Academic and Clinical Outcomes

In Context of a Theoretical Framework of Attachment, Positive Psychology and Emotional Intelligence

December, 2013

Larry S. Sanders, Ph.D., F.A.P.A.

Chairman/CEO Bayfield

sanders@bayfield.net

Robert J. Fulton, M.S.W.

PRO.FILES

Social Research and Outcomes Evaluation

r.fulton@rogers.com

Acknowledgements

Bayfield has been conducting evidence- based practice since 1994. For the past 20 years many dedicated staff have been an integral part of administrating hundreds of psychometric and sociometric tests in the pursuit to improve the quality of treatment and outcomes for the children we serve. We are very grateful for the careful attention our child and youth workers, teachers and clinicians have paid to implementing our theoretical framework of attachment in the context of positive psychology and development of emotional intelligence. Their continuity in treating our children has resulted in exceptional outcomes, academically and clinically.

We are also grateful to our partners from the child welfare and children's mental health sectors for providing us with individual and organizational feedback. And, most importantly the lessons our children have taught us as they allowed us to become part of their lives.

And, our appreciation to our research partner of 25 years, Robert J. Fulton, for his statistical expertise, creativity and dedication to numerous projects.

Larry S. Sanders, PhD

Chairman, CEO

Background

Our quantitative research (Sanders and Fulton, 2012) clearly indicated the 417 children studied achieved exceptional outcomes as measured by psychometric and sociometric instruments. Our qualitative research demonstrated the complexity of relationships through our analysis of the language used by staff and children. The importance of acknowledging the unique presentations of the population studied indicates our theoretical framework of attachment includes approaches based in positive psychology and emotional intelligence. We have developed a systematic approach to collecting reliable data with the Bayfield Information System (BIS), designed to acquire data to support evidence-based practices (EBP) and produce measurable outcomes.

Outcome Data Coverage for Residential Care

Since 1994 measurements of evidence based outcomes drawn from the Bayfield Information System (BIS) and are listed under the research tab at www.bayfield.net. On January 4, 2006, Bayfield started collecting outcome data on the academic performance of the children attending the Bayfield School. The outcome evaluation is based on a longitudinal design of multiple testing within the first 30 days following admission and every nine months thereafter.

The evidence for this study was drawn from Jan 1, 1994 to May 15, 2013 consisting of 712 unique children. Of these children, this research study has outcome evidence on 551 children. There were 161 children served who were not tested:

(a) 131 children were discharged before March 15, 2001, when clinical outcome test protocol was fully implemented;

- 28 children were not in care long enough to be tested;
- 7 children were reviewed for treatment and were not admitted
- 9 children remained in treatment for less than 42 days
- 12 children were admitted after March 15, 2013; they are still awaiting their 1st testing cycle.

(b) 2 children were admitted but were never tested

The evidence reviewed on 551 children, between March 15, 2001 and March 15, 2013 were tested on the outcome measures, with a compliance rate of 99.6%.

Outcome Data Coverage for Academic Program

The outcome data for the academic program started several years after the residential outcome study was launched. This study reviews outcome data on 354 unique children with the first measure taken on January 4, 2006. The outcome measure selected in 2006 was the Kaufman Test of Academic Achievement (K-Tea). On May 4, 2011, the instrument for measuring academic outcomes was changed to the Wechsler Individual Achievement Test (WIAT). The WIAT has Canadian norms and includes more comprehensive scales.

Since January 4, 2006, 831 tests have been administered to 354 children. Since the academic testing started 423 children were tested. The testing completion rate in the school is 84%. Academic testing is completed twice a year, September and May. Due to the pattern of admission and discharges, 69 children were not available for testing.

Clinical Outcomes

Clinical Functioning Level on Admission

The clinical functioning of children are assessed on admission using the following evidence based instruments:

(1) *Conners' Global Index*: measures hyperactivity, inattention and impulsiveness;
(2) *Children's Global Assessment Scale*: measures how well the child functions in his major life roles;
(3) *SA-45*: measures psychiatric symptoms;
(4) *FAB-C:* measures emotional problems, anti-social attitudes and behaviour;
(5) *Objective Stressors Checklist:* identifies how many things distress the child;
(6) *Caring Scale of the PBI:* measures how much affection the child feels he received from the "closest person" in his life, a critical element of attachment; and,
(7) *Over-Protection Scale of the PBI:* measures how much the child feels neglected or conversely treated unfairly by the "closest person" in their life. For analytical purposes, this scale is split into two sub-scales, neglect and unfairness.

How Progress is Demonstrated

The statistic used to measure the success of the children admitted to Bayfield is Cohen's d or the standardized treatment effect[1]. This statistic is used in meta-analysis of published research to compare the outcomes from other programs and instruments. Different tests use unique numbering systems to measure the outcomes. Cohen's d was created to change the different numbering systems into a standard format on a bell curve.

Cohen's d is also known as the *effect size* or the *treatment effect.* The *effect size* is the raw difference between the mean score on admission and the mean score nine months later, divided by the standard deviation of the corresponding test scores at baseline. This formula converts the raw difference in scores into a standardized value.

By using the baseline standard deviation (SD) as the denominator, the effect size is also standardized across time as well as across different instruments and different subgroups of children. This allows one to compare the difference in scores for any instrument, any type of treatment or any group of children in any publication on the same metric. A typical effect size, such as 0.75 indicates the change in the score across time periods is equal to ¾ of the standard deviation of the scores. The guidelines for interpreting the effect size are specified in footnote[2].

Hyperactivity, Inattention and Impulsivity

Certain symptoms are more predictive of serious behavioural misconduct across the life span. The strongest predictors[3] are poor peer relationships, measured by the FAB-C and hyperactivity/inattention, measured by the Conners' Global Index.

The Conners' Global Index (CGI) is a norm referenced screening test for symptoms of hyperactivity, inattention and impulsivity. These symptoms are especially important in academic accomplishment[4].

---

[1] Jacob Cohen (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates

[2] Cohen gives the following guidelines for the social sciences: small effect size, r = 0.1 − 0.23; medium, r = 0.24 − 0.36; large, r = 0.37 or larger

[3] Rutter, Michael and Seija Sandberg (1985), "Epidemiology of Child psychiatric Disorder: methodological Issues and Some Substantive Findings", *Child Psychiatry and Human Development*, 15(4), 209-233

[4] Galéra, C., Melchior, M., Chastang, J. Bouvard, M. & Fombonne, E. (2009), "Childhood and adolescent hyperactivity-inattention symptoms and academic achievement 8 years later: the GAZEL Youth study", *Psychological Medicine*. 39(11): 1895–1906

"Hyperactivity-inattention symptoms independently predicted grade retention, failure to graduate from secondary school, obtaining a lower-level diploma, and lower academic performances. These results remained significant even after accounting for school difficulties at baseline", Galéra et al, ibid, page 1895.

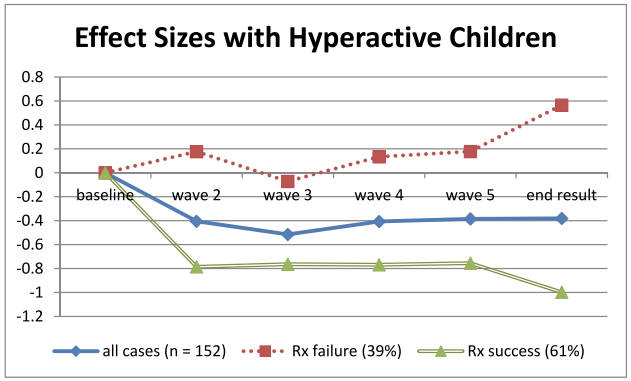
Making changes in the client level of hyperactivity is clinically important however very difficult to achieve. Hyperactivity, inattention and impulsiveness are symptoms of other problems. ADHD, the DSM diagnosis related to this set of symptoms, has the highest degree of co-morbidity with other psychiatric disorders. Children who suffer from adversity and trauma are known to experience symptoms of ADHD; these symptoms are also associated with psychological distress. When children eat too much sugar or lose too much sleep, they can acquire ADHD symptoms. ADHD is associated with learning disability, acquired brain injury and developmental conditions, such as autism. Several aspects of life can cause these symptoms; it is very difficult to treat and reduce the symptoms once they become entrenched.

