Pathways of Care Longitudinal Study: Outcomes of Children and Young People in Out-of-Home Care

Measuring Child Developmental Outcomes: Approaches and Methods

BillyBlack

Pathways of Care Longitudinal Study: Outcomes of Children and Young People in Out-of-Home Care in NSW

Technical Report No. 9

Measuring Child Developmental Outcomes:
Approaches and Methods

Contents

List of figures

List of tables

Preface

The Pathways of Care Longitudinal Study (POCLS) is funded and managed by the New South Wales Department of Family and Community Services (FACS). It is the first large-scale prospective longitudinal study of children and young people in out-of-home care (OOHC) in Australia. Information on safety, permanency and wellbeing is being collected from various sources. The child developmental domains of interest are physical health, socio-emotional wellbeing and cognitive/learning ability.

The overall aim of this study is to collect detailed information about the life course development of children who enter OOHC for the first time and the factors that influence their development. The POCLS objectives are to:

- describe the characteristics, child protection history, development and wellbeing of children and young people at the time they enter OOHC for the first time.

- describe the services, interventions and pathways for children and young people in OOHC, post restoration, post adoption and on leaving care at 18 years.

- describe children's and young people's experiences while growing up in OOHC, post restoration, post adoption and on leaving care at 18 years.

- understand the factors that influence the outcomes for children and young people who grow up in OOHC, are restored home, are adopted or leave care at 18 years.

- inform policy and practice to strengthen the OOHC service system in NSW to improve the outcomes for children and young people in OOHC.

The POCLS is the first study to link data on children's child protection backgrounds, OOHC placements, health, education and offending held by multiple government agencies; and match it to first-hand accounts from children, caregivers, caseworkers and teachers. The POCLS database will allow researchers to track children's trajectories and experiences from birth.

The population cohort is a census of all children and young people who entered OOHC for the first time in NSW over the 18 month period between May 2010 and October 2011 (n=4,126). A subset of those children and young people who went on to receive final Children's Court care and protection orders by April 2013 (2,828) were eligible to participate in the study. For more information about the study please visit the study webpage www.facs.nsw.gov.au/resources/research/pathways-of-care.

The POCLS acknowledges and honours Aboriginal people as our First Peoples of NSW and is committed to working with the FACS Aboriginal Outcomes team to ensure that Aboriginal children, young people, families and communities are supported and empowered to improve their life outcomes. The POCLS data asset

will be used to improve how services and supports are designed and delivered in partnership with Aboriginal people and communities.

FACS recognises the importance of Indigenous Data Sovereignty (IDS) and Indigenous Data Governance (IDG) in the design, collection, analysis, dissemination and management of all data related to Aboriginal Australians. The POCLS is subject to ethics approval, including from the Aboriginal Health & Medical Research Council of NSW. FACS is currently in the process of scoping the development of IDS and IDG principles that will apply to future Aboriginal data creation, development, stewardship, analysis, dissemination and infrastructure. The POCLS will continue to collaborate with Aboriginal Peoples and will apply the FACS research governance principles once developed.

1 Introduction

This paper aims to provide researchers using the POCLS data some guidelines on the use of methods and approaches in their analysis, interpretation and reporting of children's developmental outcomes. The paper provides a summary of some of the advantages of age standardised measures being used in the POCLS to maximise measurement equivalence over time.

For POCLS, in-depth data is collected from the caregivers of participant children who went on to Final Orders and agreed to be interviewed (Wave 1 n = 1285; Wave 2 n = 1200; Wave 3 n = 1033; Wave 4 n = 962). The caregivers of children under the age of six years completed standardised outcome measures in relation to the children in their care. Children aged six years or older in this group were also interviewed and completed selected standardised tests. This paper pertains to the group of interviewed children for whom standardised measures were administered.

2 Use of standardised measures to assess developmental outcomes

The primary concern of any statutory child protection agency is to keep children safe from abuse and neglect. It follows that any examination of the impact of child protection interventions should include a measure to reflect subsequent child safety[1]. However, this is not the only concern. As children who have been removed from their parents' care are generally more likely to experience poor developmental outcomes, child protection agencies also have a responsibility to ensure that their actions do not have unintended negative consequences that would compromise developmental outcomes for these already vulnerable children, as well as to ensure that interventions and supports are appropriate and lead to better developmental outcomes. For this reason, the POCLS has included standardised measures at baseline, and at each subsequent wave, corresponding to the major developmental domains of cognitive learning ability, socio-emotional adjustment and physical development to examine outcomes for children and young people aged 9 months to 17 years.

Critical to the strength of the study is the robustness of the outcome measures selected. The confidence in the conclusions drawn about the differential effects of child protection history and OOHC experience is closely linked to the reliability and validity of these measures. The standardised developmental outcome measures for the POCLS study were selected on the basis of the strength of their psychometric properties, their wide use in other studies, and their ability to draw comparisons with the general population and other groups (see Appendix 1 for an overview of all measures used[2])[3].

Two main issues arise in building a coherent set of published studies with the use of multiple measures in the POCLS:

1. Multiple measures of a particular domain are used for some children in a single wave because measures of specific domains often overlap by age. For instance, children aged 2-3 years olds have multiple sources of data on their

---

[1] See appendix 2 for a discussion of safety measures for the POCLS.

[2] For more details, see the POCLS Measures Manual, which is a compilation of all the outcome measures (for both children and carers) that are collected in the POCLS. It provides detailed information about each measure, such as the outcome domain that the measure falls into, age range of children, for which wave data was collected, scoring information and references. In terms of children's wellbeing, there are sixteen standardised outcome measures across different outcome domains and age groups in the Measures Manual. The Manual can be found on the POCLS study website https://www.facs.nsw.gov.au/resources/research/pathways-of-care).

[3] A majority of these measures have also been used in other large scale longitudinal studies such as the Canadian National Longitudinal Study of Children and Youth (NLSCY) and the longitudinal Study of Australian Children (LSAC) enabling research comparisons with general population data.

verbal ability as their verbal ability has been assessed using, the MacArthur-Bates Communicative Development Inventories (MCDI-III), the Communication and Symbolic Behaviour Scales (CSBS) and the Ages and Stages Questionnaire (ASQ) Communication Scale (Fenson et al., 2007; Wetherby & Brizant, 2003; Squires & Bricker, 2009).

2. When development in a particular domain is compared for the same children over time using different tests. These tests often use different population cut-offs to signify need for services (as specified by the producers of the test). An example would be comparing the ASQ results (where the manual suggests those more than two standard deviations below the general population mean are considered to be in need for referral for professional support in language development) with the Peabody Picture Vocabulary Test (PPVT) (where the manual indicates that those one standard deviation below the general population mean are vulnerable).

