]	Specialist SUD Cohort at Trimester 3 (n=65)[a]	Australian population
	n (%)	n (%)	
Female	696 (47.9)	28 (43.1)	In 2012, 48.5% of live births were females[19].
Twins/triplets	77 (5.3)	0 (0.0)	In 2012, 3.0% of births were twins or other multiple births[19].
Mean gestation in weeks	39.2 (SD = 1.82) Range: 27-43	38.3 (SD = 2.6) Range: 30-42	In 2012, the mean gestational age for all babies was 38.7 weeks [19].
Mean birthweight in kilograms	3.42 (SD = 0.55) Range: 0.98-5.70	3.13 (SD = 0.67) Range: 1.33-5.24	In 2011, the mean birthweight of liveborn babies was 3.37kgs[37].

[a] Individual sample sizes for each characteristic vary slightly due to missing data. SUD = substance use disorder.

Table 2: Partner cohort characteristics and comparison with Australian population data

Partner characteristics	**General Antenatal Cohort at Trimester 3 (n=1,245)** [a, b]	**Australian population**
Age in years	35.0 (SD = 5.9) Range: 17-59 Median: 35.0	The median age of fathers for births registered in 2013 was 33.0 years[29].
	n (%)	
Same sex (female)	15 (1.2)	In 2011, 0.7% of Australian couples were same-sex couples[38].
Education		
Year 10 or below	111 (8.9)	In 2014, 65.6% of males aged 20-64 had a post-high school qualification, with 26.0% having a university degree[31].
Year 12	154 (12.4)	
Diploma, trade qualification	260 (20.9)	
University/college degree	692 (55.6)	
Employment		In 2012-2013, 78.8% of males aged 20-74 were employed[32].
Employed (full or part time)	1,145 (91.9)	
Country of birth		In 2011, 27% of people living in Australia were born overseas, and 15.7% of the population were born in Non main English-speaking countries[34 39].
Australia	651 (52.3)	
Other English-speaking	284 (22.8)	
Non main English-speaking	284 (22.8)	
Aboriginal/Torres Strait Islander	19 (1.5)	In 2011, 2.5% of Australia's population identified as being of Aboriginal and/or Torres Strait Islander origin[36].

[a] Individual sample sizes for each characteristic vary slightly due to missing data.
[b]Includes participating and non-participating partners, where data is available.

Table 3. Assessment schedule and methods of the Triple B Pregnancy Cohort Study

	Pregnancy Trimester 1	**Pregnancy Trimester 2**	**Pregnancy Trimester 3**	**Postnatal 8-week**	**Postnatal 12-month**
Mother	Interview, Questionnaire,	Interview, Questionnaire,	Interview, Questionnaire, Urine sample	Interview, Questionnaire, Blue Book, Buccal swab	Interview, Questionnaire, Observational Assessment
Partner	--	--	Interview, Questionnaire,	Interview, Questionnaire, Buccal swab	Interview, Questionnaire, Observational Assessment
Infant Offspring	--	--	--	Blue Book, Developmental Assessment, Buccal swab	Developmental/ Observational Assessments, Buccal swab

Note. In instances where women commenced participation after Trimester 1 or 2, pregnancy assessments were completed for earlier waves retrospectively.

Table 4. Mother, partner and infant measures at each assessment

Parent measures	Trimester 1	Trimester 2	Trimester 3	8 week postnatal	12 month postnatal
Parent and household demographics					
Birthdate, gender, country of birth, family composition, education, SEIFA[40]	-	-	✓	-	-
Employment, income	✓	✓	✓	✓	✓
Religiosity	-	-	✓	-	✓
Maternal substance use and mental health					
Age of first alcohol, tobacco, illicit substance use	-	-	✓	-	-
Alcohol, tobacco, illicit substance, caffeine use – Q/F*	✓	✓	✓	✓	✓
Heavy episodic alcohol use – Q/F	✓	✓	✓	✓	✓
Lifetime/past 12 month mental health diagnosis [41 42]	-	-	✓	-	-
Depression, stress, anxiety [43 44]	✓	✓	✓	✓	✓
Antisocial behaviour	-	-	✓	-	-
Social functioning	-	-	✓	-	✓
Current treatment	✓	✓	✓	✓	✓
Paternal substance use and mental health					
Age of first alcohol, tobacco, illicit substance use	-	-	✓	-	-
Alcohol, tobacco, illicit substance, caffeine use – Q/F	-	-	✓	✓	✓
Heavy episodic alcohol use – Q/F	-	-	✓	✓	✓
Lifetime/past 12 month mental health diagnosis [41 42]	-	-	✓	-	-
Depression, stress, anxiety [43]	-	-	✓	✓	✓
Antisocial behaviour	-	-	✓	-	-
Social functioning	-	-	✓	-	✓
Current treatment			✓	✓	✓

Parent measures	Trimester 1	Trimester 2	Trimester 3	8 week	12 month

				postnatal	postnatal
Preconception					
Alcohol, tobacco, illicit substance use, maternal – Q/F	-	-	✓	-	-
Alcohol, tobacco, illicit substance use, paternal – Q/F	-	-	✓	-	-
Pregnancy planning	-	-	✓	-	-
Parent relationship functioning					
Relationship adjustment/satisfaction, maternal [45]	-	-	✓	-	✓
Relationship adjustment/satisfaction, paternal [45]	-	-	✓	-	✓
Spousal abuse, maternal [46]	-	-	✓	-	✓
Spousal abuse, paternal [46]	-	-	✓	-	✓
Maternal general health					
Diet (24hr food diary)/vitamin/supplement use	-	-	✓	✓	✓
Physical health	-	-	✓	✓	✓
Physical activity	-	-	✓	-	✓
Sexual health	-	-	✓	✓	✓
Medical treatment	-	-	✓	✓	✓
Pregnancy complications	-	-	✓	✓	-
Sleep	-	-	-	✓	✓
Paternal general health					
Diet (24hr food diary)	-	-	✓	-	✓
Physical health	-	-	✓	-	✓
Physical activity	-	-	✓	-	✓
Medical treatment	-	-	✓	-	✓
Sleep	-	-	-	✓	✓

Infant measures	Trimester 1	Trimester 2	Trimester 3	8 week postnatal	12 month postnatal
Infant demographics					
Date of birth, gender	-	-	-	✓	-
Birth/postnatal outcomes					
Gestational age	-	-	-	✓	-
Weight	-	-	-	✓	✓
Head circumference	-	-	-	✓	✓
Length / height	-	-	-	✓	✓
APGAR	-	-	-	✓	-
Delivery	-	-	-	✓	-
Feeding	-	-	-	✓	✓
Parent infant relationship					
Fetus/infant bond, maternal [47]	✓	✓	✓	✓	✓
Fetus/infant bond, paternal [47]	-	-	✓	✓	✓
Emotional availability/caregiving [48]	-	-	-	-	✓
Parenting stress [49]	-	-	-	-	✓
Infant behaviour and development					
Diet (24hr food diary)	-	-	-	-	✓
Sleep	-	-	-	✓	✓
Temperament	-	-	-	✓	✓
Health	-	-	-	✓	✓
Cognition [50 51]	-	-	-	✓	✓
Motor – gross, fine [50 51]	-	-	-	✓	✓
Language/communication – receptive, expressive [50 51]	-	-	-	✓	✓
Socio-emotional [50 51]	-	-	-	✓	✓
Childcare	-	-	-	✓	✓

Infant measures	Trimester 1	Trimester 2	Trimester 3	8 week postnatal	12 month postnatal
Media exposure	-	-	-	-	✓
Biological data					
Urine samples, maternal^	-	-	✓	-	-
Buccal samples (mother/partner 8-week only; infant) ^^	-	-	-	✓	✓

*Q/F = quantity/frequency; ^Conducted randomly in 85 mothers. ^^Buccal samples available: Mother: n=1,274; Partner: n=647; Infant 8-week: n=1,268; Infant 12-month: n=1,066).