Five hundred and forty eight (548) children served by Bayfield have been tested multiple times with the CGI, which is the gold standard instrument for measuring ADHD symptoms. Of these children, 40% had scores on admission to Bayfield that were in the clinical range, i.e., t-score >70. Comparatively, in any given school in Ontario, 2% of children would score above this threshold.

The group of children in this study, in the clinical range, show an average score of 81.6. Less than 1 in a 1,000 children in the public school system have a score this range.

Graph 1 is a portrait of these high risk children over time during their treatment.

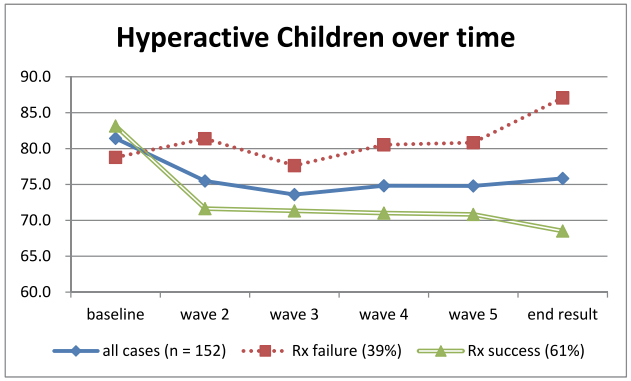
Graph 1 Effect Sizes of the Treatment with Hyperactive Children



The graph charts the changes for three groups of children: (a) all children in the clinical range; (b) a sub-group of treatment failures whose scores became consistently worse, and (c) a sub-group of treatment successes with scores consistently better.

This graph shows 61% of children improved, starting in the first nine months, referred to as wave 2 on the graph. The effect size of the change within the first few months of treatment is minus 0.786, which is considered a large treatment effect. This large improvement was maintained over time and the last test for each child shows more improvement (d = 0.998). Graph 2 shows the same data using the actual scores from the CGI.

Graph 2 Actual Results of Testing ADHD Over Time



This graph follows the actual test scores over time and indicates 61% of the children, labelled as *Rx success*, who were in the clinical range on the CGI at admission improve to a normal clinical threshold.

As indicated in the literature by Galéra et al, achieving positive outcomes on this domain is critically important to achieving academic results.

Social Functioning

Some children fail to adapt to the social demands of being home with parents and siblings, being in school with teachers and peers, or being in the community at play or with adults. The failure to fit in socially and perform the major roles of family member, student and community member is the strongest predictor in childhood of life span dysfunction[5] regardless of the specific diagnosis or even the absence of a diagnosis.

The Children's Global Assessment Scale (CGAS) is a clinician rated measure of how well the child is adapting to his major roles of life, i.e., son, daughter, student, friend and neighbour. The CGAS is embedded in Axis V of the DSM-IV. It is a better predictor of life span outcomes than the Axis I diagnosis. In other

[5] Sroufe, L. & Rutter, M (1984), "The Domain of Developmental Psychopathology", *Child Development*, 55, 17-29

words, children can have a major diagnosis, such as anxiety, depression, ASD or ADHD, which may affect them all of their lives, but if they learn to "fit in" despite their diagnosis, they will have a much better quality of life. Therefore, achieving results on this domain is a critically important outcome for residential treatment centres.

The CGAS has been used as a gating instrument in tertiary level services across the USA and found to be a reliable and valid indicator of the intensity of service needs[6]. The threshold for admission to psychiatric hospital is generally set at CGAS = 50[7] (Newman et al, 1987).

The average score for children admitted to Bayfield (CGAS = 47.9) is below the threshold for admission to psychiatric hospital. The median score for the Bayfield residents on admission is CGAS = 45. This is the dividing line for all residents; 50% of residents are below this point and 50% are above this point upon admission to Bayfield. Children with this level of functioning are described on the test as:

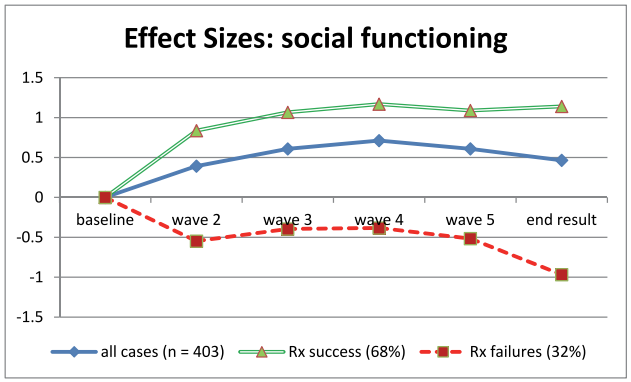
> *Major impairment in functioning in several areas or unable to function in one of these areas*, i.e., disturbed at home, at school, with peers, or in society at large, (persistent aggression without clear instigation; markedly withdrawn and isolated behaviour due to either mood or thought disturbance, suicidal attempts with clear lethal intent; such children are likely to require special schooling and/or hospitalization or withdrawal from school, however this is not a sufficient criterion for inclusion in this category).

Graph 3 shows the path children take as they receive treatment at Bayfield, measured as an effect size. All children receiving treatment and assessed on this instrument more than once are represented on this graph.

[6] Guzder, J., Paris, J., Zelkowitz, P. & Marchessault, K. (1996), "Risk Factors for Borderline Pathology in Children", *Journal of the American Academy of Adolescent Psychiatry*, 35 (1), 26-33

[7] Newman, F.L., Richard H. Hunter, R.H. & Irving, D. (1987) "Simple Measures of Progress and Outcome in the evaluation of Mental Health Services", *Evaluation and Program Planning*, 1987, vol 10, 209-218

Graph 3 Effect Sizes: Social Functioning of all Residents

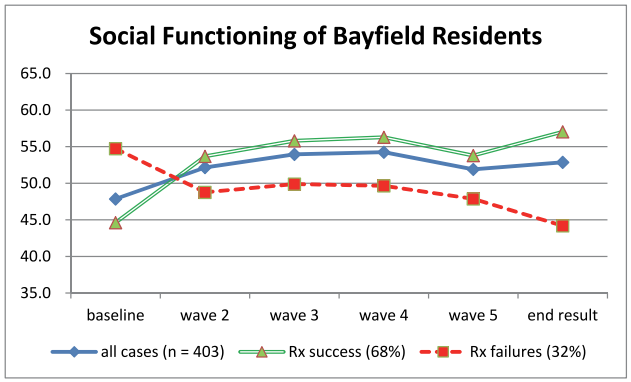


Graph 3 shows all children make a large improvement in social functioning as measured by the standardized treatment effect. This occurs within the first nine months and is maintained during the child's treatment. The outcome is unusually high in treatment programs[8]. The meta-analysis by Gresham and Horner (2001) was conducted on students with diagnosed disabilities, although not as severe as the children in this study. Meta-analysis showed a relatively weak effect size of d = 0.20 for social skills training with children having diagnosed mental disorders. The outcomes indicate effect sizes as high as 0.712 for all children at wave 4 after 27 months in treatment. The effect sizes for 62% of children who were a treatment success were very large (d = 1.169) after 27 months of treatment.

Gresham and Horner state that, "there was no difference between effect sizes based on the quality of research, length of SST, the method of measuring social skills and the construct used to measure social skills" (ibid, page 336). The children in this study are achieving more than twice the positive outcome compared to other studies.

[8] Gresham, Elliott & Horner, Robert (2001), "Interpreting Outcomes of Social Skills Training for Students with High-Incidence Disabilities", *Exceptional Children*, 67 (3), 331-344

Graph 4 Actual Scores Measuring Social Functioning



The children who deteriorated in social functioning were admitted with a significantly higher average level of functioning (CGAS = 55) compared with the children that improved consistently and strongly over time (CGAS = 44.5). Secondly, both groups Rx success and Rx failures actually end up in the reverse position. This unusual pattern suggests children functioning at a higher level on admission, need a special treatment protocol that starts immediately upon admission to prevent social functioning deterioration.

Emotional Disturbance

The measure used to assess emotional disturbance is the Feelings, Attitudes and Behaviour Scale for Children (FAB-C). The FAB-C is appropriate for school age children to measure conduct problems, poor self image, worries, negative peer relations, antisocial attitudes and lie-defensiveness. In addition, the FAB-C provides a total problem score. Negative peer relations, one of the FAB-C scales, is among the most powerful predictors of serious behavioural misconduct across the life span (Rutter and Sandberg, 1985, ibid). The FAB-C is a measure of the global condition described as "emotionally disturbed". It is not mapped to psychiatric disorders, however children who are in the clinical range are at risk of developing a psychiatric disorder.