At the time of writing, the POCLS has completed four waves of data collection and Wave 5 is currently underway. Consideration is being given to exploring the feasibility of deriving a composite developmental outcome index with scaling properties harmonised across the age ranges assessed at each wave. If we do not equate or harmonise the scales, compring the scores over time to understand change in the construct of interest presents some difficulties.

In the meanwhile, this discussion paper aims to identify a small subset of standardised outcome measures used in the POCLS which would allow for the effective comparison of developmental outcomes not only across studies but also over time.

The process of selecting this smaller subset of measures was to:

- examine the validity and reliability of each measure

- select those with the most robust psychometric properties ensuring that all domains and ages were covered

- align means and distributions for each measure with normed general population scores

- select cut-offs to define levels of developmental status (e.g., typical versus atypical).

This selection provides POCLS data users with confidence that the indicator measures have closely equivalent underlying constructs, robust psychometric properties, and appropriate coverage for different age ranges.

This paper presents two possible approaches in terms of determining cut-offs to allow more reliable comparisons between studies and over time: the established cut-offs from each measure's author(s) versus consistent cut-offs based on means and standard deviation (SD) measures. The pros and cons of each approach are discussed. The recommendations of this paper are not prescriptive and researchers can use any approach that suits the purpose of their research. However, when other measures or cut-off points are used, the method needs to be clearly explained in any resulting research report.

3 Rationale for the selection of the POCLS standardised measures subset

This section provides detailed information about the rationale for the selection of measures in each outcome domain (i.e., socio-emotional, cognitive and physical development).

Socio-emotional development

Measure of socio emotional development selected for the POCLS subset:

- BITSEA[4] behaviour problem score based on percentile rank (Wave 1: 9-36 months)

- CBCL[5] total behaviour problem T score (Wave 1: 3-17 years; Wave 2 onwards: 1.5-17 years).

Rationale

The two main measures administered at baseline (Wave 1) to measure socio-emotional development were the BITSEA for children aged 9 to 35 months and the CBCL for children aged 3 to 17 years. The Personal-Social Scale of the ASQ was also administered (one of five scales of the ASQ) but only provides a standardised cut-off rather than generating standardised scores or percentiles for all raw scores. As a result scores remain raw and increases may only reflect maturation. Briggs-Gowan et al. (2004) pointed out that, although the Person-Social Scale has acceptable test-retest reliability and sensitivity, more information is needed with regard to its sensitivity to specific types of socio-emotional disorders and its validity over the full five-year range.

On the other hand the BITSEA and the CBCL scales both generate a total behaviour problem score and a social competence score.[6] All children in the study age range (9 months to 17 years) therefore have a behaviour problems score and a social competence score either on the BITSEA or the CBCL.

The BITSEA behaviour problem score is standardised to percentile ranks. It has been shown to correlate with the longer 1.5-5 year CBCL scale. The latter accurately detects 80 to 95 per cent of toddlers identified as having social-emotional problems.

---

[4] BITSEA – Brief Infant Toddler Social Emotional Assessment
[5] CBCL – Child Behaviour Checklist
[6] The CBCL Social Competence scale is made up of three subscales (Activities, Social and School Scales). The CBCL Total Behaviour Problem Scale is made up of an Internalisation Problem scale, an Externalisation Problem scale and Other Problems Scale.

It also predicts the CBCL behaviour problem score one year later (correlation = 0.71; Briggs-Gowan, Carter, Irwin, Wachtel & Cochetti, 2004). Generally the BITSEA problem scale has stronger psychometric properties than the BITSEA social competence scale (Briggs-Gowan et al , 2004).

Similarly the CBCL behaviour problem scale classifies children more accurately in terms of their later social emotional adjustment than the CBCL social competence scale (Achenbach & Rescorla, 2001). It has also been more commonly used in longitudinal studies where children's developmental courses have been tracked to adulthood (Achenbach & Rescorla, 2001).

Adding weight to the decision to choose the CBCL behaviour problem scales over the CBCL social competence scales was the fact that the CBCL social competence scale was not asked in relation to 3-6 year olds in this study. The CBCL social competence scale is also not available for children aged 6-11 at Wave 1 due to a program/administrative error. The total behaviour problem scores have thus been selected from each measure to maximise reliability, validity, and available data for analysis.

Cognitive development

Cognitive measures can be divided into verbal and non-verbal development. If a child was determined to be developing atypically on either verbal ability or non-verbal reasoning they were considered to be needing some support. Verbal ability is particularly important for school success.

Rationale for verbal ability

Measures of <u>verbal ability</u> selected for the POCLS subset:

- CSBS total composite standard score for less than two years (Wave 1 only, with limited validity due to age).

- MCDI-III vocabulary percentile rank for 24 to 35 months (Wave 1 and Wave 2)

- PPVT standardised IQ deviation score for 36 months and over (Wave1 onwards)

For children under three years of age, measurement of verbal ability relies on carer report. Direct assessment at this age is expensive and has low reliability and validity so that parent or carer reports are considered to be a more accurate and cost-effective way of measuring ability and change over time (Fenson et al. 2007). Even so, when devising the MacArthur-Bates Development Inventories (MDI-III) for children Fenson et. al. (2007) pointed out that "due to the enormous variability in language and communicative development in infants and toddlers, it is virtually impossible to obtain a definite diagnosis of specific language impairment in the first

three years of life ….. and identification is likely to be more solid when a child is 4 years old than three years old" (p.40). Researchers should keep this limited reliability and validity in mind when interpreting analyses that include verbal ability scores for children aged three and under.

For children aged between 9 and 23 months, the CSBS was used to assess communication skills and the more detailed MDCI-III was administered for children aged 24-35 months. The ASQ communication scale was also administered to all children as part of a broader ASQ assessment for children aged 9 to 35 months. As mentioned, the ASQ age-related standardisation is limited to cut-off points, which makes change over time difficult to assess except in relation to cut-off points.

Bearing in mind the limitations of standardisation, the total CSBS standardised score was selected for infants under two years of age while recognising the very young age (9-23 months) of the children involved. The MCDI-III, which is standardised for age, was selected for toddlers aged 24 to 35 months.[7]

For children aged 3 to 17 years, a single verbal test was administered and thus selected to measure verbal ability; specifically, the well-standardised and widely used PPVT.

Rationale for non-verbal reasoning ability

Measures of <u>non-verbal ability (reasoning)</u> selected for the POCLS subset:

- ASQ problem solving scale standardised cut-off points for under 66 months (Wave 1 onwards)

- Matrix Reasoning (WISC-IV) standard score for 6-17 years (Wave 1 onwards).

The problem solving scale of the ASQ was used to measure non-verbal ability for children aged between 9 and 66 months. It is limited by the standardisation related to a specific cut-off rather than generating a percentile rank or standardised score for each raw score. This limits analysis to above or below the various age related cut-offs for the ASQ.

