

Accepted Manuscript

Tarren-Sweeney, M. (2017). Rates of meaningful change in the mental health of children in long-term out-of-home care: A seven- to nine-year prospective study. *Child Abuse & Neglect*, 72, 1-9. <http://dx.doi.org/10.1016/j.chiabu.2017.07.002>

Abstract

Children residing in long-term out-of-home care have high rates of clinical-level mental health difficulties. However, the stability of these children's difficulties throughout their time in care is uncertain. This paper reports estimates of the seven- to nine-year stability of carer-reported scores on the Child Behavior Checklist (CBCL) and Assessment Checklists for Children (ACC) and Adolescents (ACA) for 85 children in long-term foster or kinship care. Prospective score changes on the CBCL total problems and ACC-ACA shared-item scales were assigned to one of four change groups: 'sustained mental health'; 'meaningful improvement'; 'no meaningful change'; and 'meaningful deterioration'. On each of the two measures, more than 60% of children manifested either sustained mental health or meaningful improvement in their mental health, while less than a quarter showed meaningful deterioration. Mean mental health scores for the aggregate sample did not change over the 7-9 year period. Findings discount the presence of a uniform, population-wide effect – suggesting instead, that children's mental health follows several distinct trajectories. Rather than asking whether long-term care is generally therapeutic or harmful for the development of previously maltreated children, future investigations should focus on the questions “...*what are the systemic and interpersonal characteristics of care that promote and sustain children's psychological development throughout childhood, and what characteristics are developmentally harmful?*” and “...*for which children is care therapeutic, and for which children is it not?*”

Keywords:

Looked after children; out-of-home care; foster care; mental health; long-term stability; prospective design

Introduction

Within a family preservation framework, the primary purpose of out-of-home care is to provide maltreated children temporary protective care, with restoration to their parents being the ultimate goal. Since the 1970's it has been apparent that an increasing proportion of children placed into care either cannot, or should not be returned home. The observation that many of these children subsequently 'drift' in care without acquiring relational permanence, highlighted a concern for the developmental wellbeing of children growing up in *impermanent* out-of-home care (Fein & Maluccio, 1992; Rowe & Lambert, 1973). While this has prompted a policy shift in favour of legally permanent placements for children who cannot be safely returned to their parents, large proportions of children placed into legally impermanent out-of-home care remain thus until adulthood (Biehal, Ellison, Baker, & Sinclair, 2010). This reality raises important questions about the developmental wellbeing of children who grow up in care, and the extent to which our present models of care support or hinder children's recovery from early developmental adversity.

The most important marker of the developmental wellbeing of children growing up in long-term care is their mental health. Numerous cross-sectional studies conducted in North America, Europe, Australia and elsewhere have established that children placed in out-of-home care manifest high levels and rates of mental health difficulties (Oswald, Heil, & Goldbeck, 2010; Pecora, White, Jackson, & Wiggins, 2009). Though rates vary a little by survey and location, up to half of children in care have clinical-level mental health difficulties, and another 15% to 25% have difficulties approaching clinical significance (Tarren-Sweeney, 2008a). Furthermore, the types and patterns of difficulties experienced by

children in care and by child populations with similar psychosocial background (maltreated children, children adopted from care) differ somewhat from those observed among clinic-referred children at large. The most defining features are not the forms of their mental disturbance, but their complexity and severity (DeJong, 2010; Tarren-Sweeney, 2013a). The social and interpersonal relationship difficulties that are typically observed among children in care greatly adds to symptom complexity. We can anticipate that some attachment- and trauma-related difficulties follow a longer-term developmental course, and that even with optimal developmental conditions (sensitive, loving care) recovery from such difficulties for many children will be slow (Sonuga-Barke et al., 2017).

Surveys have consistently found that older children and young people in care tend to have more difficulties than younger children (Armsden, Pecora, Payne, & Szatkiewicz, 2000; Dubowitz, Zuravin, Starr, Feigelman, & Harrington, 1993; Heflinger, Simpkins, & Combs-Orme, 2000). This might suggest that children's mental health deteriorates in care. However, closer examination indicates that this age effect is largely accounted for by children entering care at older ages with greater mental health difficulties, due to their longer pre-care exposure to maltreatment (Hukkanen, Sourander, Bergroth, & Piha, 1999; Tarren-Sweeney, 2008b). There is nevertheless the possibility that older age at entry into care not only accounts for poorer mental health at the time of entry into care, but also moderates children's response to care. In other words, the therapeutic potential of care may vary according to such factors as age at entry into care, the extent that children's psychological development is compromised at entry into care, and the strength of children's pre-care attachments to their birth families.

Beyond these cross-sectional data, what do we know about the mental health trajectories of children growing up in care? A recent series of meta-analyses pooled prospective mean score changes in externalizing difficulties (21 studies, combined N = 1,729), internalizing difficulties (24 studies, combined N = 1,984), and total difficulties i.e.

global mental health (25 studies, combined N = 2,523) from all prospective studies completed to date (Goemans, van Geel, & Vedder, 2015). These meta-analyses showed no mean effect over time. Various moderator analyses failed to show effects when comparing studies on study length, sample size, publication type, attrition, or mean age. Instead, the three meta-analyses identified considerable heterogeneity across the various study findings, with some studies reporting large mean increases in mental health scores over time, and others reporting large reductions (Goemans et al., 2015).