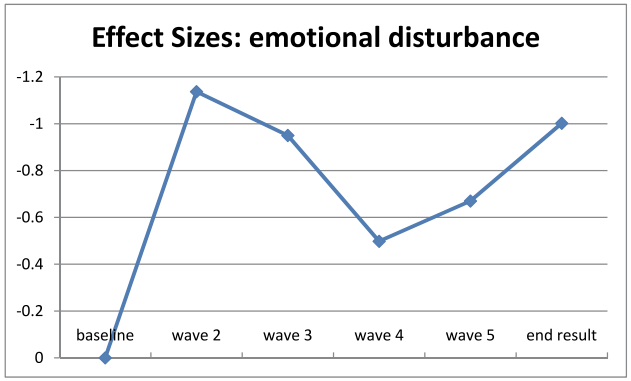
The relation between the FAB-C and various Psychiatric Diagnostic measures and the Childhood Depression Inventory (CDI) were clearly demonstrated. For example, 78% of children with clinically elevated scores on the

CDI were identified by children with a T-score of 65 or greater in the problem index[9].

Individual effect sizes for each child were computed based on the standard deviation at baseline for all children. The average score (t = 58.8) at baseline and across all waves for all children was in the normal range. The average effect sizes across five waves of testing was zero or non-significant. This is inevitable because children who are in the normal range on the FAB-C cannot get "more normal".

Due to the ceiling effect, the effect sizes on the FAB-C are reported below only for children who are in the clinical range on admission. Twenty-three percent (23%) of children below the age of 13 years, were in the clinical range; this high-risk group of children had an average t-score of 72.3.

Graph 5 Effect Sizes for Emotionally Disturbed School Age



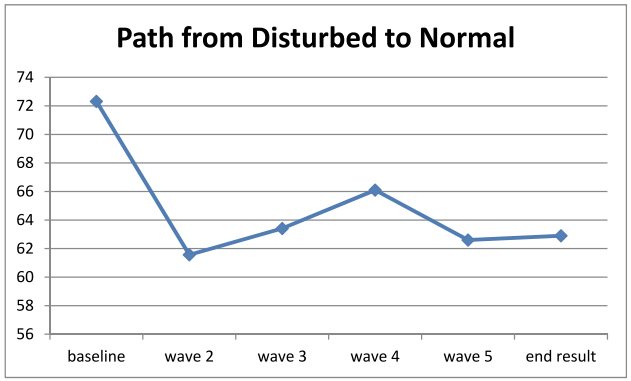
School age children admitted with clinically elevated scores on the FAB-C made significant improvements after nine months of treatment (d = minus 1.137). The improvement was maintained with a very large effect size on discharge (d = minus 1.002). Based on the domains assessed by the FAB-C, emotionally disturbed school aged children made significant progress:

[9] Beitchman, J. H. (1996), *Feelings, Attitudes and Behaviors Scale for Children (FAB-C)*, Technical Manual, Toronto: MultiHealth

- fewer conduct problems
- better self image
- fewer worries
- better peer relations, and
- fewer antisocial attitudes.

Graph 6 indicates the scores obtained by the subgroup assessed as high risk on this measure, at baseline.

Graph 6 Path to Recovery for Disturbed School Aged



School age children admitted with clinically elevated scores on domains of emotional disturbance return to the normal range of emotions, peer relations and conduct within nine months of admission and remain in the normal range.

Symptoms of Psychiatric Disorder

Symptoms of psychiatric disorder were assessed using the SA-45. This is a one-page form with 45 questions completed by the child. The longer version of this test, SCL-90R by Derogatis, is the major epidemiological instrument for adult

mental health. Borduin[10] et al (1995) found that the Global Severity Index of the SA-45 "represents the best single indicator of the respondent's psychiatric functioning" (page 572) The questions on the SA-45 are mapped to the categories of the DSM IV, the classification system for psychiatric disorders.

In the case of the SA-45, the standard deviation was computed for the first wave in each pair of waves as illustrated in the table below. The table below shows information about children in this study that had an SA-45 score above 65, our threshold for the clinical range.

Listed below is a summary of the outcomes at each wave of testing:

a) Baseline condition:

- 314 adolescents assessed with the SA-45
- Standard Deviation (SD) of baseline sample = 9.118[11]
- 72 cases (23%) found to be in the clinical range (t score >65)
- Mean score of baseline clinical cohort: t = 71.1

b) Wave 2 compared to Baseline:

- 48 pairs of cases from the baseline clinical cohort were compared
- Effect size: d = -0.770 a very large treatment effect
- Mean score at Wave 2: t = 63.2 within the normal range
- T-score of difference in means = -6.279, sig = .000

c) Wave 3 compared to Baseline:

- 34 pairs of cases from the baseline clinical cohort
- Effect size: d = -0.987 a very large treatment effect
- Mean: t = 60.7
- T-score of difference in means = -8.766, sig = .000

[10] Borduin, C., Mann, B., Cone, L., Hengeller, S., Fucci. B., Blaske, D., & Williams, R. (1995), "Multisystemic Treatment of Serious Juvenile Offenders: Long-Term Prevention of Criminality and Violence", *Journal of Clinical and Consulting Psychology*, 63, 4, 569-578

[11] Denominator for computing the effect size

d) Wave 4 compared to Baseline:

- 20 pairs
- Effect size: d = -1.409 a very large treatment effect
- Mean: t = 57.1
- T-diff = - 5.986, sig = .000

e) Wave 5 compared to Baseline

- 11 pairs
- Effect Size: d = -1.426 a very large treatment effect
- Mean: t = 57.7
- T-diff = - 5.900, sig = .000

f) Last test[12] before Discharge

- 51 pairs
- Effect size: d = -1.140 a very large treatment effect
- Mean: t = 59.7
- T-diff – 8.061, sig = .000

Considering the original group of 72 cases who were in the clinical range at baseline, 6 were discharged in the clinical range. 92% of adolescents tested in the clinical range on admission were discharged in the normal range.

Eighteen or 7% of the children whose tests results were in the normal range at baseline experienced a psychiatric deterioration, and therefore fell into the clinical range. A total of 112 children, or 36% of the cases tested with the SA-45 were in the clinical range.

Twenty percent of all children in the clinical range at some point in their treatment were discharged without a positive treatment effect. Graph 7 shows the path to recovery for three groups of children who were tested on the SA-45 across time:

[12] The last test before discharge may range from Wave 2 to Wave 14

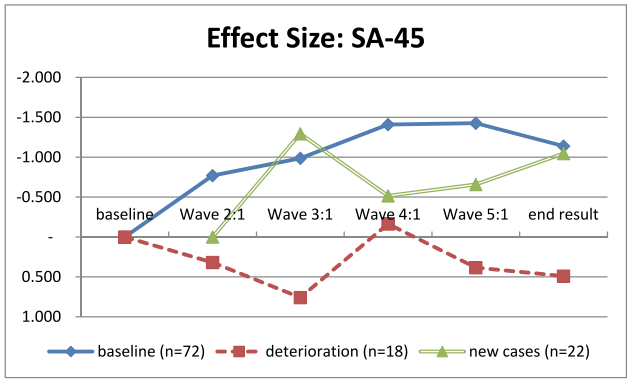
(a) Baseline: tested on admission in the clinical range;

(b) New cases: testing in subsequent waves after admission were in the clinical range;

(c) Deterioration: tested on admission in the normal range, and subsequently deteriorated or fell into the clinical range.

The cases labelled "Deterioration" are not considered treatment failures. Most psychiatric illnesses cannot be diagnosed until late adolescence or early adulthood. Before symptoms appear, the child appears normal in relation to psychiatric disease.

Graph 7 The Path to Recovery for Psychiatric Illness



Graph 7 shows the baseline group of children in the clinical range recover quickly with a very large treatment effect with nine months of treatment. This group returns to the normal range at wave 2 and continue developing within this range.