The WISC-IV for children aged 6 to 17 years is a direct measure, well-standardised and widely used. There are five months between the ages of 67-71 months that are

---

[7] The Pearson correlation between the CSBS and the ASQ communication scale was 0.59 (n=366; p=0.000). For children over 24 months the Macarthur-Bates vocabulary scales (MCDI-III) also correlated strongly with the ASQ communication scale (between 24-29 months n=83, correlation =0.57, p=0.000; 30 to 35 months, n=77, correlation = 0.36 p=0.001).

not covered by either of these tests, so numbers may be reduced in any analysis that includes children in this age bracket.

Physical development

Measure of physical development selected for the POCLS:

- ASQ fine motor scale 65 months or less (standardised cut-offs)

- ASQ gross motor scale 65 months or less (standardised cut-offs).

Rationale

While the measurement of motor development often dominates the assessment of babies and toddlers, after the age of five or six the focus of developmental assessment shifts to cognitive and social domains. Measurement of developmental motor ability after this age requires a battery of observations, extended time to administer, and assessors who are qualified in physical assessment. Assessments can include variables such as running speed, agility, response speed, reaction times and visual-motor ability that are costly to measure. Physical development was therefore only assessed in young children, as levels of fine and gross motor co-ordination have important influences on children's ability to explore their surroundings at this developmental stage. The ASQ is the only standardised measure of physical development used in this study, and is therefore the key outcome measure for this domain.

4 The POCLS standardised measures subset

The subset of standardised measures that have been selected for the POCLS is summarised below (see also Figure 1). These measures are selected by outcome domain and age, so that each domain is measured across the entire age range.

Socio-emotional development

- Brief Infant Toddler Social Emotional Assessment (BITSEA) Behaviour Problem Score based on percentile rank (Wave 1: 9-24 months)

- Child Behaviour Checklist (CBCL) Total Behaviour Problem T score (Wave 1: 3-17 years; Wave 2 onwards: 1.5-17 years)

Cognitive ability – verbal

- Communication and Symbolic Behaviour Scale (CSBS) total composite standard score (Wave 1: 9-23 months)

- MacArthur Bates Communicative Developmental Inventories (MCDI-III) vocabulary percentile rank (Wave 1 and 2: 24-35 months)

- Peabody Picture Vocabulary Test (PPVT) IQ deviation score (Wave 1 onwards: 3-17 years)

Cognitive ability – non-verbal

- Ages and Stages Questionnaire (ASQ) Problem Solving Scale standardised cut-off points (Wave 1 onwards up to 66 months)

- Matrix Reasoning Test standard score - MR WISC-IV (Wave 1 onwards aged six years or over)

Physical development

- ASQ Fine Motor Scale standardised cut-off points (Wave 1 onwards: up to 66 months).

- ASQ Gross Motor Scale standardised cut-off points (Wave 1 onwards: up to 66 months)

Figure 1: Selected measures for child developmental outcome domains by age

OUTCOME DOMAINS	9-35 mths	3-5 yrs	6-11 yrs	12-17 yrs
Physical development	ASQ (Fine and Gross Motor Scales)	ASQ	-	-
Socio-emotional development	BITSEA (W1) CBCL (W2) (Behaviour Problem Scale)	CBCL	CBCL	CBCL
Cognitive development - non verbal	ASQ (Problem-Solving Scale)	ASQ	MR-WISC	MR-WISC
Cognitive development - verbal	CSBS MCDI-III	PPVT	PPVT	PPVT
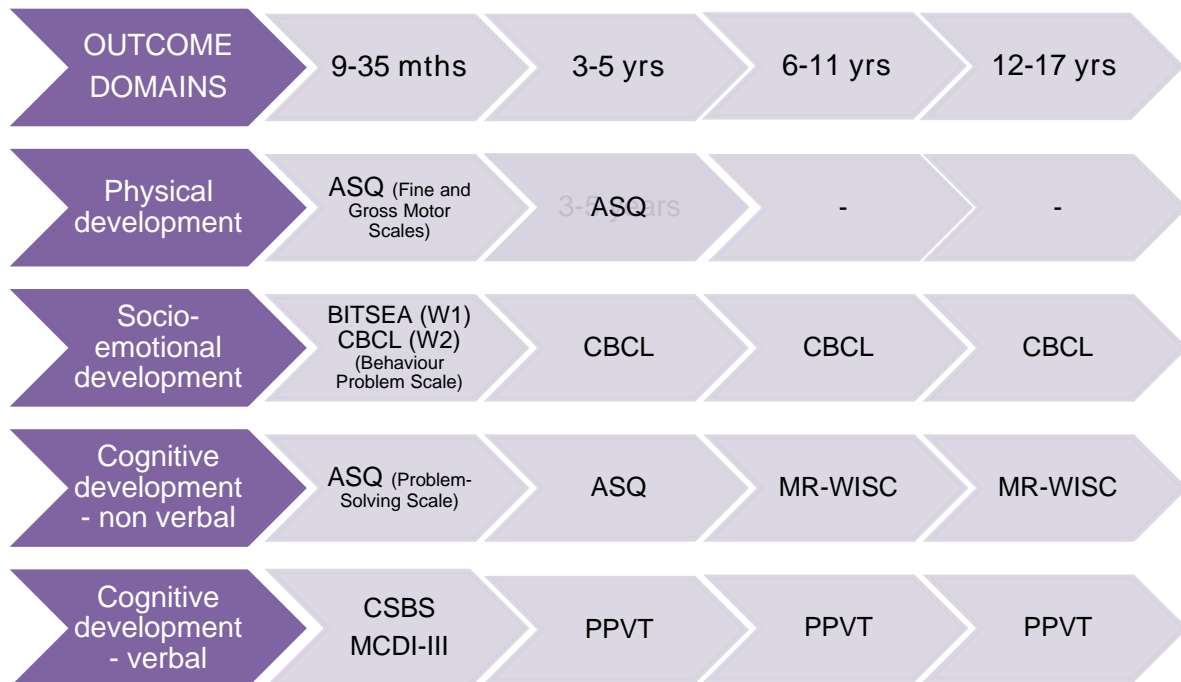
5 Determining cut-offs to define levels of developmental status

Different measures adopt different cut-off points to define vulnerability. In general, the purpose of an assessment may influence the stringency of the cut-off point. Where the measure's aims are preventative such as identifying children who may be at risk of later difficulties, the cut-off points may be more liberal than when measures are used to identify eligibility for immediate services. For analyses of the POCLS data, the specific research question may also determine the cut-off points adopted.

When researchers use measures consistently across a range of research questions, it permits more conclusive interpretations of results than if a variety of cut-off points are used. It is however understood that some research questions will require more stringent or liberal cut-off points. In these cases the cut-offs should be clearly stated.

It should be noted that the screening tests in relation to the younger children:

- are less reliable and valid as a function of the limited predictability of developmental outcomes at such young ages.