Table 5: Alcohol and other substance use in the Triple B Cohort during pregnancy and following delivery

	Trimester 1 Pre-awareness N=1,389[a] (n, %)		Trimester 1 Post-awareness N=1,599[b] (n, %)		Trimester 2 N=1,554 (n, %)		Trimester 3 N=1,447 (n, %)		8-weeks Postnatal N=1,449 (n, %)	
Alcohol	**General**	**Specialist**	**General**	**Specialist**	**General**	**Specialist**	**General**	**Specialist**	**General**	**Specialist**
Typical frequency of consumption										
Never	503 (38.3)	36 (46.8)	1229 (80.9)	62 (78.5)	1038 (70.6)	69 (82.1)	977 (70.4)	50 (83.3)	530 (38.1)	35 (59.3)
Less than once per month	158 (12.0)	7 (9.1%)	100 (6.6)	6 (7.5)	129 (8.8)	4 (4.8)	104 (7.5)	5 (8.3)	225 (16.2)	14 (23.7)
Once per month	35 (2.7)	6 (7.8%)	34 (2.2)	2 (2.5)	72 (4.9)	5 (6.0)	52 (3.8)	1 (1.7)	82 (5.9)	3 (5.1)
2-3 times per month	74 (5.6)	-	55 (3.6)	1 (1.3)	87 (5.9)	1 (1.2)	86 (6.2)	-	105 (7.6)	1 (1.7)
1-2 times per week	308 (23.5)	13 (16.9)	84 (5.5)	6 (7.6)	121 (8.2)	3 (3.6)	134 (9.7)	4 (6.7)	273 (19.6)	5 (8.5)
3-4 times per week	133 (10.1)	9 (11.7)	11 (0.7)	2 (2.5)	18 (1.22)	1 (1.2)	22 (1.6)	-	118 (8.5)	-
5-6 times per week	40 (3.1)	3 (3.9)	1 (0.1)	-	1 (0.1)	1 (1.2)	4 (0.3)	-	29 (2.1)	1 (1.7)
Daily	61 (4.7)	3 (3.9)	6 (0.4)	-	4 (0.3)	-	8 (0.6)	-	28 (2.0)	-
Among drinkers:	**n=809**	**n=41**	**n=291**	**n=17**	**n=432**	**n=15**	**n=410**	**n=10**	**n=860**	**n=24**
Median number of standard drinks consumed per typical occasion	3.0 (IQR=3.0)	4.5 (IQR=6.0)	1.5 (IQR=0.4)	3.0 (IQR=3.0	1.5 (IQR=0.5)	2.0 (IQR=1.5)	1.5 (IQR=0.3)	2.3 (IQR=3.0)	1.5 (IQR=0.5)	2.7 (IQR=4.1)
Binge drink during period (> 4 drinks on one occasion)	416 (51.4)	30 (73.2)	26 (8.9)	7 (41.2)	17 (3.9)	4 (26.7)	14 (3.4)	4 (40.0)	158 (18.4)	10 (41.7)
Median quantity per week (standard drinks)	4.5 (IQR=9.4)	6.8 (IQR=24.8)	0.6 (IQR=2.1)	2.3 (IQR=4.1)	0.5 (IQR=2.1)	0.8 (IQR=6.4)	0.9 (IQR=2.1)	1.5 (IQR=2.1)	1.9 (IQR=4.1)	0.5 (IQR=4.0)
Tobacco	**General**	**Specialist**	**General**	**Specialist**	**General**	**Specialist**	**General**	**Specialist**	**General**	**Specialist**
Any smoking during period	185 (14.1)	67 (87.0)	86 (5.7)	67 (84)	62 (4.2)	63 (74.1)	50 (3.6)	44 (73.3)	88 (6.3)	49 (83.1)
Median number of cigarettes per week (among smokers)	35.0 (IQR=75.5)	82.5 (IQR=84.0)	28.0 (IQR=49.0)	70.0 (IQR=77.5)	28.0 (IQR=66.5)	49.0 (IQR=49.0)	21.0 (IQR=67.0)	42.0 (IQR=42.0)	17.5 (IQR=52.8)	56.0 (IQR=56.0)
Illicit drug use	**General**	**Specialist**	**General**	**Specialist**	**General**	**Specialist**	**General**	**Specialist**	**General**	**Specialist**
Used cannabis during period	40 (3.1)	38 (49.4)	14 (0.9)	30 (38.0)	14 (1.0)	26 (30.6)	5 (0.4)	12 (20.0)	13 (0.9)	9 (15.3)
Used other illicit drugs	21 (1.6)	26 (34.2)	0 (0.0)	11 (13.9)	1 (0.1)	8 (9.4)	-	2 (3.3)	4 (0.3)	-

[a]Sample size is reduced as questions regarding pre- and post-awareness were introduced after the study had commenced (questions not included for n=221). Also excludes 8 women who had no pre-awareness data as they were reportedly aware of their pregnancy immediately.

[b]Post-awareness data was not available for 18 women who did not know they were pregnant in Trimester 1.

Appendix B – Ethics Approval

Macquarie University Ethics Secretariat
Research Office
Research Hub, Building C5C East
MACQUARIE UNIVERSITY NSW 2109

Phone **+61 (0)2 9850 4194**
Fax +61 (0)2 9850 4465
Email ethics.secretariat@mq.edu.au

07 February 2014

Dr Melanie Porter
Department of Psychology
Faculty of Human Sciences
Macquarie University
NSW 2109

Dear Dr Porter

Re: The Bayley Scales III - A Quantitative Analysis for Application in an Australian Population (REF: 5201400106)

The above application was considered by the Executive of the Human Research Ethics Committee (Medical Sciences) (HREC (Medical Sciences)).

In accordance with Chapter 5.3 of the *National Statement on Ethical Conduct in Human Research* (2007) the Executive noted the approval from Sydney Local Health District Ethics Review Committee (RPAH Zone) and your right to proceed under their authority.