Level of Care

The Level of Care is a measure of adult support required in 22 daily living skills of the child. As children get older, they become more independent; although, some daily living skills take longer to acquire full independence. Daily living skills that require daily supervision until adolescence include:

- Manages own behaviour without instruction from others
- Follows rules, guidelines and routines of activities
- Cleaning up after an activity
- Going from place to place in the community
- Using community recreation facilities

The full set of daily living skills include those pertaining to *basic self care* (eating, grooming, bathing, toileting), *self-regulation* (following rules and routines, coping with negatives) *communication* (providing positive and negative feedback, making choices), *social interaction* (joining in a conversation, offering and accepting assistance) and *participating in the community* (going to and from places, attending community recreation, going to school).

Each skill is rated on a five point scale from *independent,* worth 0 points, to *total care*, worth 4 points.

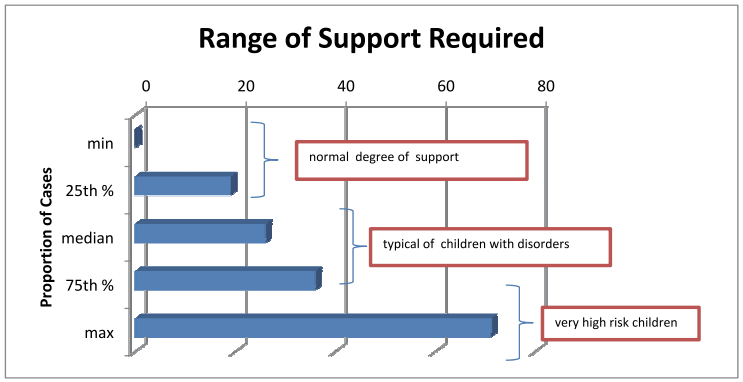
A medically fragile, quadriplegic child with no speech and severe intellectual deficit would probably score the maximum or 88 points showing that he/she requires 100% adult support to cope with 22 daily living skills. It is normal for most children and youth to need *daily supervision* for 4 or 5 of these skills, resulting in a score of 5%. Children rarely become absolutely independent before their late teens. For this reason, a score of 15% or lower means the child or youth imposes a relatively normal burden on the caregiver. At this level, the child has sufficient capacity in daily living skills to function, in this respect only, in his own home or in regular foster care.

Scores between 15% and 30% indicate the child has a number of serious deficiencies in daily living skills that require adult support. These children benefit from treatment foster care when the challenge becomes too great for the parents of origin. When the scores are above 30%, the amount of time required of the caregiver becomes significant.

Baseline Scores and Variation

The mean baseline score for the residents of Bayfield is 28.2 or 32% of the maximum amount of support possible. The effect sizes for all children tested on the LOC was computed based on the standard deviation of the baseline scores (SD = 12.666). This is a large degree of variance in relation to the average score, which reflects the variety of the degree of care and support the residents need from the caregivers.

Graph 8 Variation in the Level of Support Required



The values in the horizontal axis are the raw scores on the Level of Care instrument. This graph shows that the first 25% of residents require a "normal" degree of support from their caregivers that is typical of any child in the community. The next group of children, from the 25th percentile to the 75th percentile, require a substantial degree of support that is typical of children who have diagnosed disorders. The last 25th percentile require a degree of support of very high risk children.

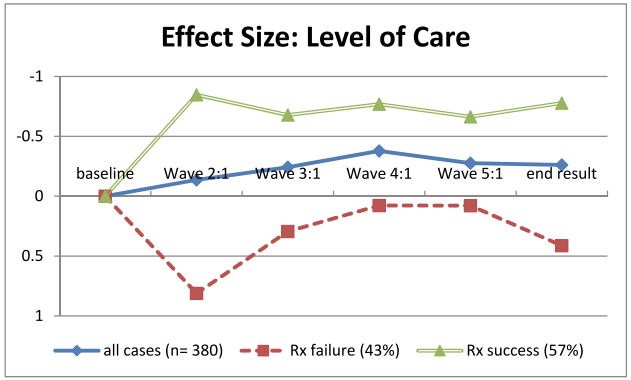
The Level of Care is moderately correlated with the child's CGAS scores (r = - .436, n = 461). High scores on the Level of Care instrument are associated with low scores on the CGAS. A regression analysis concluded that every increase on 1 point on the Level of Care produces a ½ point drop in the CGAS (standardized regression co-efficient or B = -0.436, t = -10.380, df = 459, sig= .000)

It is also correlated with the post treatment effect size for the CGI (r = 0.280, n = 383). Children who require high level of support on admission are more likely to have poor outcomes on the ADHD measure.

Graph 9 is referred to as the path to independence in the sense that a positive treatment effect indicates the child is more independent of the staff's involvement.

The graph shows the effect sizes for three groups of children: (a) all children tested, (b) Rx failures and (c) Rx success.

Graph 9 Path to Independence



Unlike the very large effects observed with psychiatric symptoms, children make small, but consistent steps towards independence. Fifty-seven percent of cases are treatment successes in the sense that they made a very large degree of progress (d = -0.847) from the baseline score and maintained this progress throughout their treatment.

Conversely, 43% of children are treatment failures because they required higher levels of care in the first nine months; these cases reversed direction at wave 3 slowly returning to their baseline level, and fall back again on discharge.

Attachment

All human beings yearn for a secure base and once achieved this is the most powerful source of resilience. Bayfield measures how much the children perceive themselves as having a secure base through the Parental Bonding Instrument (PBI). The PBI measures how securely the child is bonded to whatever *target-person* is identified on the test. The target may include foster parent, childcare worker or family of origin. The security of attachment is a critical element of resilience protecting disadvantaged children from adverse life span outcomes and improving the prognosis for change in children with symptoms of emotional disturbance or behaviour problems.

It is less evident to practitioners that attachment is subject to learning and insecure children can learn to become securely attached in as little as six months. (Crittenden[13], 2000, p. 355) Related to this finding is children usually form multiple attachments during their lifetime and this is especially true of children in high risk situations. Fonagy[14] (1993) followed 200 infants over time and found a child's ability to form a secure attachment to the *second* parent was not limited or enhanced in any way by the *first* parent. In other words, a child's ability to attach securely to his father, his foster parents or his adoptive parents is not influenced in any way by the quality of the mother-child attachment. Each parent or parent substitute has an independent chance to form a secure attachment to the child. "It appears that even the small child has the capacity independently to encode, differentiate and isolate the internal working models of primary caregivers" (Fonagy, 1993: page 240). Even the lack of secure attachment during infancy does not prevent the child from attaching to subsequent caregivers much later on in childhood (Stroufe[15], 1979; Goossens and van IJzendoorn[16], 1990; Howes and Hamilton[17], 1992a, 1992b).

The PBI has two scales: caring and over-protection. The adolescent is asked to identify the person in their life that they feel to be their closest caregiver. The form allows the adolescent to rate how affectionate this person is and how fairly he/she treats the teenager on issues of control.

Norms from cross cultural studies involving 800 teenagers have been established for the PBI. The scores obtained by the youth at Bayfield have been transformed to a t-score, with the normative mean of 50 and a normative standard deviation of 10.

[13] Crittenden, P. (2000), "A Dynamic-Maturational Approach to Continuity and Change in Pattern of Attachment", in *The Organization of Attachment Relationships*, Patricia Crittenden & Angelika Claussen (eds), Cambridge University Press: Cambridge, UK, pages 343-357

[14] Fonagy, P., Steele, M., Steele, H., Higgitt, A. & Target, M. (1993), "The Emanuel Miller Memorial Lecture: The Theory and Practice of Resilience", *Journal of Child Psychology and Psychiatry*, 35(2), 231-257

[15] Stroufe, L. A. (1979), "Early Experience: Evidence and Myth", *Contemporary Psychology* 22: 878-80

[16] Goossens, F.A. and van IJzendoorn, M.H. (1990) "Quality of infants' attachments to professional caregivers: relation to infant-parent attachment and day-care characteristics", *Child Development* 61, 832-7

[17] Howes, C, and Hamilton, C.E. (1992a), "Children's relationships with caregivers: mothers and child care teachers". *Child Development* 63, 856-9.