- have less exact cut-off points often incorporating a wide range of raw scores or the same score can reflect many percentile ranks . For example, a score of 14 on the BITSEA behaviour problem scale at 12-17 months is considered to be any percentile from the 15th to 24th percentile rank.

- do not necessarily standardise all scores. For example, the ASQ is normed across 21 age ranges with the raw scores only standardised in terms of cut-offs at the 1, 1.5 and 2 standard deviations of each age range. The raw scores themselves are not standardised and include the effects of maturation. When using the ASQ and looking at change over time relative to the general population cut-offs should be used.

The established cut-off approach

This approach uses the established cut-offs, which are defined by the author of each measure as discussed below.

Established cut-off points for the levels of socio-emotional development

CBCL: The CBCL Total Behaviour Problem Score is standardised by age and gender, generating a T score with a mean of 50 and a standard deviation of 10. According to the CBCL user manual, a child is considered to be:

- in the normal range if they have a T score within 1 standard deviation of the general population mean (i.e. <60)

- borderline/vulnerable if they have a T score between 1 and 1.3 standard deviations above the general population mean (i.e. 60-63), and

- in the clinical range if they have a T score more than 1.3 standard deviations above the mean (i.e. ☐64).

BITSEA: The BITSEA provides standardisation in terms of percentile ranks, using the 96th percentile or above (score of 1), the 91st to 95th percentile (score of 2), 86th to 90th percentile (score of 3), 76th to 85th percentile (score of 4), 75th percentile (score of 5) and 74th percentile or lower (score of 6). A child is considered to be:

- in the normal range if the score is below the 75th percentile range

- in the possible problem range if the score is in the most difficult 25% (i.e., at the 75th percentile rank or above).

To align the CBCL and BITSEA, the borderline and clinical ranges of the CBCL can be collapsed so that both measures can be formed into a single binary variable ('normal/typical' and 'at risk/atypical') covering everyone (1-17 years) in the interview cohort.

Established cut-off points for levels of verbal cognitive development

CSBS: The CSBS provides standard scores and percentiles, with percentile scores at or below the 10th percentile considered of concern. Hence, the normative cut-off for CSBS is:

- in the typical range if the percentile scores are above the 10th percentile

- in the atypical range or of concern if the percentile scores are at or below the 10th percentile.

MCDI: The cut-off for MCDI based on percentile scores is:

- in the typical range if the percentile scores are at or above the 15th percentile

- in the atypical range or of concern if the percentile scores below the 15th percentile.

PPVT: The mean standard score for the normative sample[8] is 100 and the standard deviation is 15. The normative cut-offs for the PPVT, therefore, are:

- above the normal range if the standard scores are above 115

- within the normal range if the standard scores are between 85-115

---

[8] This is referred to as the US normative sample.

- below the normal range if the standard scores are below 85.

As with the socio-emotional domain, there are only two possible categories from the established cut-offs for the young children. To align CSBS, MCDI and PPVT, the above normal and normal ranges of the PPVT can be collapsed so that the three measures can be formed into a single binary variable (with two categories being 'normal/typical' and 'at risk/atypical') covering everyone (9 months – 17 years) in the interview cohort.

Established cut-off points for levels of non-verbal reasoning ability

ASQ: The ASQ has standardised cut-off points. According to the ASQ user manual, the normative cut-offs for the ASQ are:

- in the typical range if scores are within one standard deviation below the mean

- need monitoring or follow-up if scores are between one and two standard deviations below the mean

- Referral for professional support if scores are below two standard deviations below the mean.

WISC: The WISC has the normative mean of 10 with a standard deviation of 3. The normative cut-offs, therefore, are:

- above the normal range if the standard scores are above 13

- in the normal range if the standard scores are between 7-13

- below the normal range if the standard scores are less than 7.

To align ASQ and WISC, the above normal and normal ranges of the WISC can be collapsed into one category (i.e., being 'typical'). Similarly, the 'monitoring/follow-up' and 'referral' categories of the ASQ can be combined into one category (i.e., being 'atypical'). A single binary variable ('normal/typical' and 'at risk/atypical') can then be created to cover everyone (9 months – 17 years except for 67-71 months) in the interview cohort.

Established cut-off points for levels of physical development

ASQ: The normative cut-offs for the ASQ fine motor and gross motor scales are the same as above.

Consistent cut-offs based on means and standard deviations

Although the established cut-off approach aims to align different measures across the entire age range in a domain, different measures often use different normative cut-offs to signify level of vulnerability or need for services (as specified by the producers of the test), which might be based on different standard deviations from the general population

mean. Hence, using consistent cut-off points based on means and standard deviations is valuable where consistency is sought to provide comparability between studies, over time, or across measures for the same child.

The suggested consistent cut-off points based on means and standard deviations are:

- up to one standard deviation from the mean to categorise a child's development as being typical

- more than 1 to 1.3 standard deviations to identify a child's development as being at risk and needing support

- more than 1.3 to 2 standard deviations as signifying the **'clinical' r**ange or children needing professional intervention

- more than two standard deviations from the mean as indicating that a child is in need of ongoing intensive professional support.

In the domain of socio-emotional development, the cut-off score of the BITSEA for a greater risk for later problems is seen as being in the most difficult 25% (at the 75th percentile rank or above) compared with the CBCL's one standard deviation or most difficult 16%. A further complication is that although the BITSEA norms use the most difficult 25% as the cut to indicate a risk of later problems, the raw age-related cut-off for behaviour problems for most age ranges is the same for the 15th as the 24th percentile, reflecting its status as a screening instrument. For this reason, and to align with the CBCL cut-offs, those who had a score in the highest scoring 15% of the general population were considered to be at risk of behaviour problems (i.e. a score of 3 or less) on the BITSEA. Again to align more closely (although still not perfectly) with the CBCL's use of 1.3 standard deviations as the clinical range cut-off, children were considered in the clinical range if they scored in the highest 9% of the general population (i.e. a score of 2 or less) on the BITSEA.

In the domain of cognitive verbal development, the cut-off point used for 'atypical' development was at least one standard deviation below the mean. In the general population 85% of children are seen as having 'typical' development and 15% as having atypical development. All verbal/communication measures indicated similar percentages of children in the POCLS sample who were considered to be developing 'atypically'[9].

---

[9] At baseline (Wave 1) the CSBS indicated that around 26.5% of children had communication skills that were 'of concern', which is in line with the (MDCI-III) which indicated that 28% of 30-35 months olds were around one standard deviation below the mean for their age. (The ASQ also placed 28% in the 'atypical' range). In the smaller group of children aged between 24-29 months (n=83) the MacArthur-Bates indicated that only 13% were below one standard deviation below the mean. The PPVT reflected a steady decrease in measured cognitive verbal ability with age of entry into OOHC in this sample. For children aged three to five years when they entered OOHC, 83% (compared with 85% in the general population) were considered to show 'typical' development. By 12-17 years at age of entry only 58% were showing 'typical' development. The rest were below average.