Please do not hesitate to contact the Ethics Secretariat should you have any questions.

The HREC wishes you every success in your research.

Yours sincerely

[redacted]

Professor Tony Eyers
Chair, Macquarie University Human Research Ethics Committee (Medical Sciences)

cc. Ms Ingrid Honan, combined PhD/Master of Clinical Neuropsychology Candidate, Department of Psychology, Macquarie University

This HREC is constituted and operates in accordance with the National Health and Medical Research Council's (NHMRC) *National Statement on Ethical Conduct in Human Research* (2007) (the National Statement) and the CPMP/ICH *Note for Guidance on Good Clinical Practice*.

ADDRESS FOR ALL CORRESPONDENCE
RESEARCH DEVELOPMENT OFFICE
ROYAL PRINCE ALFRED HOSPITAL
CAMPERDOWN NSW 2050

TELEPHONE: (02) 9515 6766
FACSIMILE: (02) 9515 7176
EMAIL: [redacted]
REFERENCE: X12-0232 & (prev. X08-0127) & HREC/08/RPAH/218
9.132/DEC13

Health
Sydney
Local Health District

25 November 2013

Professor R Mattick
C/- Ms M Gomez
National Drug & Alcohol Research Centre
King Street
UNIVERSITY OF NEW SOUTH WALES
NSW 2052

Dear Professor Mattick,

Re: Protocol No X12-0232 & (prev. X08-0127) & HREC/08/RPAH/218 - "Impact of parental substance use on infant development and family functioning (The Triple B Study: Bumps, Babies and Beyond)"

Thank you, on behalf of the Ethics Review Committee, for Ms I Honan's correspondence of 18 June 2013 and Dr M Porter's emailed correspondence of 22 November 2013.

The participation of Dr M Porter (Macquarie University) in the above study as Ms I Honan's supervisor is noted and approved.

Yours sincerely,

[redacted]

Lesley Townsend
Executive Officer
Ethics Review Committee (RPAH Zone)

HERC\EXECOR\13-11

ADDRESS FOR ALL CORRESPONDENCE
RESEARCH DEVELOPMENT OFFICE
ROYAL PRINCE ALFRED HOSPITAL
CAMPERDOWN NSW 2050

TELEPHONE: (02) 9515 6766
FACSIMILE: (02) 9515 7176
EMAIL: [redacted]
REFERENCE: X11-0111 & HREC/11/RPAH/153
9.33/MAY14

Health
Sydney
Local Health District

17 April 2014

Dr D Hutchinson
National Drug and Alcohol Research Centre
King Street
UNIVERSITY OF NEW SOUTH WALES NSW 2052

Dear Dr Hutchinson,

Re: Protocol No X11-0111 & HREC/11/RPAH/153 - "Impact of parental substance use on family functioning and child development: Wave III follow-up of preschoolers"

The Executive of the Ethics Review Committee, at its meeting of 17 April 2014 considered Professor R Mattick's correspondence of 24 March 2014 and gave its approval of the following:

- Information for Participants (Master Version 7, 24 March 2014)
- Mother Consent Form (Master Version 7, 24 March 2014)
- Wave III Mother Interview (Version 3, 24 March 2014)

The removal of the following components of the study is noted:

- Cheek swabs for genetic testing
- Adult Attachment Questionnaire (AAQ)
- ASQ-3 36 Month Questionnaire (© 2002)
- ASQ-SE 36 Month / 3 Year Questionnaire (© 2002)
- Parent and Stress Index (© 1995)
- Wave III Partner Indirect Interview
- Wave III Partner Questionnaire
- Wave III Partner Interview
- Wave III Mother Questionnaire
- Partner Consent Form
- Family Functioning Assessment using the Emotional Availability Scale (EAS)
- Strange Situation Paradigm

Yours sincerely,

[redacted]

Lesley Townsend
Executive Officer
Ethics Review Committee (RPAH Zone)

HERC\EXECOR\14-05

ADDRESS FOR ALL CORRESPONDENCE
RESEARCH DEVELOPMENT OFFICE
ROYAL PRINCE ALFRED HOSPITAL
CAMPERDOWN NSW 2050

TELEPHONE: (02) 9515 6766
FACSIMILE: (02) 9515 7176
EMAIL: [redacted]
REFERENCE: X08-0127 & HREC/08/RPAH/218
5.0/9.63/SEP11

Health
Sydney
Local Health District

26 August 2011

Professor R Mattick
C/- Ms M Gomez
National Drug and Alcohol Research Centre
King Street
UNIVERSITY OF NEW SOUTH WALES NSW 2052

Dear Professor Mattick,

Re: Protocol No X08-0127 & HREC/08/RPAH/218 - "Impact of parental substance use on infant development and family functioning (The Triple B Study: Bumps, Babies and Beyond)"

Thank you, on behalf of the Ethics Review Committee, for your correspondence of 23 August 2011.

The inclusion of Ms I Honan (as a research officer) and Ms S Brann (as research student) as researchers in the above study is noted and approved.

Yours sincerely,

[redacted]

Lesley Townsend
Executive Officer
Ethics Review Committee (RPAH Zone)

HERC\EXECOR\11-09

General Correspondence
PO Box M30
Missenden Road, NSW, 2050
Email: slhn.esu@sswahs.nsw.gov.au
Website: www.health.nsw.gov.au/sydlhn/

Sydney Local Health District
ABN 17 520 269 052
Level 11 North, King George V Building
83 Missenden Rd
CAMPERDOWN, NSW, 2050
Tel 612 9515 9600 Fax 612 9515 9610

Appendix C – Information and Consent Triple B Study

The Triple B Study: Bumps, Babies and Beyond

INFORMATION FOR PARTICIPANTS

Introduction

You, your partner and your baby (when he/she is born) are invited to take part in a research study examining infant development and family functioning (the Triple B Study). The objective is to investigate the impact of factors such as parental substance use and psychological health, social support and health care access on infant development and family functioning. This is the first large-scale Australian study to examine these issues. The study will improve knowledge of these effects to direct public health and treatment initiatives that improve the health and well-being of Australian children and families.

The study is being conducted by the National Drug and Alcohol Research Centre at the University of New South Wales, and the National Drug Research Institute at Curtin University of Technology, Perth.

The researchers involved in this project include:

- Professor Richard Mattick, National Drug and Alcohol Research Centre, University of New South Wales, Sydney
- Dr Delyse Hutchinson, Research Fellow at the National Drug and Alcohol Research Centre, University of New South Wales, Sydney
- Professor Steve Allsop, Director at the National Drug Research Institute, Curtin University of Technology, Perth
- Professor Jake Najman, Director at the Queensland Alcohol and Drug Research and Education Centre, University of Queensland, Brisbane
- Professor Elizabeth Elliott, Discipline of Paediatrics and Child Health, University of Sydney, Australia
- Dr Lucy Burns, Lecturer at the National Drug and Alcohol Research Centre, University of New South Wales, Sydney
- Dr Sue Jacobs, Visiting Medical Officer in Obstetrics and Gynaecology at Royal Prince Alfred Hospital Women and Babies, Camperdown, and the Mater Hospital, Crows Nest
- Dr Craig Olsson, Murdoch Children's Research Institute, Melbourne
- Dr Anne Bartu, Curtin University of Technology, Perth

Study Procedures

If you agree to participate in this study, you will be asked to sign the Participant Consent Form. You will then be asked to participate in the following procedures over a 12 month period:

1. You will be asked to invite your partner to participate in the study. It will not affect your participation in the study if you choose not to invite your partner to participate, or if you do not have a partner.

