



# Young Offenders (Serious Crimes) Bill

---

## **Youth Horizons Addendum to Oral Evidence to the Law and Order Select Committee**

Contact Person:

James Liddell  
Ph: (04) 8030200

[jamesl@youthorizons.org.nz](mailto:jamesl@youthorizons.org.nz)

## **Introduction**

1. We are grateful for the opportunity extended by the Chairman of the Committee to present addenda to our oral submission. We would like to restate our opposition to this Bill and confirm that we stand by the contents of our original written and oral submissions. These addenda focus on the two major concerns the Committee had during our oral submission: a conflict of interest, and the (supposed) inefficacy of Multisystemic Therapy as detailed in the Reducing Youth Offending Programme (RYOP) evaluation report.

## **Concerns over conflict of interest**

2. We inferred from one member's questions that that we stand to make substantial gains (financial and reputational) from the unaltered continuation of the youth justice system, and that this was a significant motivation for our opposition to the Bill. We would like to explicitly state that our motivation for opposing this Bill is unrelated to any losses Youth Horizons might incur should this Bill become law. On the contrary, the alternative interventions we suggested in our submission would result in no financial gains to the organisation and could conceivably result in a loss of income.
3. We advocate a number of earlier interventions focused on at-risk mothers, families and pre-school children. Youth Horizons' services for conduct disordered children and young people do not accept individuals under the age of 10. The services we provide to children under this age are for the purposes of addressing and treating emergent mental health problems such as psychosis (Early Intervention Service, Emergency Respite Foster Care). Our organisation currently does not have the expertise nor the resources to deliver services to very young conduct disordered children at risk of becoming offenders. Hence, should government implement programmes to address and treat these would-be offenders as we advocate, enrolments in our programmes for young offenders aged 10 years or over could conceivably drop markedly (assuming, as research suggests, that these programmes would be more effective when targeted at a younger age group). Despite this we are supportive of such programmes and believe it is incumbent on government to identify and treat these children as early as possible.
4. Our opposition to this Bill is predicated on the belief that children and young people should be provided with the very best resources and opportunities to prevent their offending. As our original submission shows, the research is explicit that psychological and ecological treatments have far greater efficacy than punitive measures; punitive measures are ineffective and in many instances result in serious iatrogenic effects that see a young person's offending worsen (both in

total number of offences and in severity of offences). Boot camp, prison and prosecution as an adult simply do not work to lower the severity or rate of offending.

### **Multisystemic Therapy and the Reducing Youth Offending Programme (RYOP)**

5. One member questioned our advocacy of Multisystemic Therapy (MST) in light of an evaluation of the Reducing Youth Offending Programme (RYOP) authored by Grace R, McLean A, & Warren J in April 2006, herein referred to as the 'RYOP Report'. We are grateful to have the opportunity to comment on the RYOP Report, which has not been published and was not previously made available to us.
6. The RYOP pilot was a joint initiative of the Department of Corrections and the Department of Child, Youth and Family. It should be clarified that Youth Horizons did not have a role in this project.
7. We would like to emphasise that the RYOP Report does not conclude that MST is an ineffective treatment but rather that there have been problems with the programme's implementation which we believe are explained by poor adherence to the treatment model. We applaud the initiative in establishing the Reducing Youth Offending Programme but believe that adherence to the treatment model ought to be strengthened in order that the clinically significant positive outcomes experienced by other MST programmes (both in New Zealand and around the world) can be replicated by RYOP.
8. We would like to state explicitly that our organisation has no financial stake in MSTNZ or any other professional group responsible for the advocacy or oversight of MST in New Zealand. Youth Horizons uses MST as one therapeutic tool in its treatment arsenal and in our experience MST has produced generally good outcomes for the young people whom we treat with this approach.
9. Our analysis of the RYOP Report is critical on six points: treatment integrity, the statistical model implemented, partial treatment at the time of data collection, the use of psychometric data, incongruent offence data from MSTNZ, and the age of participants. These are discussed briefly below; a fuller critique of the RYOP Report is included as an appendix at the back of this document for the interested reader.
10. *Treatment integrity.* The evaluation highlighted deficiencies in adherence to the therapeutic model by therapists, managers and supervisors. The average caseworker spent only up to one hour per week with each family (four hours is the required minimum for MST) and had little or no psychology training. Most caseworkers came from social work or probation backgrounds; the usual requirement in the US

for an MST caseworker is a postgraduate qualification in psychology and clinical training. Many RYOP caseworkers were resistant to clinical supervision. Linking their analysis of Therapeutic Adherence Measures (TAMs) to the rate of offending, the authors hypothesise that better adherence is linked with reduced offending. This backs up the findings of other studies into MST therapist adherence (e.g. Hengeller et al. 1997). We are, however, critical of the methods and data used in the analysis and thus believe that the RYOP outcomes may have been misrepresented.