Given the general vulnerability of this population, the cut-off point for professional support was set at 1.3 standard deviations below the mean (around the 9th percentile) in line with the cut-off for a clinical category on socio-emotional development. This is less stringent than the two standard deviations (around the lowest-performing 2% of population) used to determine eligibility for funded services on the ASQ (Squires, Twomby, Bricker and Potter, 2009).

For the ASQ problem solving scale, a score of 1.5 standard deviations[10] below the mean is used as the indicator of the need for more professional support, with the general vulnerability of the population dictating a more liberal cut-off than the commonly cited two standard deviations. Similarly with the Matrix Reasoning (WISC-VI), a score of less than 7 (1 standard deviation below the mean) was considered 'atypical' and needing support and those with a score below 6 (1.3 standard deviations) were considered to need professional support[11]. This is in line with the cut-offs selected for socio-emotional development. For the analysis of the POCLS data, it is recommended that scoring more than one standard deviation below the mean on either verbal or non-verbal cognitive ability indicates vulnerability in the cognitive domain.

In the domain of physical development[12], the suggested consistent cut-offs for the ASQ are the same as stated above, that is, scores within one standard deviation below the mean indicate typical development; scores from more than one standard deviation below the mean to 1.5 standard deviations below the mean are considered to 'need monitoring', scores from more than 1.5 standard deviations below the mean to 2 standard deviations below the mean are in the clinical range, and scores below 2 standard deviations below the mean are considered to need intensive services. The established cut-offs and the consistent cut-offs can be aligned exactly because ASQ is the only measure used in this domain.

Table 1 summarises the consistent cut-off points (based on means and standard deviations) and established cut-off points to describe child development in the domains

---

[10] Noting that it is not possible to obtain the score for 1.3 standard deviations. The ASQ is a screener and only 1, 1.5 and 2 standard deviations from the mean are published.

[11] At baseline (Wave 1) on the ASQ Problem Solving Scale, 68% of the POCLS children scored within the typical range (one standard deviation below the mean or better) compared with 85% of children of the same age (9-65 months) in the general population. For children aged six and over 29% of the POCLS children (compared with 15% in the general population) were more than one standard deviation below the mean (a standard score of 7) and only 4% were more than one standard deviation above the mean on the WISC Matrices Reasoning Test.

[12] According to carer report at baseline, 73% of surveyed children showed typical and 27% atypical gross motor development. Of those showing atypical gross motor development, the majority (82%) were more than 1.5 standard deviations below the mean, indicating that they needed more professional support (i.e. 22% of all children under 5.5 years). Fine motor development also appeared compromised for many children, with more than double the percentage expected in the general population showing atypical development at Wave 1 (34% observed, versus 15% expected). Further, almost a quarter of the children aged 65 months and under were more than 1.5 standard deviations below the mean, which is about three times the rate of the general population considered to need professional support.

of socio-emotional development, cognitive learning ability and physical development in the POCLS.

Table 1: Suggested consistent cut-off points versus established cut-off points

Established cut-offs			Consistent cut-offs		
Categories by manual	Cut-off points by manual	Categories based on established cut-offs	Proposed categories	Consistent cut-off points	Standard deviation
Socio-emotional development					
CBCL					
Typical	<60	Typical	Typical	<60	1 SD
Borderline	60-63	Atypical	Borderline	60-63	1-1.3 SD
Clinical	>63		Clinical	>63-70	>1.3-2 SD
			Intensive	>70	>2 SD
BITSEA*					
Typical	6	Typical	Typical	4-6	1 SD
At risk	1-5	Atypical	Borderline	3	1-1.3 SD
			Clinical	2	>1.3-1.7 SD
			Intensive	1	>1.8 SD
Verbal cognitive development					
CSBS					
Typical	>81 (10 percentile)	Typical	Typical	85-115	1 SD
At risk	<=81	Atypical	Borderline	81-84	1-1.3 SD
			Clinical	70-80	>1.3-2 SD
			Intensive	<69	>2 SD
MCDI					
Typical	>=15 percentile	Typical	Typical	16-84 percentile	1 SD
At risk	<15 percentile	Atypical	Borderline	10-15 percentile	1-1.3 SD
			Clinical	2-9 percentile	>1.3-2 SD
			Intensive	<2 percentile	>2 SD
PPVT					
Above normal	>115	Typical			
Normal	85-115		Typical	85-115	1 SD
Below normal	<85	Atypical	Borderline	81-84	1-1.3 SD
			Clinical	70-80	>1.3-2 SD
			Intensive	<69	>2 SD

Note: * BITSEA and ASQ don't follow the consistent cut-off pointes listed above for reasons stated earlier.
     ^ The cut-offs are the same for ASQ used in the domains of non-verbal cognitive and physical development.

Table 1: Suggested consistent cut-off points versus established cut-off points (contd.)

Established cut-offs			Consistent cut-offs		
Categories by manual	Cut-off points by manual	Categories based on established cut-offs	Proposed categories	Consistent cut-off points	Standard deviation
Non-verbal cognitive development					
ASQ*^					
Typical	1 SD	Typical	Typical	-	1 SD
Monitor	1-2 SD	Atypical	Borderline	-	1-1.5 SD
Refer	>2 SD		Clinical	-	>1.5-2 SD
			Intensive	-	>2 SD
WISC					
Above normal	>13	Typical			
Normal	7-13		Typical	7-13	1 SD
Below normal	<7	Atypical	Borderline	6	1-1.3 SD
			Clinical	4-<6	>1.3-2 SD
			Intensive	<4	>2 SD

Note: * BITSEA and ASQ don't follow the consistent cut-off pointes listed above for reasons stated earlier.
^ The cut-offs are the same for ASQ used in the domains of non-verbal cognitive and physical development.

Comparison of the two approaches

There are advantages and disadvantages with each of the two approaches. These are summarised in Table 2 below.

Table 2: Pros and cons of the established versus consistent cut-off approach

Approach	Pros	Cons
Established cut-offs	• Base on each measure's manual • Tested and validated with known sensitivity/ specificity	• Different measures use different cut-offs • Binary outcome only (i.e., typical versus atypical)
Consistent cut-offs	• Use of a common set of cut-offs based on the mean and standard deviation across measures • Four outcome categories (up to five for cognitive development)	• Assuming that the means and standard deviations of the scales are the same thing, and the underlying constructs are equivalent • Not validated

The strength of the consistent cut-off approach is that its derived outcome measure has at least four categories providing more information than the established cut-off approach. However, it is based on the assumption that the distributions and underlying constructs are equivalent, which may or may not have strong empirical support. The established cut-off scores were likely developed against outcome measures, and based on population norms. Changing them may undermine the sensitivity, specificity and

validity of the categories. However, it is acknowledged that there might be situations where the use of one approach may be preferred over the other. For example, if the focus of the research is on the distinction between clinical versus borderline ranges or clinical versus intensive services, then the consistent cut-off approach may be preferred.