2. [Mothers only] During your first and second trimester, you will also be asked to complete a brief 15-minute telephone interview about your first and second trimester. This interview will contain demographic questions, questions related to substance use and how you have been feeling in the first and second trimester.

3. [Mothers and partners] You and your partner will be asked to complete an interview and self complete questionnaires during your third trimester [baseline interview]. If you are already in your third trimester this interview will also incorporate the first and second trimester interviews. This interview will take approximately 1-1½ hours to complete, and will ask questions relating to demographics, substance use, general health and psychological well-being, health problems associated with pregnancy and your relationship with each other.

4. [Mothers and partners] You will be asked to complete an interview after you have had your baby. This interview will again contain questions regarding your wellbeing, including substance use, mental and physical health, antenatal care received and birth outcomes. The mother interview is approximately 45 minutes in length, and the partner interview takes approximately 20 minutes.

5 [Mothers, partners and infant] We would also like to collect a genetic sample from you, your partner and your baby. A simple cheek swab is used to collect DNA. We are collecting DNA to see if genes influence how babies react to events during their first year of life. We also want to see if events during pregnancy change the way genes work. This will help us identify those mothers who may need special care during pregnancy. Our work may also lead to new ways of repairing faulty genes in the future. If you would not like to participate in this component of the research please indicate in the appropriate place on the consent form.

6 [Mothers and partners] You and your partner will be asked to complete the final follow-up interview when your baby is 12 months old. This interview will be similar in content to the baseline interview. Additionally, an infant development and an observational assessment will be conducted. The final assessments will take approximately 3 hours to complete (for mothers approximately 2 hours – typically split across two different days, according to what works best for you and 1 hour for partners). This includes the interview, infant development assessment and observational component. The observational assessments will be videotaped to allow for accurate coding by an experienced researcher. The observational assessment will not be conducted if you do not wish to be recorded. You will still receive the results from the infant development assessment.

7 [Infants only] At this time we would also like to collect another cheek sample from your baby for genetic testing. This will allow us to investigate how events in the first year of life change the way genes work. This information holds considerable promise for finding new ways of treating common conditions of childhood in the future.

8 It is not possible for us to provide individual genetic results for you or your family. However, a summary report (that does not disclose individual results) will be sent to all study participants on completion.

9 [Mothers and infants] You will be asked to allow researchers to have access to your medical record to obtain information relevant to this study. The researchers

will only record information on antenatal care and health outcomes. No identifying information will be recorded.

10 [Mothers and partners] We may ask you for permission to record one or more of your interviews or developmental assessment on a Dictaphone (voice) recorder, or video camera. This is for quality and training purposes only, to ensure that interviews are being conducted correctly. We will record 40% of participants, all of whom will be chosen at random. If used, the video camera will be focused on the interviewer only: you and your infant will not be visible. If you do not wish the interview to be recorded, you can decline. We will not record interviews without your permission.

11 [Mothers only] We may ask you for a confidential urine drug screen at your baseline (first) interview. This is to compare reliability between the survey and biological sampling. We are asking 5% of participants to do this and people are selected at random. If you do not wish to provide this sample, you may decline. If you do provide a sample, the results will be identified by a study number only (not your name or other identifying information) and will only be used by the project team. Results will not be provided to your health service provider.

12 [Mothers and partners]. Should we receive funding to continue the study into the future, we would like to contact you again to invite you to participate. Participation, as with all aspects of this study, would be entirely voluntary and you would be under no obligation to participate.

Risks

There is minimal risk associated with participation in this study. Some of the interview questions involve sensitive topics; you may as a result experience emotional distress when answering these questions. If any particular question makes you uncomfortable, you may decline answering the question. We do not expect there to be any risks or side-effects from giving a cheek swab.

Benefits

The results of this study will probably be of no benefit to you. We do, however, hope that you gain value from the opportunity to participate in research that helps understand the role of infant development and family functioning. You will also receive a report on your infant's developmental assessment (the Bayley Scales of Infant and Toddler Development) results at 12 months, which includes feedback and recommendations, and is normally only available privately.

Costs

Participation in this study will not cost you anything, and you will be reimbursed for your time and out-of-pocket expenses. You will receive $50 (two payments of $25) upon completion of the first interview, $25 on completion of the 8 week interview, and $40 upon completion of the last interview and infant development assessment when your child is aged 12 months. If both you and your partner are involved you will both receive this amount.

Voluntary Participation

Participation in this study is entirely voluntary. You do not have to take part in it. If you do take part, you can withdraw at any time without having to give a reason.

Whatever your decision, please be assured that it will not affect your medical treatment or your relationship with the staff who are caring for you.

Confidentiality
All information collected for the study will be treated confidentially, and only the researchers involved will have access to it. The study results may be presented at a conference or in a scientific publication, but individual participants will not be identifiable.

There is a very unlikely possibility of a breach of confidentiality regarding the information that you provide us. Researchers may be required to report a serious crime or if they suspect harm to yourself or to another person. All the information will be stored with a number that we assign to you. Your interview will only be identified by this number. Your contact details will be kept separately in a locked filing cabinet or a password protected file that is only accessible by the researchers.

Please note that DNA samples will only be identified by a participant number, ensuring complete confidentiality of participant information. These de-identified samples will be stored in a secure DNA Data Bank and analysed in accordance with strict ethical guidelines set by the University of New South Wales, and the Sydney South West Area Health Service Human Ethics Review Committees.

Data from the study may be reused for research purposes; however all participant details will remain confidential, and data will only be used for studies relating to psychological and behavioural health.

Further Information
When you have read this information, the interviewer will discuss it with you and answer questions you may have. If you would like to know more at any stage, please contact Dr Delyse Hutchinson or Dr Richard Mattick on (02) 9385 0333. Alternatively you may email the project staff on antenatalstudy@unsw.edu.au if you have questions at any time.

This information sheet is for you to keep.

Ethics Approval
This study has been approved by the Ethics Review Committee (RPAH Zone) of the Sydney South West Area Health Service. Any person with concerns or complaints about the conduct of this study should contact the Secretary on 02 9515 6766 and quote protocol number X08-0127.

The conduct of this study has been authorised by the South Western Sydney local health Network. Any person with concerns or complaints about the conduct of the study may also conduct the Research Governance officer on (02) 9612 0614 and quote the project number 11’064.