Most of the prospective studies published to date followed children over relatively short time periods (most were six months to two years). These timeframes are too short to predict longer-term mental health trajectories within this population. Only four studies have tracked the mental health of children in care over periods of five or more years. The first of these estimated children's mental health over five years from social worker reports (Fanshel & Shinn, 1978; Frank, 1980). However, it is doubtful that social workers have sufficient proximal engagement with children in care to be reliable informants of their mental health, and there is no research supporting the validity of this method. A second study compared five-year changes in the mental health of orphans growing up in foster care versus those placed in children's homes (Bulat, 2010). However, the pre-care development of orphans is not comparable to that of seriously maltreated children. It is important to keep in mind that the effects of growing up in care are not shaped in isolation from the exceptional developmental context of pre-care maltreatment experienced by the vast majority of children placed into care in western, democratic jurisdictions. A third study recruited children shortly after entry into care, but obtained baseline mental health scores from the children's parents rather than their foster carers, and then 7-8 year follow-up scores from their foster carers, residential carers or parents (depending on whether they remained in foster care, moved to residential care, or returned to their parents' care) (Havnen, Breivik, & Jakobsen, 2014). In the absence of inter-

rater reliability estimates, the reliability of prospective estimates of the stability of children's mental health over time based on parent- and foster carer-reported scores at different time points is uncertain. Furthermore, the prospective cohort included children who had left family-based care, such that the findings aren't specific to children growing up in care. The fourth study obtained baseline carer-reported Child Behavior Checklist (CBCL) (Achenbach & Rescorla, 2001) scores for a large sample (n=233) of children in care, and 8-year follow-up measures for 111 of the young people (48% retention), by which time many of the participants were adults (Vis, Handegård, Holtan, Fossum, & Thørnblad, 2016). Carer-reported CBCL scores were obtained for 38 young people under 18 who were still in care. Instead of reporting prospective mean score changes for these 38 young people, the study reported rates of meaningful change in CBCL total problems scores, as defined by the Reliable Change Index (RCI): 26% (n=10) showed meaningful improvement, 26% (n=10) showed meaningful deterioration, and 47% (n=18) showed no meaningful change.

This brief overview suggests that as yet we have insufficient knowledge of the stability of children's mental health as they grow up in long-term alternate care. Nevertheless, both developmental theory and developmental psychopathology research would predict that these children experience a range of mental health trajectories, due to heterogeneity in: 1. children's pre-care exposure to severe social adversity; 2. children's pre-care attachment development; 3. children's age at entry into long-term care; 4. various characteristics of foster or kinship caregiving; and 5. social care systemic pressures on children's felt security. We should also anticipate that transactional interactions between each of these sets of developmental influences will increase variance in children's mental health trajectories.

On this basis, while it remains useful to establish whether long-term care is generally therapeutic or harmful, the presence of a homogenous effect (as measured by mean score changes, and a unimodal distribution of score changes) is unlikely. The present paper seeks to

expand on this limited knowledge, by reporting seven- to nine-year changes in mental health estimates from a prospective, epidemiological study carried out in the Australian state of New South Wales (NSW). The analyses address two related research questions – *Does long-term out-of-home care have a homogenous effect on the stability of children’s mental health? And, do children residing in long-term care manifest different mental health trajectories over the course of their time in care?*

Method

The Children in Care Study (CICS) is a prospective epidemiological study of the mental health and developmental risk exposure of children in long-term foster and kinship care, in New South Wales (NSW), Australia. It consists of a baseline cross-sectional survey of 4-11 year-olds conducted between November 1999 and January 2003 (N=347); a follow-up survey of those participants conducted in 2009 (N=85); and a cross-sectional adolescent survey of 11-17 year-olds (N=232). The latter included the 85 follow-up survey participants, and 147 newly-recruited adolescents.

A large number of demographic, developmental, mental health, and likely risk and protective factors were measured retrospectively and prospectively in the baseline and follow-up surveys. These are described previously in detail (Tarren-Sweeney, 2008b). Factors pertaining to children's mental health, development, education and present status (e.g. type and makeup of present placement, recent life events) were measured from carer questionnaires at both time points. Historical data (e.g. maltreatment history, care history, and birth family factors) were measured retrospectively from the state child welfare database, again at both time points. Since historical events were mostly recorded on the child welfare database shortly after they occurred, these data are thought to be more reliable than those typically obtained in a retrospective design. Children and young people were not directly involved in the study.

Recruitment of follow-up participants

The CICS follow-up survey attempted to obtain follow-up social care and developmental data for all participants in the CICS baseline survey who remained in foster or kinship care in early to mid-2009, and who had verifiable contact details. The research team was unwilling to mail survey materials containing personal information to unverified locations. Of the 347 baseline survey participants, 231 remained in court-ordered foster or kinship care at follow-up. Of these, 66 were residing in placements that did not have verifiable contact details, and whose carers could not be located by telephone. There were thus 165 young people that were eligible for inclusion in the follow-up survey, and could be reliably located. Of these, questionnaires were returned for 85 young people, representing a 51.5% response rate. However, these participants represent only 37% of baseline participants remaining in care.