Howes, C. and Hamilton, C.E. (1992b) "Children's relationships with child care teachers: stability and concordance with parental attachments", *Child Development* 63,867-78

Caring Scale

The caring scale is a measure of how much affection the young person feels he receives from his closest caregiver. This scale is derived from a series of questions that the children rated on a four point agreement scale referring to their closest caregiver. The questions from the caring scale include:

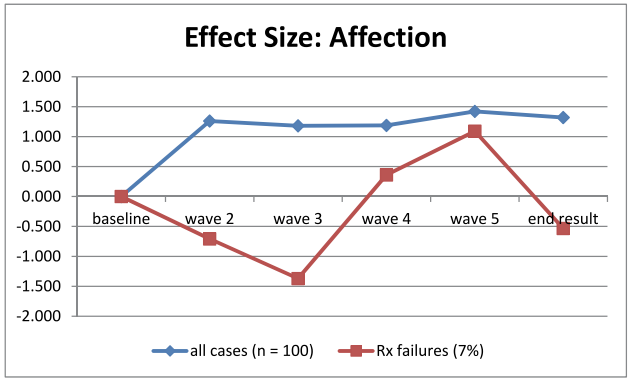
Spoke to me with a warm and friendly voice.

Enjoyed talking things over with me.

Could make me feel better when I was upset.

The effect sizes are graphically portrayed on the next page apply to children in the clinical range (t < 45). One hundred children (25% of the total tested) have very low scores on the caring scale. This sub-group of children have a mean caring t-score of 32. Less than 4% of children world wide perceive themselves to be so disliked by their parents.

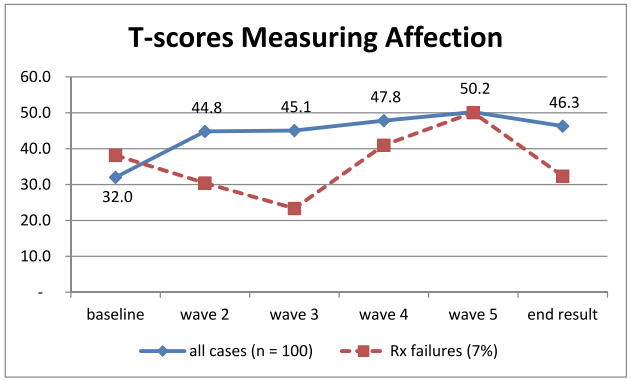
Graph 10 Children Feel They Receive More Affection over Time



Graph 10 demonstrates children admitted to Bayfield believing they are deeply disliked by their closest caregiver change this perception to a large degree (d = 1.262) in the first nine months of treatment. All of the changes recorded on this chart are statistically significant in addition to being clinically significant. The at-risk group of children who felt deeply disliked on admission continue to make progress while living at Bayfield, showing an effect size of d = 1.422 at wave 5. Over time, the group that were discharged showed a final effect size of d = 1.320.

Only 7 children from the original cohort of 100 cases did not follow this path. These children, labelled Rx failures, deteriorated markedly but even they recovered by wave 5. Unfortunately, most were discharged before the turnaround could be achieved, which is why the "end result" is d =minus 0.5.

Graph 11 The Path to a Secure Base



The path to a secure base is clearly demonstrated through the actual t-scores, showing this cohort actually achieves the normative mean at wave 5. The end result dips a little lower as many children are discharged before this path is completed.

Over-Protection Scale

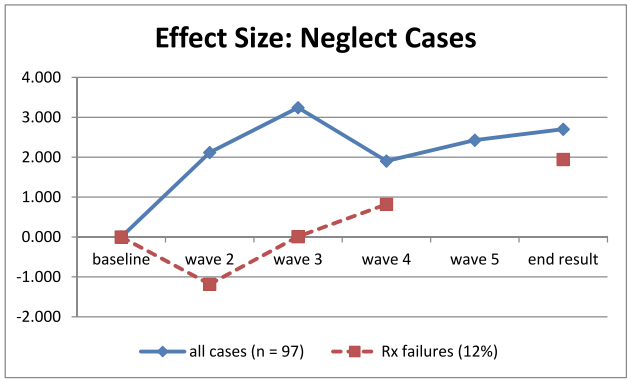
The over-protection scale of the PBI splits into two distinct sub-groups of clinical cases: neglect and unfairness. Neglect applies when the over-protection scale is very low (t < 45) compared to the norms indicating a belief that his/her caregiver did not provide enough supervision, rules and expectations to keep them safe. Unfairness applies when the over-protection scale is very high (t >65) compared to the norms indicating a belief that his/her caregiver was over-controlling and not respectful of his need for autonomy and capacity to act responsibly. Between the two extremes is the normal range t >= 45 and t <= 65. When the child perceives his relationship is within this range on the over-protection scale, the relationship is perceived as safe, respectful and authentic.

Neglect

Children with very low scores (t < 45) believe their caregiver did not keep them safe, and are considered neglected; 25% of the total number of children who completed the PBI meet the criteria of neglect. Neglected children have a mean t-score of 40.6 on admission.

The two distinct clinical ranges identified by the over-protection scale use a standard deviation specific to each clinical range as the denominator for the effect sizes. The denominator of the neglected cases is SD = 3.426.

Graph 12 Neglected Children can Feel Safe

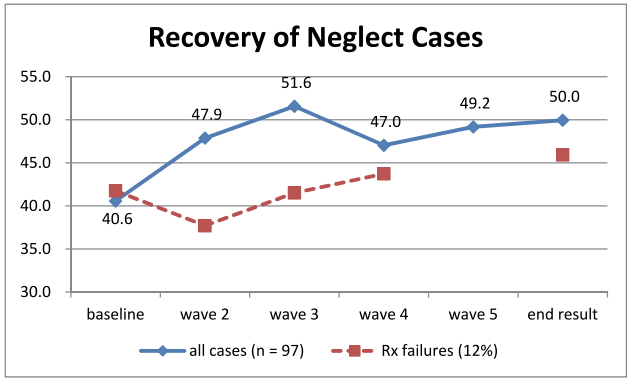


Children who feel neglected on admission make significant progress clinically in the first 18 months of treatment and this is maintained until discharge. The differences on the graph are statistically significant, indicating this result cannot be explained as a random variation.

Bayfield staff change the inner working model of the children to an exceptional degree on both perceptions: being the object of affection and being safe. The majority of special needs children (Groze, 1996; Goldberg, 2000; Crittenden, 2000) can form secure attachments, but they often retain dysfunctional aspects of their inner working model. Crittenden (2000) also points out that no matter what happens, the attachment relationship and the inner working model change continuously over time with each maturational stage and with significant

disruption and periods of stress. This insight provides the opportunity for treatment that is at the core of reciprocal influence.

Graph 13 The Path to a Feeling of Safety



This graph shows the children who meet the criteria of neglected on admission recover within nine months to the normal range and by 18 months, they are slightly above the average score for the normative reference group.

Children labelled as treatment failures deteriorate in the first nine months, but rebound at 18 months and by discharge they are within the normal range. (Note that there was one case at wave 5, therefore the dotted line does not go all the way to the end result).

Fairness

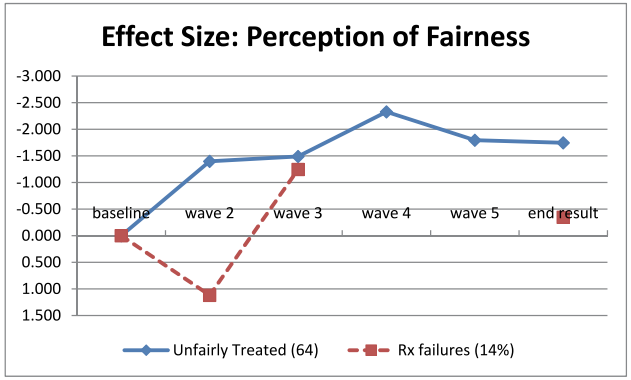
The over-control questions assess the child's perception of the level he/she was over protected by his closest caregiver. Children whose t-score is above 65 perceive their caregiver as intrusive, violated their boundaries, infantilized them and treated them unfairly. This type of relationship predicts children who have an insecure attachment pattern with their caregivers.

Fifteen percent (15%) of children who completed the PBI perceive their closest caregiver to be unfair ($t > 65$). The mean score for children in this clinical range is $t = 73.7$. Fewer than 1 in 1000 teenagers world wide perceive their

caregiver to be unfair. The denominator of being unfairly treated has a SD = 7.723.

The pathway to recovery for children who perceive their caregivers as unfair is highly correlated with the pathway to recovery for children who perceive their caregiver has having little affection for them. ($r = - 0.507$).