The Triple B Study: Bumps, Babies and Beyond

PARTICIPANT CONSENT FORM

I, ... [name]

of.....[address]

have read and understood the Information for Participants on the above named research study and have discussed the study with .....

I have been made aware of the procedures involved in the study, including any known or expected inconvenience, risk, discomfort or potential side effect and of their implications as far as they are currently known by the researchers.

I understand that my participation in this study will allow the researchers to have access to my medical record, and I agree to this.

I freely choose to participate in this study and understand that I can withdraw at any time.

I also understand that the research study is strictly confidential and any information will be disclosed only with my permission, except as required by law.

I hereby agree to participate in this research study.

Please tick the appropriate responses in the box below

	YES	NO
I would like to receive the results of the infant developmental assessment, which will be conducted when my child is 12 months old	☐	☐
I am happy to be contacted in the future regarding future participation in this study. I understand that my participation, as with all aspects of this study, would be entirely voluntary and that I would be under no obligation to participate.	☐	☐
If I am randomly selected to have my interviews recorded, I agree to this. I understand that this is for quality and training purposes.	☐	☐
I would like to participate in the epigenetic component of the study	☐	☐
(mothers only) If I am randomly selected to provide a confidential urine drug screen, I agree to this.	☐	☐

NAME: _____

SIGNATURE: _____

DATE: _____

NAME OF WITNESS: _____

SIGNATURE OF WITNESS: _____

Infant development and family functioning

PARENT / GUARDIAN CONSENT FORM

I, ... *[name of parent/guardian]*

of .....*[address]*,

parent/guardian of ……*[name of child]*

have read and understood the Information for Parent/Guardian on the above named research study and have discussed the study with .....

I have been made aware of the procedures involved in the study, including any known or expected inconvenience, risk, discomfort or potential side effect and of their implications as far as they are currently known by the researchers.

I understand that participation in this study will allow the researchers to have access to my child's medical record, and I agree to this.

I freely choose to allow my child to participate in this study and understand that I can withdraw him/her at any time.

I understand that the developmental assessment of my infant will be videotaped, and I agree to this.

I understand that I may be asked permission by the interviewer to record the interview for quality and training purposes, and I agree to this.

I also understand that the research study is strictly confidential and any information will be disclosed only with my permission, except as required by law.

I hereby agree to my child's participation this research study.

NAME OF PARENT/GUARDIAN:.....

SIGNATURE:…….....

DATE:.....

NAME OF WITNESS:.....

SIGNATURE OF WITNESS:.....

Appendix D – Information and Consent 3-year Follow-Up

The Triple B Study: Bumps, Babies and Beyond

Preschool Development Follow Up

INFORMATION FOR PARTICIPANTS

Introduction

You and your child are invited to continue your involvement in the Triple B Study examining child development and family functioning. The objective of the Triple B Study is to investigate the impact of factors such as parental substance use and psychological health, family social support, childcare quality and health care access on child development and family functioning. The study will improve knowledge of these effects to direct public health and treatment initiatives that improve the health and well-being of Australian children and families.

We invite you to participate in a PhD project which is an extension of the Triple B Study. This study will assess the expected developmental trajectories of Australian children, and will be conducted in collaboration with Macquarie University and the National Drug and Alcohol Research Centre at the University of New South Wales.

The researchers involved in this project include:

- Dr Delyse Hutchinson, Research Fellow at the National Drug and Alcohol Research Centre, University of New South Wales, Sydney
- Professor Richard Mattick, National Drug and Alcohol Research Centre, University of New South Wales, Sydney
- Professor Steve Allsop, Director at the National Drug Research Institute, Curtin University of Technology, Perth
- Professor Ann Sanson, Department of Paediatrics, The University of Melbourne
- Professor Elizabeth Elliott, Discipline of Paediatrics and Child Health, University of Sydney, Australia
- Dr Lucy Burns, Senior Lecturer at the National Drug and Alcohol Research Centre, University of New South Wales, Sydney
- Dr Sue Jacobs, Visiting Medical Officer in Obstetrics and Gynaecology at Royal Prince Alfred Hospital Women and Babies, Camperdown, and the Mater Hospital, Crows Nest
- Dr Craig Olsson, Senior Research Fellow, Murdoch Children's Research Institute, Melbourne
- Dr Anne Bartu, School of Nursing and Midwifery, Curtin University of Technology, Perth
- Ingrid Honan, PhD candidate at Macquarie University, Sydney.

Study Procedures

If you agree to participate in this study, you will be asked to sign the Participant Consent Form. You will then be asked to participate in the following procedures:

1. You will be asked to complete an interview and a self-complete questionnaire when your child is three years old. These will involve responding to questions relating to substance use; health and psychological adjustment and your child's health and developmental progress.

2. A child development assessment will also be conducted. The assessment and interview combined will take approximately 1.5 hours to complete. You will receive the results from the child development assessment.

3. [Mothers and child]. You will be asked to allow researchers to have access to your medical record to obtain information relevant to this study. The researchers will only record information on antenatal care and health outcomes. No identifying information will be recorded.

4. Should we receive funding to continue the study into the future, we would like to contact you again to invite you to participate. Participation, as with all aspects of this study, would be entirely voluntary and you would be under no obligation to participate.

Risks
There is minimal risk associated with participation in this study. Some of the interview questions involve sensitive topics; you may as a result experience emotional distress when answering these questions. If any particular question makes you uncomfortable, you may decline answering the question.

Benefits
The results of this study will probably be of no benefit to you. We do, however, hope that you gain value from the opportunity to participate in research that helps understand the role of infant development and family functioning. You will also receive a report on your child's developmental assessment (the Bayley Scales of Infant and Toddler Development) results at 36 months, which includes feedback and recommendations, and is normally only available privately at a significant cost.

Costs
Participation in this study will not cost you anything, and you will be reimbursed for your time and out-of-pocket expenses. You will receive $30 upon completion of the interview and development assessment.

Voluntary Participation
Participation in this study is entirely voluntary. You do not have to take part in it. If you do take part, you can withdraw at any time without having to give a reason. Whatever your decision, please be assured that it will not affect your medical treatment or your relationship with staff who may be caring for you or your child.

Confidentiality

All information collected for the study will be treated confidentially, and only the researchers involved will have access to it. The study results may be presented at a conference or in a scientific publication, but individual participants will not be identifiable.

There is a very unlikely possibility of a breach of confidentiality regarding the information that you provide us. Researchers may be required to report a serious crime or if they suspect harm to yourself or to another person. All the information will be stored with a number that we assign to you. Your interview will only be identified by this number. Your contact details will be kept separately in a locked filing cabinet or a password protected file that is only accessible by the researchers.

Data may be reused for research purposes; however all participant details will remain confidential. Data will only be used for studies relating to psychological and behavioural health. No other investigations will be undertaken without your written consent.