The impact on analysis of the two approaches was examined. A series of chi-squared tests were undertaken to examine the association of the outcomes with key child demographic variables using the POCLS Wave 1 data. Table 3 shows that the associations between the outcomes and the four selected demographic variables are consistent between the two approaches except for Aboriginality, which is significant with the established cut-off approach but not with the consistent cut-off approach.

Table 3: Associations between outcomes (based on established versus consistent cut-offs) and selected demographic variables, POCLS Wave 1

Domain and approach	Gender	Aboriginality	CALD	Age at entry to care
Verbal cognitive Established cut-offs Consistent cut-offs	n.s. n.s.	n.s. n.s.	n.s. n.s.	sig. sig.
Non-verbal cognitive Established cut-offs Consistent cut-offs	sig. sig.	n.s. n.s.	n.s. n.s.	sig. sig.
Socio-emotional Established cut-offs Consistent cut-offs	n.s. n.s.	sig. n.s.	sig. sig.	sig. sig.

Note: n.s.=not significant; sig.=Significant at p=.05.

Table 4 shows that using the different approaches affects a small number of children in the verbal cognitive and socio-emotional wellbeing domains. In the verbal cognitive domain, 17 children (1.3% of the total number of children) who are classified as being 'typical' using the established cut-off approach are classified as 'borderline' under the consistent cut-off approach. As for the socio-emotional domain, 25 children (2.1% of the total number of children) who are classified as being 'atypical' using the established cut-off approach would be classified as 'typical' using the consistent cut-off approach. There are no differences in the non-verbal cognitive and physical development domains between the two approaches.

Table 4: Classification of children using the established and consistent cut-off approaches, POCLS Wave 1

Consistent cut-offs	Established cut-offs				
	Atypical		Typical		
	n	%	n	%	Total
Verbal cognitive ability					
Above Average	0	-	166	16.6	166
Typical	0	-	817	81.7	817
Borderline	75	26.3	17	1.7	92
Clinical	145	50.9	0	-	145
Intensive	65	22.8	0	-	65
Total	285	100.0	1,000	100	1,285
Non-verbal cognitive ability					
Above Average	0	-	71	8.2	71
Typical	0	-	799	91.8	799
Borderline	99	27.3	0	-	99
Clinical	106	29.2	0	-	106
Intensive	158	43.5	0	-	158
Total	363	100.0	870	100.0	1,233
Socio-emotional wellbeing					
Typical	25	7.0	831	100.0	856
Borderline	74	20.6	0	-	74
Clinical	127	35.4	0	-	127
Intensive	133	37.0	0	-	133
Total	359	100.0	831	100.0	1,190

6 Use of the standardised measures for longitudinal data analysis

This section provides some discussion of some of practical considerations involved in the use of standardised measures in longitudinal data analysis and the modelling of developmental outcomes over time.

The availability of data on different outcome measures is one of the strengths of the POCLS. However, the use of multiple measures across domains and age also poses challenges for statistical analyses and the interpretation of the results, especially for analyses examining changes over time. For example, in the non-verbal cognitive ability domain, if a child was assessed using the ASQ Communication Scale at the age of five and assessed again using WISC at the age of 7, a change in the cognitive functioning of the child might be due, in part, to the use of different measurement instruments over time (i.e., ASQ vs. WISC). This is likely to be true regardless of which approach is adopted.

Psychological tests, through the process of standardisation, adhere to the rules of normal distribution. Scores are derived relative to the general population norms. A child's position relative to the general population can be assessed through this score. These interval-scale scores can be expressed in terms of z-scores, deviation IQ scores, T scores or scaled scores. They can also be ordinal-scale percentiles or scores based on these. The POCLS standardised data uses a number of these types of scores. The way scores relate to a normal distribution or align with scores obtained from other standardised measures can be seen in Brock (2019).

As children in the study age, the measures used are more reliable and valid. Tests like the PPVT, WISC, CBCL are standardised interval scales. Cut-off points can be used but analyses of improvement or deterioration are not restricted to change from one category to another and can be more fine-grained. However, there are practical issues that need to be considered when standardised measures are used. These are discussed in this section.

For the purposes of statistical analyses, both interval and categorical (i.e., cut-offs) forms of the measures can be used. The interval form may be preferred as the normal distribution theory can then apply and linear regression techniques employed. When cross-sectional analyses are being performed, the use of either the raw or the T scores should yield similar results as T scores are transformed raw scores[13] and statistical

---

[13] T scores are a transformation of individual raw scores into a standard form with a mean of 50 and a standard deviation of 10. The T scores for CBCL, for example, are based on a national sample of non-referred children according to their age and gender (Achenbach & Rescorla, 2001).

analyses (e.g., correlation) are not affected by the absolute magnitudes of the scores (Achenbach & Rescorla, 2001).

For example, both the raw and T scores of the CBCL Total Problem Scale are available in the POCLS data. Initial analyses show that the raw scores and T scores of the CBCL Total Problem Scale are highly correlated at each wave (e.g., Wave 1: Pearson correlation coefficient = 0.943, p=.000, n=714). Similar results were found for the other measures, such as PPVT and WISC.

Modelling developmental outcomes over time is much more complex. Traditional statistical approaches such as multivariate analysis of variance (MANOVA) or repeated measures designs (e.g., mixed ANOVA) are not suitable to use because they fail to recognise and/or address interactions between levels of data resulting from the serial correlation of measures on individuals made at multiple points in time. To address the issue of clustering in longitudinal data, statistical techniques, such as mixed effects modelling and Structural Equation Modelling (SEM) are increasingly being adopted by health and social science researchers. Singer and Willett (2003) provides a comprehensive account of the range of multilevel modelling methods now available which are suitable for exploring individual longitudinal change. One of the SEM techniques worth highlighting for its potential value in investigating trajectories of development is the so-called Growth Mixture Models or Latent Class Growth Models. This is particularly useful for identifying groups of children following different developmental paths. For example, those with more resilient, adverse, variable or other trajectories. Identifying such unobserved groups can be useful in gaining a better understanding of their specific demographic, clinical and other features.

At this stage, we don't have a single interval measure that covers the entire age range in an outcome domain, so it is not possible to model changes over time in an outcome domain for all children using interval measures. There may be many different possible approaches in looking at outcomes over times in terms of what form of measures can be used and when. Two approaches are discussed below.

One approach is to focus the analysis on a single outcome measure over time such as the CBCL. The advantages for this approach are: 1) avoids the issue of multiple measures being used to operationalise a construct over time, as discussed above; 2) allows freedom to choose either the interval scale or categorical cut-offs for examining change over time. The main disadvantage of this approach is the reduced sample size because no single measure is available for children across all ages and the numbers of children responding decreases across waves. In the case of CBCL, children aged less than 3 years at Wave 1 would be excluded[14]. In addition, there are different versions of the CBCL used for preschool children aged 1.5 to 5 years and for school children aged

---

[14] There are 567 children who were aged under 3 years at Wave 1, which accounted for 44% of the Wave 1 interview cohort (n=1,285).