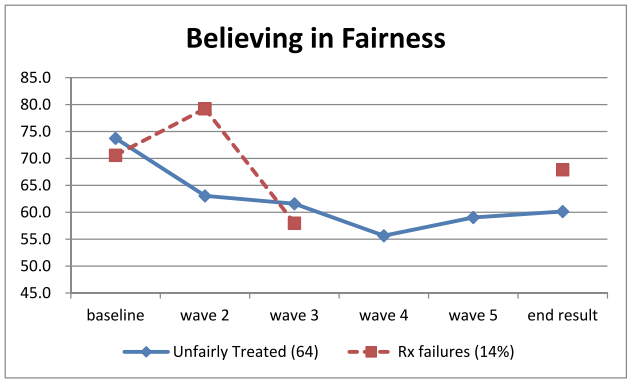
Graph 14 Effect Size: Perception of Fairness



This graph shows that children who perceive their closest caregiver as deeply unfair change this perception significantly ($d = - 1.398$) in the first nine months of treatment and again at 27 months ($d = -2.328$). We compared the overall pattern with a sub-group of children who deteriorated in the first nine months. The perception of a small group of respondents, 14% of the larger group who perceive the world as unfair, actually became significantly worse ($d = 1.1$), but then recovered by an effect size of $d = -2.3$ between wave 2 and wave 3. Because there was only case of the Rx failures left between wave 3 and wave 5, the graph does not continue.

Taken together, the pathway to recovery for children who perceive their closest caregiver as unfair is strong, positive, immediate and maintained over a long period of time.

Graph 15 Pathway to a Healthy Perception of Caregivers

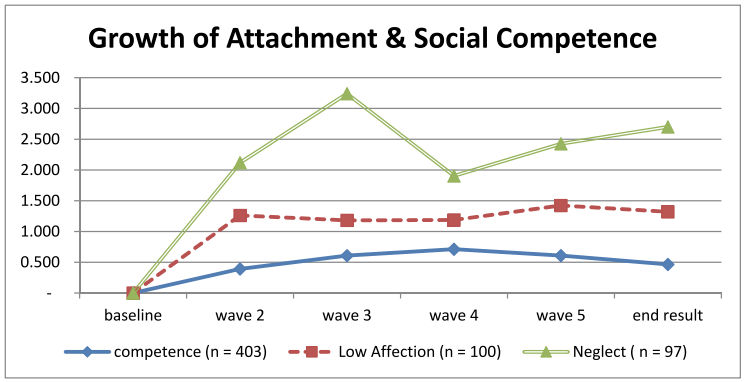


The children make steady progress towards perceiving the world as fair and respectful of their individuality. As a result, these children feel better and become more open to working with teachers and staff in improving their grades and social functioning.

Summary of Clinical Outcomes

The instruments used to assess outcomes are not correlated with each other when collected on admission, at any point in time thereafter or on discharge. This means that each instrument brings new information to help in understanding the child's needs and the impact of treatment. Graph 15 is a composite of effect sizes for different instruments. Effect sizes are standardized and allow the reader to compare results on a level playing field.

Graph 16 Growth of Attachment and Social Competence



Social competence is measured by the CGAS. The growth of all children is shown, indicating a large effect size across 45 months of treatment and concluding with the assessment near discharge. A large effect size in the growth of social competence is unusual. In most published studies, the effect size is half of the level achieved by the children in this study.

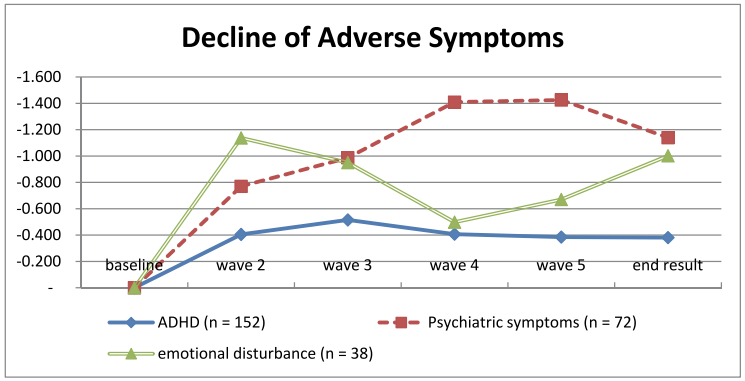
Children who believe that their closest caregiver has very little affection for them represent 25% of the child admitted to Bayfield. These children show a very large positive effect size throughout treatment. They end up feeling they are liked by their caregivers to a degree that is slightly above the normative mean. This has a significant impact on their sense of self-worth.

Children who believe their closest caregivers did not provide enough structure, supervision and expectations to keep to them safe, do not feel secure. Over the years, 25% of children admitted to Bayfield felt deeply unsafe. They experienced a profound change within the first nine months of treatment and continued to grow strongly to 27 months. The evidence indicates these children felt as protected as other children in society and were feeling safe at the time of discharge.

Growth in social competence and attachment are identified in the literature as the most powerful determinants of life span outcomes for children at-risk.

Graph 17 displays the rate of decline in adverse symptoms that persist into adulthood with treatment.

Graph 17 Decline of Adverse Symptoms



Three instruments are used to assess ADHD, mental illness and emotional disturbance. ADHD is measured by the Conners' Global Index. Forty percent (40%) of children admitted to Bayfield had clinically high scores on this instrument. These children showed a large effect size in lowering symptoms of ADHD. The decline of symptoms continued during the time in treatment and were much improved at the time of discharge.

Psychiatric symptoms were assessed using the SA-45. Twenty-three percent (23%) of adolescents had clinically high scores on admission. These children showed a very large effect size in reducing symptoms of mental illness. The decline in symptoms continued until it reached an apex at 36 months. Overall, the children were discharged in the normal range with a very large treatment effect.

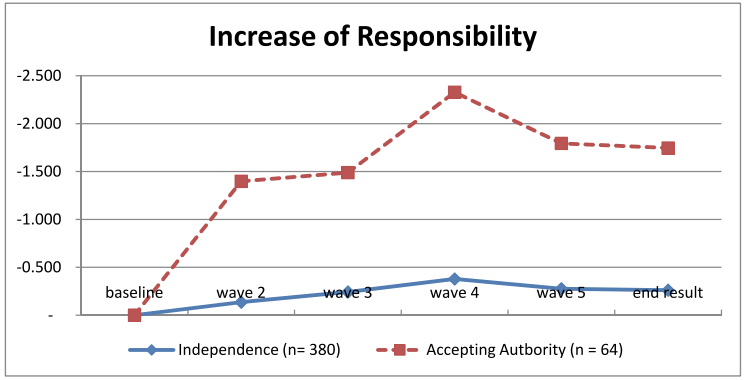
Emotional disturbance was assessed using the FAB-C. The FAB-C is used for school age children; once they reach 13 years, the children are assessed using the SA-45 for adolescents. Twenty-three percent of children were in the clinical range on admission. Based on the domains assessed by the FAB-C, emotionally disturbed school aged children admitted to Bayfield, make very large changes clinically reflecting:

- fewer conduct problems;

- better self-image;
- fewer worries;
- better peer-relations, and
- fewer antisocial attitudes.

Graph #17 shows the increase of responsibility. Children become responsible when they accept appropriate authority, supervision and rules in the social context and when they are capable of managing the demands of daily living on their own. Bayfield has two measures for assessing these domains: (a) the Level of Care and (b) the Unfairness scale of the PBI.

Graph 18 The Increase of Responsibility



The growing independence of the residents of Bayfield is demonstrated by the decline of dependence on staff support to manage the demands of daily living. The decline of the staff's support is shown by the solid blue line. The effect sizes for this construct are small but consistent and children are discharged more independent than when they were admitted. All residents of Bayfield are represented on this graph.

Conversely there is a very large effect size for the degree that children learn to accept authority, supervision and structure in their lives. This construct is indicated indirectly by the child's perception that their closest caregiver provides

appropriate structure, supervision and expectations for them. A group of children (n = 64 or 15% of pop) whose perception of the appropriate structure was so extreme in the unfair end of the continuum were included in this analysis. This high risk group of adolescents show a very large effect in changing their perception. As with psychiatric symptoms, the peak treatment effect takes place after 36 months.

Conclusion of Clinical Outcomes

Bayfield Treatment Centres provides a potent treatment service that increases social competence, attachment and personal responsibility and reduces symptoms of dysfunction. The changes over time are significant and they endure across 45 months of testing. The children are demonstrably healthier on discharge.