Further Information

When you have read this information, the interviewer will discuss it with you and answer questions you may have. If you would like to know more at any stage, please contact Dr Delyse Hutchinson on (02) 9385 0333 or Ingrid Honan (02) 93850382. Alternatively you may email the project staff on antenatalstudy@unsw.edu.au if you have questions at any time.

This information sheet is for you to keep.

Ethics Approval

This study has been approved by the University of New South Wales Human Research Ethics Committee.

This study has also been approved by the Ethics Review Committee (RPAH Zone) of the Sydney Local Health District. Any person with concerns or complaints about the conduct of this study should contact the Executive Officer on 02 9515 6766 and quote protocol number X11-0111.

Triple B Study

MOTHER CONSENT FORM

1.I,_____of_____
age _____ years agree to participate in the study described in the participant information sheet.

2. I acknowledge that I have read the Participant Information Sheet, which explains why I have been selected, the aims of the study and the nature and the possible risks of the investigation, and the statement has been explained to me to my satisfaction.

3. Before signing this Consent Form, I have been given the opportunity to ask any questions relating to any possible physical and mental harm that my child or I might suffer as a result of participation. I have received satisfactory answers to any questions that I have asked.

4. I understand that I can withdraw from the study at any time, without prejudice to my relationship to the University of New South Wales.

5. I understand that one of the named researchers will contact me via the phone numbers I have provided to arrange a time for my child and I to attend the three year assessment. I understand that I will be reimbursed for my time and out of pocket expenses following completion of the assessment.

6. I understand that research data gathered from the results of the study may be published, provided that I cannot be identified.

7. I understand that all information that I give in this study is completely anonymous and confidential, except as required by law.

8. I understand that my participation in this study will allow the researchers and others, as described in the Information for Participants, to have access to me and my child's medical record, and I agree to this.

9. I understand that if I have any questions relating to my participation in this research, I may contact either Dr Delyse Hutchinson (02 9385 0333) or Ingrid Honan (02 9385 0382) who will be happy to answer them.

10. I acknowledge receipt of a copy of this Consent Form and the Subject Information Statement.

11. If I have any complaints or concerns about any ethical aspect of this study, I have been provided details of whom these may be directed to.

_____	_____
Signature of participant	Signature of witness
_____	_____
Please PRINT name	Please PRINT name
_____	_____
Date	Nature of witnes

Appendix E – Interview Questions

SECTION A: Demographics (asked at every time point)

A1. What is your date of birth? _____

A2. How old are you now? _____ yrs

A3. Are you Aboriginal and/or Torres Strait Islander?

No	☐ 0
Yes	☐ 1
Not stated	☐ 9

A4. In which country were you born?

Australia	☐ 0
England	☐ 1
Italy	☐ 2
Greece	☐ 3
New Zealand	☐ 4
Vietnam	☐ 5
Scotland	☐ 6
Other	☐ 7 (*Please specify*_____)

A5. What is your highest level of education?

Did not complete Year 10	☐ 0
Completed Year 10	☐ 1
Completed Year 12	☐ 2
Completed Tafe/technical	☐ 3
Completed University/college	☐ 4

A6. What is your **current employment status**?

☐ 1 Full time employment
☐ 2 Part time/casual employment
☐ 3 Unemployed (pension, unemployed)
☐ 5 Student
☐ 6 Home duties
☐ 9 Other: (*Please specify*_____)

A9. Where are you currently living *(tick one only):*

Rented house or flat	☐ 1
Privately owned house or flat	☐ 2
Staying with family/friends	☐ 3
Other (*please specify*_____)	☐ 4

A10. What is your current marital status?

Never married	☐ 1
Widowed	☐ 2

Divorced	☐ 3
Separated but not divorced	☐ 4
Married	☐ 5

A11. Are you currently living with your partner?
Yes……☐ 1 No ……..☐ 0

A12. Is your current partner biologically related to this baby?
☐ 0 No, he/she is not biologically related
☐ 1 Yes, he is the father
☐ 2 *(same sex couples)* Yes, she donated the egg

C14. How many biological children do you have? _____ (if no children skip to *non-biological children* below)

Now I'm going to ask you some questions about your drug use. I'll emphasize again that the information you give me is confidential in the same way that the rest of the interview is.

D1. Are you currently in any type of drug treatment?
☐ 1 Not in treatment
☐ 2 Methadone
☐ 3 Buprenorphine
☐4 Detoxification
☐ 5 Therapeutic community /Residential Rehab
☐ 6 Narcotics anonymous
☐ 7 Drug counselling
☐ 8 Other (specify _____)
(includes nicotine replacement; SMART recovery, etc)

D2. How long have you been in your current treatment? _____ months

D3. What is your current dose? _____ (specify in mgs)

8 WEEKS- SECTION F: ABOUT YOUR INFANT

F1. What gender is your baby? (code silently if known) ..... Male…….. ☐0 Female ……☐1

F2 What is your infant's D.O.B? _____

F4. How many babies did you have? E.g. twins (code silently if known)

1……☐ 1
2……☐ 2
3+……..☐ 3

F5 How many weeks gestation was your child when he/she was born? _____ weeks

F6 What was your baby's birth weight? (from blue book where possible) _____ kgs

F7. What was your baby's length at birth? (from blue book where possible) _____ cms

F8. What was your baby's head circumference at birth? (from blue book where possible) _____ cms

F9. What Apgar score did your baby receive at 1 and 5 minutes? (from blue book where possible) a)_____ at 1 min and b) _____ at 5 minutes

F10. What was the outcome of your infant's hearing (SWISH) test (from blue book where possible)? Pass…….☐ 1 Fail.....☐0 Don't know/not completed...☐ 1

10a. Referral to audiologist? No……..☐ 0 Yes…..☐ 1

F11. Was your baby transferred to a special care nursery or neonatal intensive care unit?

No ☐0
Yes, special care nursery ☐1
Yes, neonatal intensive care nursery ☐2

a. If yes, please provide reasons for transfer, length of stay and discharge status_____
_____

F12 Did your baby require oxygen after the birth?

No ☐0
Yes ☐1 (If yes, record details of how much, for how long and at what pressure, if known) _____
_____
_____

1-YEAR- SECTION C: Children

C1. Since the birth of your baby have you had another child?
No……..☐ 0Yes……..☐ 1

C2. Have FACS (previously DOCS) had any intervention in your family's life over the last 12 months?
No……..☐ 0Yes……..☐ 1

C3. Are you aware of any reports to FACS (previously DOCS) being made over the last 12 months?
No……..☐ 0Yes……..☐ 1

G1. Infant weight ..... _____kgs
..... Interviewer please note where measurement came from: blue book ☐ 1 interviewer administered ☐ 2
..... How old exactly was the infant when this measurement was taken? _____ months
.....Was the infant weighed with or without clothes: with clothes ☐ 1 without clothes ☐ 2

G2. Infant height ..... _____cms
..... Interviewer please note where measurement came from: blue book ☐ 1 interviewer administered ☐ 2
..... How old exactly was the infant when this measurement was taken? _____ months

G3. Infant head circumference _____cms
..... Interviewer please note where measurement came from: blue book ☐ 1 interviewer administered ☐ 2
..... How old exactly was the infant when this measurement was taken? _____ months

G7. Does your child currently need or use medicine prescribed by a doctor, other than vitamins?
No……..☐ 0 (skip to next question) Yes……..☐ 1

a) If yes, Is this because of any medical, behavioural or other health condition?
No……..☐ 0Yes……..☐ 1

b) Is this a condition that has lasted or is expected to last for at least 12 months?