Mental health measures

The mental health and socialization of subject children were measured prospectively using two sets of caregiver-report checklists; the Child Behavior Checklist (CBCL) (Achenbach, 1991; Achenbach & Rescorla, 2001), and the Assessment Checklists for Children (ACC) and Adolescents (ACA) (Tarren-Sweeney, 2007, 2013b).

CBCL. The CBCL measures child problem behavior across eight empirically-derived clinical sub-scales; two higher-order, broadband scales approximating spectrums of depressive / anxious symptoms (internalizing) and disruptive behavioral symptoms (externalizing); and a total problems score that provides a measure of global psychopathology. In addition to measuring child problem behavior in terms of continuously distributed scores, the CBCL defines scores as being within ‘normal’, ‘borderline clinical’ and ‘clinical ranges. The CBCL was selected for the present study: 1. because it yields valid and reliable mental health estimates (Halle & Darling-Churchill, 2016); 2. because of the availability of a large amount

of comparative data for high-risk populations (Armsden et al., 2000; Heflinger et al., 2000); and 3. because there are Australian normative data (Hensley, 1988; Sawyer et al., 2000).

While the social, attention, and thought problems factors did not constitute a separate higher-order factor in the development of the CBCL, they diverge from the internalizing and externalizing factors (Bor et al., 1997), and have particular relevance for the present study population. First, baseline survey mean scores of each of these three syndromes were high, and scores were strongly inter-correlated. Second, a higher-order factor analysis of the present study baseline CBCL scores yielded a four-factor higher-order model for boys (86% of score variance), that includes a Social-Attention-Thought problems (SAT) factor; as well as two-factor higher-order model for girls (75% of score variance), that includes a combined externalizing—SAT factor (2013a). Hence, a composite SAT (Social-Attention-Thought problems) scale was included in this present analysis.

The CBCL generates both raw scores and age- (4-11 and 12-18) and gender-standardized T-scores. Raw scores more precisely estimate uniform change in mental health difficulties (than do T-scores), in that they more closely approximate an interval measurement scale, whereas T-scores represent unit changes in the population distribution of such difficulties. Raw score points provide uniform, calibrated measurement of the prevalence and frequency of children's mental health difficulties – whereas T-score points do not. However, age-standardized T-scores adjust for normative, developmental shifts in the population distribution of mental health difficulties, such as in adolescence. Given the present study spans pre-adolescent and adolescent development, we need to account for some small 'developmentally-expected' increases in difficulties as identified in age-standardized normative data. Otherwise, such increases could be interpreted as deterioration in mental health. The respective normative score distributions for ages 4-11 and 12-18 on the CBCL total problems, internalizing, externalizing, withdrawn-depressed, and rule-breaking scales are

sufficiently dissimilar to use T-scores in the present analyses. For all other CBCL scales, raw scores are used in the analyses.

ACC-ACA. The ACC is a 120-item carer-report checklist that measures behaviors, emotional states, traits, and social relatedness difficulties, as experienced by children in care aged 4 to 11 years, and related populations. It was designed for the present study to measure a range of mental health difficulties (mainly attachment- and trauma-related) observed among children in care that are not adequately measured by standard rating instruments, such as the CBCL, the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 2001) and the Conners scales (Conners, Sitarenios, Parker, & Epstein, 1998). The ACC has 10 clinical sub-scales derived empirically through factor analysis, as well as two low self-esteem sub-scales derived using a different procedure. The ACA is a 105-item adolescent (ages 12-17) variant of the ACC that was initially designed for use in the CICS. While its content is mostly derived from the Assessment Checklist for Children, its factor structure differs from that of the ACC. Initial data indicate that the ACC and ACA have good content, construct and criterion-related validity, as well as high internal reliability (Tarren-Sweeney, 2007, 2013b).

As with the CBCL, each ACC and ACA scale has two cut-points that identify children who have ‘possible’ and ‘probable’ clinical-level difficulties. Scores above the higher cut-points constitute *clinical* ranges that are strongly predictive of psychiatric impairment (highly specific), while scores above the lower cut-point ranging up to and including the high cut-point constitute *elevated* ranges that retain high sensitivity for detecting psychiatric impairment. The clinical and elevated (i.e. borderline clinical) ranges for the ACC and ACA clinical scales were defined from criterion-referenced procedures and clinical reasoning. The ACC and ACA share 64 clinical items. For the present analysis, baseline and follow-up scores on these shared clinical items were summed to yield a nominal ACC-ACA *shared-item* score,

allowing for prospective comparison of children's global mental health from childhood through adolescence.