6-18 years. Given the longitudinal nature of the data, many children in the POCLS will have completed both versions. For these children the use of the CBCL raw scores are not appropriate and T scores should be used instead. Changes in functioning over time, would be reflected in T scores (i.e., we would expect to see their relative position in the population distribution to change).

An alternative approach for modelling changes over time is to derive a categorical outcome variable by aligning different measures across different age groups at each wave using either the established cut-offs or the consistent cut-off points that are discussed in this paper. The established cut-off approach only allows the use of a binary outcome variable while the consistent cut-off approach has up to five proposed categories – above normal, typical, at risk/borderline, clinical and intensive support. These categories can be collapsed into three (typical, at risk/borderline, clinical) or binary (typical and atypical) depending on the research questions and/or for modelling approach. Children are grouped into these categories based on their scores at each wave. Given the discrete nature of the outcome measures, non-linear modelling approach may be employed. An advantage of this approach is that a single (categorical) measure to track changes over time is able to be constructed. A disadvantage is the loss of statistical efficiency resulting from continuous outcome scores being collapsed into ordinal or dichotomous categories. Also, the use of non-linear mixed models is quite complex and may pose additional complications in parameter estimation and interpretation. It is advisable that researchers use both approaches in their analysis if possible so that findings can be cross-validated. Another disadvantage is that children may improve or decline in scores but not change from one category to another (e.g., from clinical to borderline). This won't be picked up with the use of categorical outcome variables.

Finally, as mentioned earlier, because the ASQ raw scores are not standardised, it would be misleading if the ASQ raw scores are compared/presented over time. As children grow older, the increase in the ASQ raw scores might well be due to the effects of maturation. Therefore, when using the ASQ and looking at change over time only the ASQ cut-offs should be used.

Appendix 1

The questions and measures used in the POCLS interviews to examine children's wellbeing and caregiver and placement characteristics, including the respondent type, the age range, and availability of norms or whether used in other studies

Domain	Questions and standardised measures	Respondent type	Study age range	Used in other studies/norms available
Children's wellbeing				
Physical health and development	• Ages and Stages Questionnaire (ASQ3; Squires & Bricker, 2009)	Caregiver	9 months[1]–5 years	US norms- standardised cut-off
	• Additional questions about health conditions, services received, immunisation, diet, weight, sleep	Caregiver	All	Project developed and used by other studies such as LSAC, ATP
Child socio-emotional development	• Abbreviated Temperament Scales adapted from the Revised Infant Temperament Questionnaire (Carey & McDevitt, 1978), the Toddler Temperament Questionnaire (Fullard, McDevitt & Carey, 1978) and the Childhood Temperament Questionnaire (Thomas & Chess, 1977)	Caregiver	9 months–7 years	LSAC, ATP
	• School Aged Temperament Inventory (SATI; McClowry,1995) – short form	Caregiver	8–17 years	LSAC, ATP
	• Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan et al, 2004)	Caregiver	9–35 months	LSAC US norms
	• Child Behaviour Checklist 1.5–5 and 6–18 (CBCL; Achenbach & Rescorla, 2000; 2001)	Caregiver	3–17 years in Wave 1; All ages from Wave 2	NSCAW, LONGSCAN, US and Australian norms
	• Ages and Stages Questionnaire (ASQ3; Squires & Bricker, 2009)	Caregiver	9 months[1]–5 years	US norms
	• School Problems Scale (Prior, Sanson, Smart & Oberklaid, 2000)	Young person	–17 years	ATP
	• School Bonding Scale (O'Donnell, Hawkins & Abbott, 1995)	Young person	7–17 years	ATP, Seattle Social Development Project

Domain	Questions and standardised measures	Respondent type	Study age range	Used in other studies/norms available
	• Short Mood & Feeling Questionnaire 13-item scale (Angold et al, 1995) and additional questions on mood[2]	Young person	12–17 years	LSAC, ATP, ASSAD
	• Self Report Delinquency Scale 10-item scale adapted from (Moffitt & Silva,1988) [2]	Young person	10–17 years	ATP
	• Felt security activity to show who they feel close to (adapted from the Kvebaek Family Sculpture Technique; Cromwell, Fournier & Kvebaek, 1980).	Child/young person	7 years plus	Cashmore & Parkinson (2014) in family law study
	• Additional questions for caregivers about services and supports for child emotional and behavioural problems, problems at school, child psychotropic medication	Caregiver	All	Project developed and used by other studies such as LSAC, ATP
	• Additional questions for children and young people about peer relationships, friendships, school, health, caregivers and caseworkers	Child/young person	7 years plus	Project developed and used by other studies such as LSAC, ATP
Cognitive and language development	• Ages and Stages Questionnaire (ASQ3; Squires & Bricker, 2009)	Caregiver	9 months[1]–5 years	US norms
	• Communication and Symbolic Behaviour Scale Infant and Toddler Checklist (CSBS ITC; Wetherby & Prizant, 2003)	Caregiver	9[1]–23 months	LSAC US norms
	• MacArthur-Bates Communicative Developmental Inventories (MCDI-III; Fenson et al, 2007)	Caregiver	30–35 months	LSAC US norms
	• MacArthur Communicative Development Inventories— Short form (Fenson et al, 2000)	Caregiver	24–29 months	US norms
	• Peabody Picture Vocabulary Test (PPVT-IV; Dunn & Dunn, 2007)	Interviewer administered	3–17 years	Many studies; US norms
	• Matrix Reasoning Test from Wechsler Intelligence Scale for Children (WISC-IV; Wechsler, 2003)	Interviewer administered	6–16 years	LSAC
	• Additional questions about current schooling (usual grades at school, changes in schools, repeated years, school problems); for children aged 15 and older, questions on work and further education, life skills and plans for leaving care	Caregiver	All	Project developed and used by other studies such as LSAC, ATP