No……..☐ 0Yes……..☐ 1

G12. Does your child have any of these ongoing problems?

☐ 0 None
☐ 1 Hearing problems
☐ 2 Eyes or seeing properly
☐ 3 Developmental delay
☐ 4 Eczema
☐ 5 Diarrhoea or colitis
☐ 6 Anaemia
☐ 7 Ear infections
☐ 8 Other infections
☐ 9 Food or digestive allergies
☐ 10 Other illnesses
☐ 11 Other physical disabilities

G13. Since birth, how many times has your child been hurt, injured or had an accident and needed medical attention from a doctor or hospital? _____ times (If 0 skip to next question)

a) What types of injury or accident did child have that needed medical attention?

Broken or fractured bones ☐ 1
Burn or scald ☐ 2
Dislocation ☐ 3
Sprain or strain ☐ 4
Cut or scrape ☐ 5
Concussion or internal head injury ☐ 6
Internal injury (not head) ☐ 7
Dental injury ☐ 8
Accidental poisoning ☐ 9
Other (please specify _____) ☐ 10

b) Has your child stayed in hospital for at least one night because of any (of these) injuries or accidents?

No……..☐ 0
Yes…….☐ 1

G14. Not including when he/she was born, how many times has your child stayed in hospital for at least one night for any reason? (NOT HOSPITAL OUTPATIENT OR EMERGENCY DEPARTMENT)
_____ times

G15 For what main reason?

Fever or viral illness ☐ 1
Asthma …... ☐ 2
Gastroenteritis ☐ 3
Pneumonia ☐ 4
Bronchiolitis ☐ 5
Urine infection ☐ 6
Croup…….. ☐ 7
Febrile convulsion ☐ 8
Grommets/tympanostomy tube ☐ 9
Tonsillectomy and/ or adenoidectomy ☐ 10
Other illness, surgery not needed ☐ 11
Other illness/condition, surgery needed ☐ 12

J1. Have you experienced any difficulties in relation to your infant that we have not already covered, for example with feeding or sleeping? If so, please give details of these and what you did to try and address these, and whether it worked for you.

_____

_____

_____

_____

_____

3-YEAR SECTION A: Demographics

A1. What is your **current employment status**?

- ☐ 1 Full time employment
- ☐ 2 Part time/casual employment
- ☐ 3 Unemployed (pension, unemployed)
- ☐ 5 Student
- ☐ 6 Home duties
- ☐ 9 Other: (*Please specify*_____)

A2. What is the *total (before tax)* of all wages/salaries, government benefits, pensions, allowances and other income the HOUSEHOLD *usually* receives?

Combined household	Mother

☐ 1 $2400 or more per week ($124,800 or more per year)
☐ 2 $2200 - $2399 per week ($114,400 - $124,799 per year)
☐ 3 $2000 - $2199 per week ($104,000 - $114,399 per year)
☐ 4 $1500 - $1999 per week ($78,000 - $103,999 per year)
☐ 5 $1000 - $1499 per week ($52,000 - $77,999 per year)
☐ 6 $800 - $999 per week ($41,600 - $51,999 per year)
☐ 7 $700 - $799 per week ($36,400 - $41,599 per year)
☐ 8 $600 - $699 per week ($31,200 - $36,399 per year)
☐ 9 $500 - $599 per week ($26,000 - $31,199 per year)
☐ 10 $400 - $499 per week ($20,800 - $25,999 per year)
☐ 11 $300 - $399 per week ($15,600 - $20,799 per year)
☐ 12 $200 - $299 per week ($10,400 - $15,599 per year)
☐ 13 $100 - $199 per week ($5,200 - $10,399 per year)
☐ 14 $50 - $99 per week ($2,600 - $5,199 per year)
☐ 15 $1 - $49 per week ($1 - $2,599 per year
☐ 16 Nil income
☐ 17 Negative income

A2a). What was the *total (before tax)* of all wages/salaries, government benefits, pensions, allowances and other income the HOUSEHOLD *usually* received when you were pregnant?

Combined household	Mother

☐ 1 $2400 or more per week ($124,800 or more per year)
☐ 2 $2200 - $2399 per week ($114,400 - $124,799 per year)
☐ 3 $2000 - $2199 per week ($104,000 - $114,399 per year)
☐ 4 $1500 - $1999 per week ($78,000 - $103,999 per year)
☐ 5 $1000 - $1499 per week ($52,000 - $77,999 per year)
☐ 6 $800 - $999 per week ($41,600 - $51,999 per year)
☐ 7 $700 - $799 per week ($36,400 - $41,599 per year)
☐ 8 $600 - $699 per week ($31,200 - $36,399 per year)
☐ 9 $500 - $599 per week ($26,000 - $31,199 per year)
☐ 10 $400 - $499 per week ($20,800 - $25,999 per year)
☐ 11 $300 - $399 per week ($15,600 - $20,799 per year)
☐ 12 $200 - $299 per week ($10,400 - $15,599 per year)
☐ 13 $100 - $199 per week ($5,200 - $10,399 per year)
☐ 14 $50 - $99 per week ($2,600 - $5,199 per year)
☐ 15 $1 - $49 per week ($1 - $2,599 per year
☐ 16 Nil income
☐ 17 Negative income

A3. Where are you currently living *(tick one only):*

Rented house or flat ☐ 1
Privately owned house or flat ☐ 2
Staying with family/friends ☐ 3
Other (*please specify*_____) ☐ 4

A4. Are you currently living with your partner? Yes ☐ 1 No ☐ 0

A5. What is your current marital status?