Survey procedure

The legal and care status of baseline participants at follow-up was obtained from the state child welfare database. While these details were reliably extracted from the database, foster and kinship family contact details were less reliable. The research team cross-checked the carers' contact details with the electoral roll and white pages telephone directory. Where addresses remained unverified, the research team attempted to confirm the address by telephoning foster carers. Consent forms for young people and their carers together with the questionnaires were mailed in March 2009 to verified addresses. Carers were assured that the state foster care agency had no knowledge of who was being contacted, and which individuals had decided to participate or not. Foster and kinship carers had the option of either completing the questionnaire, returning it unanswered (thus indicating non-participation), or doing nothing. Late responders were contacted, initially via a reminder letter, and later by telephone, to determine their participation status. Questionnaires were returned between March and November 2009. The study was approved by the Human Ethics Committees of the University of Newcastle, Australia (baseline and follow-up surveys), and the University of Canterbury, New Zealand (follow-up survey) [institutions masked for peer review]. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committees and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Defining meaningful prospective changes in mental health scores

While prospective changes in group mean scores provide some sense of how children's mental health changes over time during their time in care, they don't provide a sense of what proportions of children's mental health *meaningfully* improves, worsens, or

stays much the same. This latter information is fundamental for investigating the therapeutic potential of foster care. In practice, however, prospective changes in mental health estimated by way of repeated measures are continuously distributed, such that categorizing score changes as either ‘meaningful improvement’, ‘no meaningful change’, or ‘meaningful deterioration’ will always be imperfect (and somewhat artificial). A second challenge is distinguishing between statistical significance, clinical significance, and a level of change that is perceptible to children and/or families, and/or that has developmental and social meaning.

There is no accepted convention for defining such categories in population research. For this reason, the method used to define meaningful change in the present paper requires some justification. In psychotherapy research, clinically significant change is typically defined *statistically* rather than by way of clinical reasoning, using the Reliable Change Index (RCI). The RCI describes the magnitude of change that is statistically reliable – thus identifying score changes that are larger than what can be statistically attributable to internal measurement error (Jacobson & Truax, 1991). Jacobson and Traux (1991) originally proposed that the RCI be used as part of a ‘twofold’ procedure for establishing a level of change that is *clinically significant*, without elaborating on what other criterion might be used to define clinically significant change. In practice, however, the RCI has since been employed in treatment evaluation as the *sole* criterion for defining clinically significant change. This approach is limited in at least two ways. First, the RCI does not account for regression to the mean effects, or other external sources of error, such as systematic responder bias on pre- and post-measures. More important for the present study, there is no logical connection between the scale of score differences that is accounted for by measurement error, and that which represents clinically and developmentally meaningful change.

The measurement of meaningful change in clinical research proceeds with the premise that all of the study participants have clinical-level difficulties requiring intervention. The

goal of intervention is to maximize the rate of clinical improvement, and to minimize rates of both clinical deterioration and lack of change. However, population cohort studies include participants who have normative (i.e. good) mental health at baseline, many of whom will continue to manifest good mental health at follow-up. This is true even for higher risk populations, such as children in care. The lower a child's symptom scores are at baseline, the less scope they have to experience improvement in their mental health – there is a 'floor' effect. For these children, a lack of meaningful change translates as sustained mental health.

Another approach for defining clinically meaningful change is where children's scores shift from normative to clinical score ranges, and vice versa. Both of the measures used in the present study have three such ranges (normal range, borderline / elevated range, and clinical range). Population cohort studies sometimes report rates at which scores shift from one range to the next over time. While this may be useful for indicating changes in the rates of children who are in need of clinical interventions, the method yields inaccurate estimates of the rates of meaningful change across a population. This is because small, imperceptible changes in symptomatology (as little as a single raw score point) can push scores across a single clinical threshold.

With this in mind, the present analyses firstly differentiated between those children whose scores were in the normal ranges on both occasions, and other participants. The former constitute a *sustained mental health* group. For the latter, meaningful change on the two sets of measures was defined from the spread of scores traversing *both* the borderline and clinical range cut-points. The reasoning for this was that a shift in scores from a normal range to a clinical range, and vice-versa, which traverses the borderline/elevated range, is clinically meaningful. This magnitude of change is also unlikely to be exclusively accounted for by a regression to the mean effect. The relevant spread of raw scores traversing both the borderline and clinical range cut-points of the boy and girl profiles for the 4-11 and 12-18 age groups

ranges from 10 to 13 points. A 12-point score differential was selected as the universal cut-off for defining meaningful change on the CBCL total problems scale (this compares with an estimated RCI of 8 points for this scale). The method thus yields more conservative estimates of meaningful change than does the RCI. The equivalent score differentials for the ACC and ACA total problems scales are 7 and 8 points respectively. Given that these scales are constructed from 102 and 87 clinical items respectively, a conservative equivalent score differential for identifying meaningful change on the 64-item ACC-ACA shared-item scale is 5 points.

In the present study, prospective changes in 1. CBCL total problems scale scores, and 2. ACC-ACA scores were each (i.e. separately) allocated to four mutually exclusive mental health stability groups: A. *sustained mental health*; B. *meaningful improvement*; C. *no meaningful change*; D. *meaningful deterioration*.

Results

Follow-up participants had a mean age of 15.6 years, while gender was unevenly distributed (61% boys and 39% girls). Ninety-one percent of young people resided with foster parents, and 9% with kin. Three quarters of these young people remained with their baseline foster and kinship carers. Their mean age at entry into care was 2.8 years and their mean time in care at follow-up was 12.9 years. They thus constitute a group of young people who have largely grown up in long-term, stable care.