Academic Outcomes

Approximately 82% of the children admitted attend the Bayfield School. This is a private school accredited to teach the Ontario curricula. The teacher to pupil ratio is much higher and every class is supported with a child and youth worker.

The children admitted have experienced devastation in the public education system. Eighty-one percent have failed in school from primary grades (Sanders & Fulton, 2012).

In 2012, Bayfield conducted a qualitative analysis of five children to learn about process of treatment education in more detail. Five children were selected at random and entire Ontario Scholastic Record (OSR) was reviewed.

The OSR includes notes written by teachers starting when the child was in Junior Kindergarten. These early records clearly document severe family adversity. The Children's Aid Society (CAS) was involved in all cases before the child was in grade one. The children were identified by the Kindergarten teacher as being unable to function in school due to aggression and special needs. Three of the five children were suspended in either kindergarten or grade one.

School support staff and medical consultants at local hospitals diagnosed four out of the five children with developmental disorders before grade one including FASD, mild intellectual deficits, memory deficits and speech and language delay. All children displayed high risk behaviour including self-harming behaviour, animal cruelty and aggression. All of the children experienced multiple placements before being admitted.

Anecdotally, Bayfield staff and teachers believe this is a typical profile of the early education experience of the children admitted.

Adversity and Trauma

Bayfield School staff embrace the challenge to educate children with serious emotional and behavioural problems that started in Kindergarten and followed by years of school failure in the public school system. The core challenge of Bayfield School staff is amplified by abuse and family pathology, each contributing to the child's disability.

The charts below are indicative of the children's history.

HARDSHIP		
No	Yes	Total
206	178	384
54%	46%	100%

SEXUAL ABUSE		
No	Yes	Total
250	134	384
65%	35%	100%

PHYSICAL ABUSE		
No	Yes	Total
131	253	384
34%	66%	100%

ALL CHILD ABUSE		
No	Yes	Total
99	285	384
26%	74%	100%

Three quarters of the student population of Bayfield School suffer one or both forms of (sexual and/or physical) child abuse. The child's trauma is carried into the classroom. The trauma emerges during the curricula, especially in emotionally evocative stories from history lessons, poetry, drama, creative writing, art and discussion. It is also very difficult for the students to quiet their mind sufficiently to concentrate on math.

Many children have severe pathology in their family history that emerges in their emotional reaction to school. The highlights of the emotional overlay these children carry include:

Parent Incarcerated		
No	Yes	Total
255	129	384
66%	34%	100%

Parent Hospitalized for Mental Illness		
No	Yes	Total
336	48	384
88%	13%	100%

Parent diagnosed with Intellectual Deficits		
No	Yes	Total
344	40	384
90%	10%	100%

Parent with Substance Abuse		
No	Yes	Total
141	243	384
37%	63%	100%

Measurement of Outcomes

Bayfield started assessing academic outcomes in 2006 using the Kauffman Test of Educational Achievement (K-tea) and changed instruments in 2011 to the *Wechsler Individual Achievement Test* (WIAT). Both instruments cover the same academic skills, i.e., reading, math and oral learning; the WIAT has wider coverage. The WIAT also assesses listening skills and receptive vocabulary.

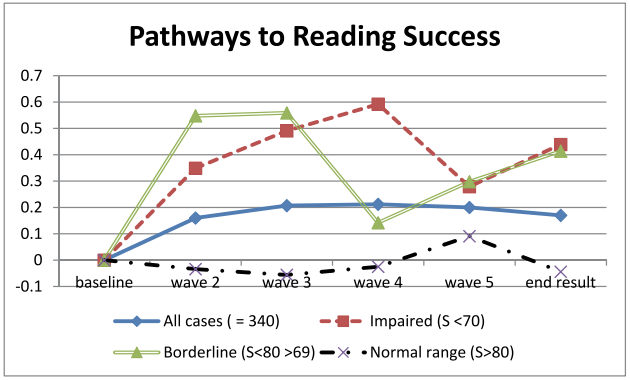
The key statistics from these tests is the standard score. The standard scores are created by comparing the child's performance in any area, e.g., reading, math, to the typical child in a classroom in Central Canada. The average standard score is 100. A child is considered to be impaired if the standard score is below 70, borderline if the standard score is between 70 and 79 and in the normal range if the standard score is above 80. Less than 2% of society is considered impaired.

A child who has the same standard score, e.g., S = 100, year after year is still learning more, and he is just learning at an average pace. A standard score of 100 is the normative mean.

Reading

The final effect size on discharge for reading comprehension was moderately correlated with success in standard scores for math problem solving. (r = -0.349, n = 220)

Graph 19 Pathways to Reading Success



Graph 19 shows children are not just keeping pace with their reading levels at baseline, but they are advancing toward greater success. However, the rate of advancement varies by the degree of impairment at baseline. Specifically,

(1) Children who were impaired (26% of cases)) made the greatest advances in the pace of learning with a moderate improvement at wave 2 (d = 0.349), a large improvement at wave 3 (d = 0.491) and a large improvement at wave 4 (d = 0.592) and eventually ending up at discharge with a large improvement overall (d = 0.414)

(2) Children who were in the borderline range (13% of cases) respond quickly and to the greatest degree at wave 2 (d = 0.548) and end up in the same place as the impaired children on discharge (d = 0.414) with a large effect size.

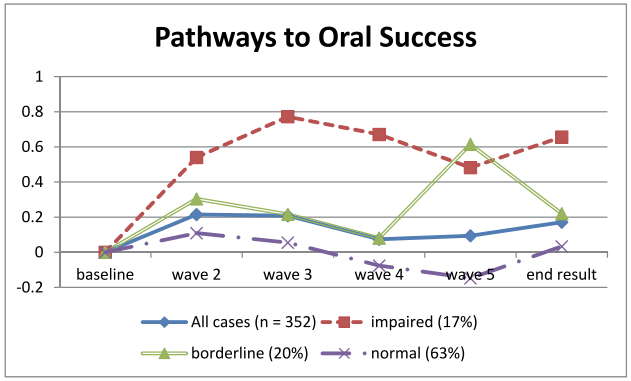
(3) Children who are in the normal range on admission (61% of cases) remain at that level throughout their service at Bayfield School. The effect sizes skirt below the zero line but this is not significant and should be interpreted as "no change".

Therefore, the program is successful in improving reading to a large degree for 39% of the children who begin their education at borderline or lower levels of functioning.

Oral Learning

The full name for this domain of learning is oral discourse comprehension. Progress on this skill is not correlated with reading or math.

Graph 20 Pathways to Oral Success



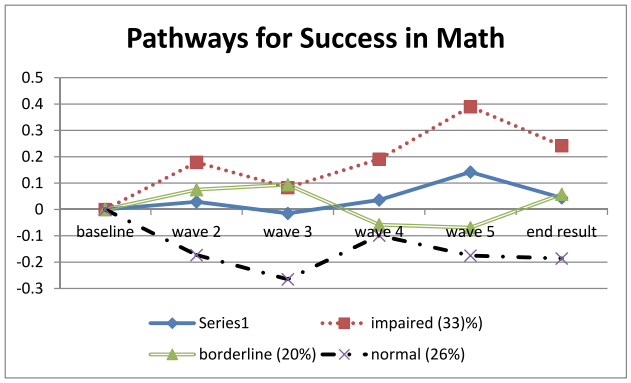
Children who are impaired on admission make very large improvements on oral learning by wave 3 (d = 0.772), and maintained to the point of discharge. The pathway for children in the borderline range follows the same track as for all cases with a small improvement overall (d = 0.171).

As with reading, children in the normal range on admission do not change in relation to the standard scores for oral discourse comprehension.

Math

Math problem solving is clearly dependent on being able to read the problem. As expected, success in math is correlated with success in reading (r = -0.349, n = 220).

Graph 21 Pathways to Success in Math



The children who start out in the normal range in math standard scores decline to a small degree by wave 3 (d = minus 0.265) before a small recovery by discharge, but still a small decline overall (d = minus 0.186). Children in the borderline range criss-cross the track for all cases and end up with no change overall.

Children who were impaired on admission in relation to math standard scores show steady improvement throughout their treatment and are discharged with a small effect size (d = 0.242).