Domain	Questions and standardised measures	Respondent type	Study age range	Used in other studies/norms available
Caregiver and placement characteristics				
Caregiver psychological distress	• Kessler K10 (Kessler et al, 2003)	Caregiver	All	LSAC, NSW Health Survey, Australian norms
Social cohesion	• Social Cohesion and Trust Scale (Sampson, Raudenbush & Earls, 1997)	Caregiver	All	LSAC
Parenting practices/ style/self-efficacy	• Parenting – Warmth (Paterson & Sanson, 1999)	Caregiver	All	LSAC
	• Parenting – Hostility (Institut de la Statistique du Québec, 2000)	Caregiver	All	LSAC
	• Parenting – Monitoring (Goldberg et al, 2001)	Caregiver	10–17 years	LSAC
	• Difficult Behaviour Self-Efficacy Scale (DBSES; Hastings & Brown, 2002) • Emotional Responsiveness Scale from the Parenting Style Inventory II, adapted version (PSI-II: Darling & Toyokawa, 1997)	Caregiver Young person	2-17 7–17 years	Study by Whenan, Oxlad & Lushington (2009) LSAC
	• Additional questions for child about relationship with caregiver	Child/young person	All	Project developed and used by other studies such as LSAC, ATP
Satisfaction with support from services	• Satisfaction with Foster Parenting Inventory (SFPI) – Social Service Support Satisfaction Scale (Stockdale et al, 1997)	Caregiver	All	–
	• Additional questions for caregiver about socio-demographic characteristics; relationship with partner; relationship with study child; caregiver experience and training; family activities; support network; caregiver physical health; cultural background and cultural activities	Caregiver	All	Project developed and used by other studies such as LSAC, ATP

[1] While children will be recruited from birth onwards, an interview with their caregiver will not be conducted until the child reaches 9 months of age, to ensure that the measures of infant development are reliable.

[2] These measures were added at Wave 2.

Note: ASSAD=Australian Secondary Students' Alcohol and Drug Survey; ATP=Australian Temperament Project; LSAC=Longitudinal Study of Australian Children; LONGSCAN=Longitudinal Studies of Abuse and Neglect (US); NSCAW=National Survey of Child and Adolescent Well-Being (US).

Appendix 2

Measures of safety (summarised in the FACS_summary file):

Appendix 2 provides a brief description about the DCJ child protection and OOHC administrative data and the key measure of safety.

Child protection and OOHC history data[15] are aggregated by creating a variety of counting rules. For example, in calculating the number of placement changes a child has had, inclusion of respite care placements in the count would inflate the number and indicate a higher level of placement instability. In general, different counting rules are likely to generate different results. Comparisons will be more meaningful if the same counting rules are used between studies.

Where consistency is required when using DCJ administrative data it may be more convenient and also more reliable for researchers to use the existing summary data files rather than carrying out their own computations. The DCJ OOHC administrative data have already been collapsed and recoded in their summarised state in the OOHC_period file. Here variables can be found such as the number of placements, the number of households experienced, the number of moves, time in each type of care and the predominant type of care within a care period.

The summarised child protection variables and the summarised OOHC variables are also included in the FACS_summary file which provides child protection and OOHC information before, during, and after each care period with the child as the unit of analysis. It includes counts of reports of significant harm, counts of urgent and substantiated reports, counts by type of maltreatment and type of parental risk. In addition, there are counts of the various assessments and substantiations, care and protection outcomes as well as the summary versions of the OOHC variables. The summary variables available in each of these files, and how they were derived, can be found in the POCLS Data Users Guide.Where more detail is required or other counting rules are called for, detailed data about child protection and OOHC is available in the CP_report and the OOHC_plcmt files.

The removal of children is a last resort intervention with the primary aim of protecting them from further abuse and neglect when other less intrusive approaches have been exhausted.

---

[15] For more information, please refer to the POCLS data user guide and data dictionary for DCJ administrative data, both available on the POCLS study website https://www.facs.nsw.gov.au/resources/research/pathways-of-care.

There is no accurate direct measure of abuse and neglect. The most reliable measure available is the abuse or neglect reported to DCJ. Some studies have also used visits to hospital emergency departments for injuries and poisonings as a proxy measure of maltreatment. While these data have also been collected in this study, it represents a more indirect measure and hospital emergency departments rely on different and at times complex diagnostic coding which require additional analysis to produce comparable data sets.

If children are safer, it is reasonable to expect there should be a reduced rate of risk of significant harm (ROSH) reports to the DCJ Helpline, a reduced rate of substantiated reports, and reports should be less urgent in nature over time. These counts have already been calculated and can be found in the FACS_ summary file.

- Number of ROSH reports. This is counted before entry into OOHC (rosh_sum_A), during a care period in OOHC (rosh_sum_1, rosh_sum_2, rosh_sum_3 etc) and after a care period in OOHC (rosh_sum_B, rosh_sum_D, rosh_sum_E etc).

- Number of substantiated ROSH reports. Again, this is counted before entry into OOHC (SUBS_A), during a care period in OOHC (SUBS_1, SUBS_2 etc) and after a care period in OOHC (SUBS_B etc).

- Number of urgent reports ROSH reports before entry into OOHC (L24hour_A), during a care period in OOHC (L24hour_1, L24hour_2, L24hour_3 etc) and after a care period in OOHC (L24hour_B, L24hour_C etc).

The three variables are, however, inter-correlated with each other (range from $r=0.56$ to $r=0.66$).

> **Measures of <u>safety</u> selected for the POCLS:**
>
> The number of ROSH reports per 100 days was selected.
>
> As children have not only experienced abuse and neglect, but have also been in OOHC for varying lengths of time, the number of ROSH reports per a time period (e.g., 100 days or 1,000 days), which adjust for the different durations of abuse or time spent in OOHC, might be a better measure.[16]

---

[16] Date of entry into care (CARE_START_1) minus date of first report before entering OOHC (cntStDat_first_A) created the variable preOOHC reflecting the duration of possible abuse before entering care in days. There is an existing variable, CARE_DUR_1, for length of the first care period in care in days (based on CARE_END_1 minus CARE_START_1). So the number of reports, divided by the defined time period (e.g., 100 or 1,000 days), is the rate.

References

Achenbach, T. M., & Rescorla, L. A. (2001). Manual for the ASEBA School-Age Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Briggs-Gowan, M.J., Carter, A. S., Irwin, J.R., Wachtel, K. & Cichetti, D. (2004). The Brief Infant –Toddler Social-and Emotional Assessment: Screening for social-emotional problems and delays in competence. Journal of Pediatric Psychology, 29(2), 143-155.

Brock, S. E. Descriptive statistics and psychological testing. California State University, accessed on 28 June 2019, https://www.csus.edu/indiv/b/brocks/courses/eds%20250/eds%20250/handouts/11/descrptive%20statistics%20and%20the%20normal%20curve.pdf

Fenson, L., Marchman, V.A., Thai, D.J., Dale, P. S., Reznick, J.S. & Bates, E. (2007). MacArthur-*Bates Communicative Development Inventories. User's Guide and Technical* Manual (2nd Ed). Baltimore: Paul. H Brookes Publishing Co.

Singer, J.D., & Willett, J.B. (2003) Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. New York: Oxford University Press.

Squires, J., Bricker, D., Twombly, E., & Potter, L. (2009). ASQ-3 User's Guide. Baltimore: Paul H. Brookes. Publishing Co.

Wetherby A.M. & Prizant , B. M. (2003). CSBS DP Infant-Toddler Checklist and Easy Score *User's Guide* Baltimore: Paul H. Brookes Publishing Co.