The time elapsed between baseline and follow-up measures ranged from 75 to 118 months (6.3 to 9.8 years). However, because baseline participants were largely recruited in two blocks (one in early to mid-2000, and the other in 2002), the resulting longitudinal timeframes separate into two discrete groups: 6.25 to 7.4 years (n=47, 55%), referred to as the 7-year follow-up group; and 9.0 to 9.8 years (n=38, 45%), referred to as the 9-year follow-up

group. Consequently, the follow-up sample had an uneven age distribution, with bi-modal peaks at ages 14 and 17 (*age 11*, n=1; *age 12*, n=0; *age 13*, n=15; *age 14*, n=23; *age 15*, n=8; *age 16*, n=11; *age 17*, n=24; *age 18*, n=3).

Follow-up sample representativeness

Representativeness of baseline participants remaining in care at follow-up.

The representativeness of the follow-up sample (N=85) was examined by comparing their demographic, mental health, developmental and risk exposure characteristics at baseline with that of other baseline participants who remained in care at follow-up (and were thus eligible to participate, N=146). Follow-up participants and ‘eligible non-participants’ had comparable mean age at baseline (7.7 versus 7.5 years). The former group had slightly lower representation of girls than the latter group, though the difference fell short of statistical significance (Girls = 39% and 51% respectively, $\chi^2 = 3.0$, $p = 0.08$). At baseline, follow-up participants had slightly lower, non-significant differences in global mental health scores than eligible non-participants. Mean CBCL total problems scores for the two groups were 46.1 and 48.7 respectively ($p = 0.55$), and mean ACC total clinical scores were 25.8 and 29.8 respectively ($p = 0.21$). Consistent with the direction of these non-significant differences, the follow-up sample’s mean age at entry into care was six months younger than that of eligible non-participants (2.8 and 3.3 years respectively, $p = 0.09$), and the mean length of the placements they were residing in at baseline was one year longer (4.1 and 3.2 years respectively, $p = 0.004$). Otherwise, there were no significant differences between the follow-up participants and eligible non-participants on rates of: reported intellectual disability (20% versus 27%, $\chi^2 = 1.6$, $p = 0.21$); placement in kinship care (9% versus 16%, $\chi^2 = 1.6$, $p = 0.17$); or prior notifications of sexual abuse (15% versus 11%, $\chi^2 = 0.92$, $p = 0.34$), physical abuse (55% versus 55%, $\chi^2 = 0.005$, $p = 0.94$), or neglect (74% versus 82%, $\chi^2 = 2.1$, $p = 0.15$).

Representativeness in relation to the adolescent survey mental health scores. The representativeness of the follow-up sample was also examined in terms of the extent to which participants resemble other adolescents in long-term care. The present study had opportunity to compare characteristics of the follow-up participants with that of additional participants in the CICS adolescent survey (N=145). Follow-up participants had comparable age distribution to that of other participants in the adolescent survey (mean age of both groups = 15 years) and comparable mental health scores. Mean CBCL total problems scores for follow-up versus other adolescent survey participants were 40.9 and 42.5 respectively ($p = 0.71$), while mean ACA total clinical scores were 20.5 and 22.2 respectively ($p = 0.56$). However, this discrepancy is smaller than what we might have expected, since other adolescent participants entered care at a much older mean age (6.8 years) than did the follow-up participants (2.8 years), and older age at entry into care is a known risk factor for mental health difficulties.

Mental health stability and change

Mean baseline, follow-up and prospective changes in children's mental health scores are listed in Table 1. The only significant changes were small reductions in the SAT and social problems raw scores. Seven- and 9-year follow-up group mean baseline, follow-up and prospective change scores are compared in Table 2, along with group mean age, and mean age at entry into care. Age at follow-up was the only significant between-group difference, with the nine-year follow-up group being on average 18 months older than the 7-year follow-up group.

[insert Tables 1 and 2 about here]

Rates of assignment to the prospective mental health change groups (sustained mental health; meaningful improvement; no meaningful change; meaningful deterioration), as derived from score changes on the CBCL total problems and ACC-ACA shared-item scales,

are listed for the aggregate sample, and for the 7-year and 9-year follow-up groups in Table 3. This tells quite a different story to that suggested by mean score changes. On each of the two measures, more than 60% of the children manifested either sustained mental health or meaningful improvement in their mental health, while less than a quarter showed meaningful deterioration. With regards to the two follow-up groups, they did not differ in terms of rates of assignment to the four change groups based on changes to CBCL total problems scores (4×2 $\text{Chi}^2 = 0.79$, $p = 0.85$) or ACC-ACA shared-items scores (4×2 $\text{Chi}^2 = 5.1$, $p = 0.17$). However, the 9-year follow-up group had a significantly higher rate of meaningful deterioration in ACC-ACA scores relative to all other outcomes (29.0% versus 71%), than did the 7-year follow-up group (10.6% versus 89.4%) (2×2 $\text{Chi}^2 = 4.6$, $p = 0.03$).