Listening Comprehension

There were only 24 cases overall with a pre-post pair to analyze in the listening comprehension variable. When the cases were split up into impaired, borderline and normal, the numbers were too small to interpret. Overall, there was a small improvement in listening comprehension (d = 0.15).

Receptive Vocabulary

There were only 23 cases overall with a pre-post pair to analyze in the receptive vocabulary variable. When the cases were split up into impaired, borderline and normal, the numbers were too small to interpret. Overall, there was a small improvement in receptive vocabulary (d = 0.21).

Conclusion of Academic Outcomes

All of the clinically significant effects described above are statistically significant, meaning they could not be explained as a random variation. The primary hypothesis for this study was that children attending school at Bayfield would do much better than simply maintain their level of academic functioning measured on admission.

The primary hypothesis was true for children who were impaired on admission. This group represents between 17% and 33% of children depending on the subject. Impaired children made consistent improvements that ranged from small to very large. Conversely, children in the normal range tended to maintain their academic functioning. All children advanced in grade levels. Children who are within the normal range do not advance in relation to their peers. This means that if their performance was "low average" it did not change. Our goal should be to improve this group to the average range or above.

Summary and Recommendations

The Bayfield outcome study for 2013 was conducted by testing hypotheses with high expectations. We were able to confirm our expectations in the domain of clinical outcomes for attachment, psychiatric symptoms and social functioning. We were less successful in reducing symptoms of ADHD. In academic outcomes, our success varies by the level of impairment on admission.

This research demonstrates that Bayfield has produced excellent outcomes for the children served, however there are a small number of areas and sub-groups of children who could do even better. The last mile is always the most difficult. In order to reach even better results, Bayfield staff require more feedback from the outcome testing, more frequently and on a case by case basis. Therefore, we recommend:

(1) Effect sizes for each child and each test after the baseline are calculated through the child's file on the BIS.

(2) The child's personal effect sizes are reported on the plans of care, every 90 days, leading to changes in instruction or treatment interventions in order to improve the effect of service.

(3) Details about changes made at each feedback loop are recorded in the child's Plan of Care and become part of the child's treatment plan.

Bayfield Treatment Centres provided residential treatment to 551 children between March 15, 2011 and March 15, 2013. All but 2 of these children were tested using the Bayfield outcome measures, a 99.6% compliance rate. The Bayfield outcome measures are evidence based instruments are well established in the literature as assessment and outcome measures. The domains measured by the six instruments are critical areas of the child's life that predict lifespan outcomes. They are: attachment, hyperactivity and inattention, social functioning, affect regulation, psychiatric symptomatology, and academic performance.

Recommendation 1

Accelerate the test cycles to quarterly intervals. This approach to outcome measurement will provide important feedback on the child's response to intervention.

Recommendation 2

Continue with the current measure for attachment, the PBI, for young people ages 13 and older. Add the Kern`s Security Scale to the attachment measures. The Kern`s Security Scale is validated and appropriate for children ages 8 years to 14 years.

Recommendation 3

Children who are functioning at a higher level on admission, need a special treatment protocol that starts immediately upon admission to stop the pattern of deterioration.

Recommendation 4

Since both good and poor responders are unable to get beyond the mean CGAS of 50 to 59, a new social competence building program should be developed to teach higher social functioning.

Recommendation 5

A substantial number of children admitted to Bayfield (43%) are very dependent on adults support through their lifespan. This is a typical pathway of very disturbed children[18]. According to Kazdin (1995), they respond best to high intensity interventions in the beginning followed by a long term, low intensity, high support maintenance program. Therefore, we should not think of the lack of

[18] Kazdin, Alan (1995), "Scope of Child and Adolescent Psychotherapy Research: Limited Sampling of Dysfunctions, Treatments and Client Characteristics", *Journal of Clinical Child Psychology*, 24 (2), 125-140

progress on independence as a treatment failure but as normal and appropriate for children with chronic disabilities.

Recommendation 6

Learning math skills is the area that challenges children to the greatest degree. We have learned that children with an emotional overlay caused by adversity are impeded from learning abstract concepts (Sanders & Fulton, 2006). This is true for the general population as well and raises questions about the Ontario curricula. Achieving a large effect size in this domain would make a great contribution to evidence based education and the residents of Bayfield. Therefore, we recommend a *Bayfield Math Innovation Project* running alongside the standard Ontario curricula.

The Bayfield Math Innovation Project would:

1. be applied to students with low normal or normal functioning at baseline
2. be driven by feedback from monthly test scores
3. be designed with instructional methods and curricula drawn from the literature whether it is within the Ontario curricula or not.

Recommendation 7

This research study demonstrates Bayfield has produced excellent outcomes for the children served. There are a small number of areas and sub-groups of children who could do better. The last mile in the marathon of treatment is the most difficult. In order to realize better results, Bayfield staff must have more feedback from the outcome testing, more frequently and focussed on a case by case basis.

References

Beitchman, J. H. (1996), *Feelings, Attitudes and Behaviors Scale for Children (FAB-C)*, Technical Manual, Toronto: MultiHealth

Borduin, C., Mann, B., Cone, L., Hengeller, S., Fucci. B., Blaske, D., & Williams, R. (1995), "Multisystemic Treatment of Serious Juvenile Offenders: Long-Term Prevention of Criminality and Violence", *Journal of Clinical and Consulting Psychology,* 63, 4, 569-578

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates

Crittenden, P. (2000), "A Dynamic-Maturational Approach to Continuity and Change in Pattern of Attachment", in *The Organization of Attachment Relationships*, Patricia Crittenden & Angelika Claussen (eds), Cambridge University Press: Cambridge, UK, pages 343-357

Fonagy, P., Steele, M., Steele, H., Higgitt, A. & Target, M. (1993), "The Emanuel Miller Memorial Lecture: The Theory and Practice of Resilience", *Journal of Child Psychology and Psychiatry*, 35(2), 231-257

Galéra, C., Melchior, M., Chastang, J. Bouvard, M. & Fombonne, E. (2009), "Childhood and adolescent hyperactivity-inattention symptoms and academic achievement 8 years later: the GAZEL Youth study", *Psychological Medicine*. 39(11): 1895–1906

Gresham, Elliott & Horner, Robert (2001), "Interpreting Outcomes of Social Skills Training for Students with High-Incidence Disabilities", *Exceptional Children*, 67 (3), 331-344

Goossens, F.A. and van IJzendoorn, M.H. (1990) "Quality of infants' attachments to professional caregivers: relation to infant-parent attachment and day-care characteristics", *Child Development* 61, 832-7

Guzder, J., Paris, J., Zelkowitz, P. & Marchessault, K. (1996), "Risk Factors for Borderline Pathology in Children", *Journal of the American Academy of Adolescent Psychiatry*, 35 (1), 26-33

Howes, C, and Hamilton, C.E. (1992a), "Children's relationships with caregivers: mothers and child care teachers". *Child Development* 63, 856-9.

Howes, C. and Hamilton, C.E. (1992b) "Children's relationships with child care teachers: stability and concordance with parental attachments", *Child Development* 63,867-78

Kazdin, Alan (1995), "Scope of Child and Adolescent Psychotherapy Research: Limited Sampling of Dysfunctions, Treatments and Client Characteristics", *Journal of Clinical Child Psychology*, 24 (2), 125-140

Newman, F.L., Richard H. Hunter, R.H. & Irving, D. (1987) "Simple Measures of Progress and Outcome in the evaluation of Mental Health Services", *Evaluation and Program Planning*, 1987, vol 10, 209-218

Rutter, Michael and Seija Sandberg (1985), "Epidemiology of Child psychiatric Disorder: methodological Issues and Some Substantive Findings", *Child Psychiatry and Human Development*, 15(4), 209-233

Sanders, L.S. & Fulton, R.J., (2008), "Analysis of Bayfield Treatment Model and Educational Outcomes", University of South Africa, Canadian Psychological Association Ambassador to South Africa.

Sanders, L.S. & Fulton, R.J., (2012), "In Search of a Healing Praxis, A Quantitative and Qualitative Study in Context of Attachment Theory, Positive Psychology and Emotional Intelligence", International Positive Psychology Association World Congress, Los Angeles, CA, USA.

Sroufe, L. & Rutter, M (1984), "The Domain of Developmental Psychopathology", *Child Development*, 55, 17-29

Stroufe, L. A. (1979), "Early Experience: Evidence and Myth", *Contemporary Psychology* 22: 878-80