Never married ☐ 1
Widowed ☐ 2
Divorced ☐ 3
Separated but not divorced ☐ 4
Married ☐ 5

C2. How many biological children do you have now? _____

C3. What is the birth order of your child (child involved in this study)? _____

SECTION G: Your child's health and nutrition

Interviewer to record details below

G1. Child weight ..... _____kgs
..... Interviewer please note where measurement came from: blue book ☐ 1 interviewer administered ☐ 2
..... How old exactly was the child when this measurement was taken? _____ months
.....Was the child weighed with or without clothes: with clothes ☐ 1 without clothes ☐ 2

G2. Child height..... _____cms
..... Interviewer please note where measurement came from: blue book ☐ 1 interviewer administered ☐ 2
..... How old exactly was the child when this measurement was taken? _____ months

G3. Child head circumference _____cms
..... Interviewer please note where measurement came from: blue book ☐ 1 interviewer administered ☐ 2
..... How old exactly was the child when this measurement was taken? _____ months

G6. In general, how would you say your child's current health is?
☐ 1 Excellent
☐ 2 Very good
☐ 3 Good
☐ 4 Fair
☐ 5 Poor

G7. Does your child currently need or use medicine prescribed by a doctor, other than vitamins?
No ☐ 0 (skip to next question) Yes ☐ 1

a) If yes, Is this because of any medical, behavioural or other health condition?
No ☐ 0 Yes ☐ 1

b) Is this a condition that has lasted or is expected to last for at least 12 months?
No ☐ 0 Yes ☐ 1

c) Does your child need or use more medical care than is usual for most children of the same age?
No ☐ 0 (skip to next question) Yes ☐ 1

d) Is this because of any specific medical, behavioural or other health condition? (NOT JUST COLDS)
No ☐ 0 Yes ☐ 1

e) Is this a condition that has lasted or is expected to last for at least 12 months?
No ☐ 0 Yes ☐ 1

G12. In the past 2 years, has your child had any of these ongoing problems?
☐ 0 None
☐ 1 Hearing problems
☐ 2 Eyes or seeing properly
☐ 3 Developmental delay
Please specify(e.g. autism, Asperger's) _____
☐ 4 Eczema
☐ 5 Diarrhoea or colitis
☐ 6 Anaemia
☐ 7 Ear infections
☐ 8 Other infections
Please specify _____
☐ 9 Food or digestive allergies
☐ 10 Other illnesses
Please specify _____
☐ 11 Other physical disabilities
Please specify _____

G13. In the past 2 years, how many times has your child been hurt, injured or had an accident and needed medical attention from a doctor or hospital? _____ times (If 0 skip to next question)

a) What types of injury or accident did child have that needed medical attention?
Broken or fractured bones ☐ 1
Burn or scald ☐ 2

Dislocation	☐ 3
Sprain or strain	☐ 4
Cut or scrape	☐ 5
Concussion or internal head injury	☐ 6
Internal injury (not head)	☐ 7
Dental injury	☐ 8
Accidental poisoning	☐ 9
Other (please specify _____)	☐ 10

b) Has your child stayed in hospital for at least one night because of any (of these) injuries or accidents?

No ☐ 0⇨skip to hearing screener
Yes ☐ 1⇨continue

G14. In the past 2 years, how many times has your child stayed in hospital for at least one night for any (other) reason? (NOT HOSPITAL OUTPATIENT OR EMERGENCY DEPARTMENT) _____ times

G15 For what main reason?

Fever or viral illness	☐ 1
Asthma	☐ 2
Gastroenteritis	☐ 3
Pneumonia	☐ 4
Bronchiolitis	☐ 5
Urinary tract infection	☐ 6
Croup.....	☐ 7
Febrile convulsion	☐ 8
Grommets/tympanostomy tube	☐ 9
Tonsillectomy and/ or adenoidectomy	☐ 10
Other illness, surgery not needed	☐ 11
Other illness/condition, surgery needed	☐ 12

Appendix F- Journal Submission Receipts

STUDY 1

Dear Mrs Honan:
Thank you for submitting your manuscript entitled "**A comparison between Australian infant performance and United States normative data at 1-year on the Bayley Scales of Infant and Toddler Development III.**," received on July 31, 2017. Your manuscript has been assigned the following manuscript number: PAM17-1631.

Please refer to the manuscript number in all communications. You may check the status of this manuscript by selecting the Check Manuscript Status link on the following Web page:

https://manuscripts.jamapeds.com/cgi- (Concealed to protect privacy as per below)
bin/main.plex?el=A1h4ISR3A1nHO6F5A9ftdzpSHn86N1EEG60ibs1EbVwZ

Do not share this encrypted link with others, as it will automatically log you into your account for JAMA Pediatrics's Web-based system.

* We agree to consider your manuscript with the understanding that its content, figures, and tables have not been published or submitted elsewhere in print or electronic format and will not be submitted elsewhere during the period of review by JAMA Pediatrics.
* If you have not already done so, please provide copies of any manuscripts on closely related topics or with possibly duplicative material that have been previously published or are under consideration for publication elsewhere.
* The information in your manuscript should not be distributed or released in hard copy or electronic form, except through presentation at scientific meetings, unless and until the manuscript is published.
* The fact that your manuscript is under consideration by JAMA Pediatrics is confidential and should not be disclosed to anyone except coauthors and contributors.

Your manuscript will be reviewed by an editor here, and possibly by two or more peer reviewers. Every effort will be made to expedite the review process and to notify you of our decision as soon as possible.

If accepted for publication, all major articles, including reports of research, reviews, and opinion pieces, will be published either Online First (before print) or Online Only to ensure that publication can occur quickly.

Sincerely,

Frederick P. Rivara, MD, MPH
Editor, JAMA Pediatrics

STUDY 3

Dear Ms. Honan,

Your manuscript entitled 'A comparison between Australian and United States normative data at 3-years of age on the Bayley Scales of Infant and Toddler Development-3rd Edition (BSID-III).' has been successfully submitted online to the Australian and New Zealand Journal of Psychiatry.

Your manuscript ID is ANP-2017-00369.

Please mention the above manuscript ID in all future correspondence or when contacting the Editorial Office. If there are any changes to your contact details, please log in to Manuscript Central at and edit your user information as appropriate. You can also view the status of your manuscript at any time by checking your Author Center.

Thank you for submitting your manuscript to the Australian and New Zealand Journal of Psychiatry.

Sincerely,
ANZJP Editorial Office

STUDY 4

Dear Mrs Honan,

Your manuscript entitled "Predictive Utility of the Bayley Scales of Infant and Toddler Development- 3rd Edition (BSID-III) from 1-year to 3-years of age." has been successfully submitted online and is presently being given full consideration for publication in Child Neuropsychology.

Your manuscript reference ID is CNY-OA 17-81.

Please mention the above manuscript ID in all future correspondence. If there are any changes in your street or e-mail addresses, please log in to ScholarOne Manuscripts at https://mc.manuscriptcentral.com/ncny and edit your user information as appropriate.

We attempt to have all reviews completed within two to three months of your submission being received, however, due to various factors it is not always possible to complete the reviewing procedure within that timescale. You may view the status of your manuscript at any time by checking your Author Centre after logging in to the website.

Please also find attached an Article Publishing Agreement that we ask corresponding authors to read through for information. The purpose of sending this form to you now is so that you may see what terms and conditions will apply on the acceptance of your paper, should that be the end result of the reviewing process. There is no need to send it back to us now. In the event of your paper being accepted we will send you another copy.

I will be in contact to inform you if your paper is sent to an action editor for reviewing, and I will keep you updated on the progress of your paper through the reviewing process though should you have any questions or concerns, at any stage of the reviewing process, please don't hesitate to contact me.

Thank you again for submitting your manuscript to Child Neuropsychology.

Sincerely,
Joey Anito
Editorial Assistant