[insert Table 3 about here]

While the CBCL total problems score and ACC-ACA shared-item score separately provide an estimation of global mental health for this population, and are strongly correlated (baseline, $r = 0.86$; follow-up, $r = 0.89$), their items nonetheless measure different types and classes of symptoms. Furthermore, these types and classes of symptoms are likely to have different natural histories and change trajectories. Therefore it is useful to examine the concordance of assignment to change groups by each measure. Table 4 lists a 4×4 matrix showing numbers of young people assigned to each combination of the four groups for the two measures. This table reveals a fairly high degree of concordance on change group assignment for the two measures. Importantly, no young person showed meaningful improvement on one measure, and meaningful deterioration on the other. In summary, the findings indicate that: 31.8% of young people ($n=27$) manifested *sustained mental health* on both measures; 32.9% ($n=28$) manifested *meaningful improvement* in their mental health on

one or both measures; 29.4% (n=25) manifested *meaningful deterioration* in their mental health on one or both measures; and the remaining 5.9% (n=5) showed no meaningful change on either measure.

[insert Table 4 about here]

Discussion

Mean score changes in the aggregate sample suggest that the distribution of mental health difficulties among children growing up in long-term, family-based care is unchanged over a seven- to nine-year period. The only exceptions were small reductions in mean CBCL SAT and social problems scores. These findings respond to the first research question in the negative, suggesting there is no homogenous, population-wide shift in children's mental health as they grow up in care. Instead, rates of meaningful change identified in the present study (using a more conservative definition than that typically employed in treatment evaluation studies) respond to the second research question in the affirmative – suggesting that substantial proportions of children in care experience very different mental health trajectories. This is not surprising, given that the experience of growing up in care is heterogeneous, and that children enter care with heterogeneous developmental context (such as varying ages at entry into care). It is very likely that for some children, long-term care provides the opportunity to grow up feeling loved, nurtured and secure, translating into relational permanence. For these children, alternate care provides the conditions for developmental recovery, whereby therapeutic effects accumulate and consolidate over time. For other children, long-term care may compromise their development and well-being. It was notable in this regard that the nine-year follow-up group had close to three times the rate of meaningful deterioration in ACC-ACA scores than the seven-year follow-up group (29.0% versus 10.6%).

This raises a question then, about how we should define and measure ‘successful’ outcomes for children whose early development is seriously compromised by severe and chronic maltreatment. Children who continue to manifest clinical-level mental health difficulties through their time in care may nonetheless have benefitted from being in care, relative to what would have unfolded had they remained in the care of their maltreating families. In other words, success might be defined as experiencing less adverse developmental outcomes than otherwise would have been the case. Unfortunately, we presently lack good comparative data on the developmental trajectories of severely and chronically maltreated children who are raised by their parents versus those raised by foster carers and kin (Goemans, van Geel, van Beem, & Vedder, 2016).

The study findings are potentially compromised by two limitations – high sample attrition, and collection of data at only two points in time. Follow-up participants constituted only 37% of the baseline participants remaining in long-term care. However, this limitation is mitigated by data showing: 1. that follow-up participants resemble other baseline participants who remained in care on outcome and study factors measured at baseline; and 2. that follow-up participants are not dissimilar to other (newly recruited) participants in the adolescent survey. These two comparisons suggest that, despite high attrition, the follow-up sample is representative of baseline participants who remained in care, and that the study findings are thus generalizable for children growing up in long-term care.

Rather than asking whether long-term care is generally therapeutic or harmful for the development of previously maltreated children, future investigations should focus on the questions “...*what are the systemic and interpersonal characteristics of care that promote and sustain children’s psychological development throughout childhood, and what characteristics are developmentally harmful?*” and “...*for which children is care therapeutic, and for which children is it not?*” These questions are critical for both social care policy and

practice, and for the design of clinical interventions. Various developmental theories (including attachment theory and social learning theory), as well as research into the neurodevelopmental effects of early maltreatment, would predict that the therapeutic and harm potentials of long-term care will be moderated by such factors as children's age when entering care, their carers' commitment and bonding to them, and the stability of their placements. There are likely to be complex transactional mechanisms that dictate children's developmental trajectories as they grow up in care. Unfortunately, there has been very little substantive research on this question to date. It is also important to keep in mind that developmental change within care is moderated by this population's earlier exposure to severe social adversity. A recent study found that the developmental effects of more than six months exposure in early childhood to profound psychosocial deprivation (institutional care) persist for many through childhood and adolescence – despite being subsequently raised by loving adoptive families (Sonuga-Barke et al., 2017). These persistent effects include symptoms of autism spectrum disorder, disinhibited social engagement, and inattention and over-activity. This supports the notion that recovery from some forms of psychopathology caused by early severe adversity tends to follow a long developmental trajectory even where a child's developmental conditions have markedly improved.

Acknowledgement

The present study was funded and facilitated by the New South Wales Department of Family and Community Services. However, the information or views contained in this study do not necessarily, or at all, reflect the views or information held by the New South Wales government, the Minister for Family and Community Services, or the Department.

Table 1. Mean prospective change in mental health scores (aggregate and gender-stratified)

Scale	Baseline score Mean (SD)	Follow-up score Mean (SD)	Change	Cohen's <i>d</i>
<i>CBCL scales (T-scores)</i>				
Total problems	59.4 (12.5)	56.9 (12.9)	-2.5	.20
Externalizing	56.8 (12.1)	57.3 (12.3)	0.5	.04
Internalizing	52.7 (11.3)	52.7 (11.5)	0.1	< .01
Withdrawn/Depressed	56.3 (8.4)	57.0 (7.9)	0.7	.09
Rule-breaking behavior	59.2 (9.0)	59.2 (9.1)	0.1	< .01
<i>CBCL scales (raw scores)</i>				
SAT ^a	17.2 (12.0)	14.4 (12.1)	-2.8	.23 *
Anxious/Depressed	3.8 (4.0)	3.4 (3.4)	-0.5	.14
Somatic complaints	1.9 (2.4)	1.9 (2.4)	-0.1	.04
Social problems	5.5 (4.6)	4.0 (4.5)	-1.5	.33 **
Thought problems	4.4 (4.3)	3.5 (3.9)	-1.0	.24
Attention problems	7.4 (5.0)	7.0 (5.2)	-0.4	.08
Aggressive behavior	10.4 (8.0)	8.7 (7.3)	-1.6	.21
<i>ACC/ACA scales (raw scores)</i>				
ACC-ACA shared-item ^b	17.0 (15.2)	14.4 (13.9)	-2.6	.18
Negative self-image	2.0 (3.6)	1.9 (2.9)	-0.1	.03
Low confidence	3.8 (3.3)	4.0 (3.5)	0.2	.06

^a SAT = Composite of Social problems, Attention problems and Thought problems syndrome scale raw scores; ^b Total raw score of 64 clinical items shared by the ACC (child) and ACA (adolescent) measures.

* $p < .05$ ** $p < .001$

Table 2. Comparison of 7-year and 9-year follow-up groups' mental health scores, age at follow-up and age at entry into care

		7-year follow-up group (n=47)		9-year follow-up group (n=38)		<i>p</i> value (2-tailed)
		Mean	(SD)	Mean	(SD)	
	Age at follow-up	14.9 yrs	(1.5)	16.6yrs	(1.4)	<0.001
	Age at entry into care	2.8 yrs	(2.2)	2.7yrs	(2.2)	0.90
CBCL total problems T-score:	Baseline	60.6	(12.4)	57.9	(12.7)	0.34
	Follow-up	56.8	(13.0)	57.0	(12.9)	0.95
	Change	-3.8	(12.0)	-1.0	(16.0)	0.36
ACC-ACA shared item score:	Baseline	18.1	(14.8)	15.7	(15.7)	0.46
	Follow-up	13.1	(13.1)	16.0	(14.8)	0.34
	Change	-5.0	(13.0)	0.3	(16.5)	0.10

Table 3. Rates of assignment to mental health change groups: 1. *sustained mental health*; *meaningful improvement*; 3. *no meaningful change*; and 4. *meaningful deterioration*

<i>Measure</i>	<i>Group</i>	Group A Sustained mental health ^a		Group B Meaningful improvement ^b		Group C No meaningful change ^c		Group D Meaningful deterioration ^d	
		N	%	N	%	N	%	N	%
<i>CBCL total</i>									
<i>problems scale</i>	Aggregate (n=85)	30	35.3%	23	27.1%	11	12.9%	21	24.7%
	7-year follow-up group (n=47)	18	38.3%	13	27.7%	6	12.8%	10	21.3%
	9-year follow-up group (n=38)	12	31.6%	10	26.3%	5	13.2%	11	29.0%
<i>ACC/ACA shared</i>									
<i>clinical items</i>	Aggregate (N=85)	32	37.7%	21	24.7%	16	18.8%	16	18.8%
	7-year follow-up group (n=47)	19	40.3%	14	29.8%	9	19.2%	5	10.6%
	9-year follow-up group (n=38)	13	34.2%	7	18.4%	7	18.4%	11	29.0%

^a Scores within normal range at baseline and follow-up; ^b CBCL total score reduction > 11, ACC-ACA shared-item score reduction > 4; ^c CBCL total score change < 12, ACC-ACA shared-item score change < 5; ^d CBCL total score increase > 11, ACC-ACA shared-item score increase > 4.

Table 4. Matrix of participants' (N=85) assignment to mental health change groups, based on CBCL versus ACC-ACA score changes

Group Assignment based on ACC-ACA score changes

		A: Sustained mental health	B: Meaningful improvement	C: No meaningful change	D: Meaningful deterioration
<i>Group Assignment based on CBCL score changes</i>	A: Sustained mental health	27	3	0	0
	B: Meaningful improvement	3	16	4	0
	C: No meaningful change	0	2	5	4
	D: Meaningful deterioration	2	0	7	12

References

- Achenbach, T. (1991). *Manual for the Child Behavior Checklist / 4-18, and 1991 profile*. Burlington, VT: University of Vermont.
- Achenbach, T., & Rescorla, L. (2001). *Manual for ASEBA school-age forms and profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Armsden, G., Pecora, P. J., Payne, V. H., & Szatkiewicz, J. P. (2000). Children placed in long-term foster care: An intake profile using the child behavior checklist/4-18. *Journal of Emotional & Behavioral Disorders, 8*(1), 49-64.
- Biehal, Nina, Ellison, Sarah, Baker, Claire, & Sinclair, Ian. (2010). *Belonging and permanence: Outcomes in long-term foster care and adoption*. London: BAAF.
- Bor, W., Najman, Jake M., Andersen, Margaret J., O'Callaghan, Michael, Williams, G. M., & Behrens, Brett C. (1997). The relationship between low family income and psychological disturbance in young children: An Australian longitudinal study. *Australian & New Zealand Journal of Psychiatry, 31*(5), 664-675.
- Bulat, L. (2010). A longitudinal study of depressiveness in children in public care. *International Journal of Social Welfare, 19*(412-423).
- Conners, C. Keith, Sitarenios, Gill, Parker, James D. A., & Epstein, Jeffrey N. (1998). The revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26*(4), 257-268.
- DeJong, Margaret. (2010). Some reflections on the use of psychiatric diagnosis in the looked after or 'in care' child population. *Clinical Child Psychology and Psychiatry, 15*(4), 589-599.
- Dubowitz, Howard, Zuravin, Susan, Starr, Raymond H, Feigelman, Susan, & Harrington, D. (1993). Behavior problems of children in kinship care. *Journal of Developmental and Behavioral Pediatrics, 14*(6), 386-393.
- Fanshel, David, & Shinn, Eugene. (1978). *Children in foster care: A longitudinal investigation*. New York: Columbia University Press.

- Fein, E., & Maluccio, A. (1992). Permanency planning: Another remedy in jeopardy? *Social Services Review*, *66*(3), 335-348.
- Frank, G. (1980). Treatment needs of children in foster care. *American Journal of Orthopsychiatry*, *50*(2), 256-263.
- Goemans, A., van Geel, M., van Beem, M., & Vedder, P. (2016). Developmental outcomes of foster children: A meta-analytic comparison with children from the general population and children at risk who remained at home. *Child Maltreatment*, *21*(3), 198-217.
- Goemans, A., van Geel, M., & Vedder, P. (2015). Over three decades of longitudinal research on the development of foster children: A meta-analysis. *Child Abuse and Neglect*, *42*, 121-134.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*, 1337-1345.
- Halle, T., & Darling-Churchill, K. (2016). Review of measures of social and emotional development. *Journal of Applied Developmental Psychology*, *45*, 8-18.
- Havnen, K., Breivik, K., & Jakobsen, R. (2014). Stability and change: A 7- to 8-year follow-up study of mental health problems in Norwegian children in long-term out-of-home care. *Child & Family Social Work*, *19*, 292-303.
- Heflinger, Craig Anne, Simpkins, Celeste G., & Combs-Orme, Terri. (2000). Using the CBCL to determine the clinical status of children in state custody. *Children & Youth Services Review*, *22*(1), 55-73.
- Hensley, V. Rae. (1988). Australian normative study of the Achenbach Child Behavior Checklist. *Australian Psychologist*, *23*(3), 371-382.
- Hukkanen, Raija, Sourander, Andre, Bergroth, Lasse, & Piha, Jorma. (1999). Psychosocial factors and adequacy of services for children in children's homes. *European Child and Adolescent Psychiatry*, *8*, 268-275.
- Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12-19.

- Oswald, S, Heil, K, & Goldbeck, L. (2010). History of maltreatment and mental health problems in foster children: A review of the literature. *Journal of Pediatric Psychology, 35*(5), 462-472.
- Pecora, P, White, C, Jackson, L, & Wiggins, T. (2009). Mental health of current and former recipients of foster care: a review of recent studies in the USA. *Child & Family Social Work, 14*, 132–146.
- Rowe, J, & Lambert, L (1973). *Children who wait*. London: Association of British Fostering Agencies.
- Sawyer, Michael G., Kosky, Robert J., Graetz, Brian W., Arney, Fiona, Zubrick, Stephen R., & Baghurst, Peter. (2000). The National Survey of Mental Health and Wellbeing: The child and adolescent component. *Australian & New Zealand Journal of Psychiatry, 34*(2), 214-220.
- Sonuga-Barke, E, Kennedy, M, Kumsta, R, Knights, N, Golm, D, Rutter, M, . . . Kreppner, J (2017). Child-to-adult neurodevelopmental and mental health trajectories after early life deprivation: the young adult follow-up of the longitudinal English and Romanian Adoptees Study. *The Lancet*. doi: 10.1016/S0140-6736(17)30045-4
- Tarren-Sweeney, M. (2013a). An investigation of complex attachment- and trauma-related symptomatology among children in foster and kinship care. *Child Psychiatry & Human Development, 44*, 727-741.
- Tarren-Sweeney, M. (2013b). An investigation of complex attachment- and trauma-related symptomatology among children in foster and kinship care. *Child Psychiatry and Human Development. 44*, 727–741.
- Tarren-Sweeney, M. (2008a). The mental health of children in out-of-home care. *Current Opinion in Psychiatry, 21*, 345-349.
- Tarren-Sweeney, M. (2008b). Retrospective and concurrent predictors of the mental health of children in care. *Children and Youth Services Review, 30*, 1-25.
- Tarren-Sweeney, M. (2007). The Assessment Checklist for Children – ACC: A behavioural rating scale for children in foster, residential and kinship care. *Children and Youth Services Review, 29*, 672-691.

Vis, S, Handegård, B, Holtan, A, Fossum, S, & Thørnblad, R. (2016). Social functioning and mental health among children who have been living in kinship and non-kinship foster care: Results from an 8-year follow-up with a Norwegian sample. *Child & Family Social Work, 21*, 557-567.