11. *Statistical model used for comparison.* We are highly critical of the methodology employed to compare the rates of reoffending of those in the RYOP sample and those in the (supposed) control group – the centrepiece of the RYOP Report which claimed that RYOP was ineffective. It is critical to note that the investigators' methodology is entirely novel (which they admit); their study was not a randomised clinical trial (essential for evaluating medical and psychological therapies), nor a matched group design (the next best option). Instead, the investigators constructed a statistical model of reoffending behaviour from archival data – a method not tested or validated on any sample other than their own. We have significant concerns about the validity of this methodology and believe that a simpler and more traditional research design (such as a randomised clinical trial) would have been a far better tool for evaluation.
12. *Partial treatment at the time of data collection.* The RYOP Report states that the great majority of youth treated by RYOP reoffended within 40 days of entering the programme and are hence judged to have failed intervention. A complete MST intervention, however, is five to six months' duration. The authors disregard this and count any reoffending after the start of the intervention as a failure, not from conclusion of the treatment. In this way MST is judged and assessed on the basis of incomplete interventions; the evaluation does not consider those individuals who may have offended after the start of treatment, but who have stopped offending or reduced their offending at the conclusion of the programme.
13. *Psychometrics.* The authors found only one dimension to be statistically significant: RYOP individuals improved on the Socialised Aggression scale. However, the basis for the authors' claims in this section (that there were no statistically significant improvements in the other dimensions) are fundamentally flawed in that a very low number of assessments were completed; for some measures only 10 completed assessments were obtained (out of the 223 RYOP participants). It is almost impossible to extrapolate findings from such a limited sample to the rest of the population with any validity.
14. *Offence data from MSTNZ.* Offence frequency data from MSTNZ shows a general trend towards reduction over the course of the study, from pre-programme to 24 months post-intervention. The authors are

critical of these data and downplay them, erroneously we believe. The authors' basis for the rejection of the significance of these data is based on several flawed assumptions (see appendix for more detail).

15. *Age of participants.* It is very likely that if, despite the many problems with methodology and assumptions contained within the RYOP Report the evaluators' conclusions are valid, a factor contributing to the inefficacy of MST in RYOP is most certainly the age group of the participants. RYOP participants had an average age of 17.1 years at referral. This is significantly higher than the average age in the well known US trials: Borduin et al. (1995) involved young people with an average age of 14.8 years and Henggeler et al. (1997) involved participants with an average age of 15.2 years. Increasing age is likely to have negative impacts upon treatment implementation because as the young person ages, the systemic and ecological importance of family association and influence diminish. This is of high importance because MST specifically exploits the family and familial networks during the young person's treatment.
16. In summary, the evaluation report of the Reducing Youth Offending Programme exhibits several major flaws and must be approached with caution. The study design, data collection, and data analysis cannot be considered to support the conclusions drawn by the authors. However, measures of adherence to the MST model were valid and suggest that, in effect, MST was not delivered by RYOP therapists. The RYOP therapists' lack of relevant training and experience prior to undertaking MST, their poor adherence to the treatment model and reported attitudes to supervisory and clinical processes, and the average age of participants being two years greater than that of international peer-reviewed studies that have found MST to be effective are factors that would have contributed to any inefficacy of MST delivered by RYOP. Thus, conclusions drawn from this study cannot be generalised to the delivery and outcome of MST provided by other agencies in New Zealand.

## **Conclusion**

17. Finally we would like to thank the Committee for the opportunity to present our oral evidence and to submit these addenda, and we reaffirm our opposition to this Bill.

## ***Appendix***

# **Commentary on *An Evaluation of the Reducing Youth Offending Programme*<sup>1</sup>**

By Charles Heywood, PhD  
Youth Horizons Psychologist

## **1. Introduction**

The Report is an evaluation of a trial of Multisystemic Therapy (MST) commissioned by Child, Youth and Family and the Department of Corrections and authored by Randolph Grave and Anthony McLean of the University of Canterbury (Canterprise) and Julie Warren of the Centre for Research, Evaluation and Social Assessment (CRESA). This is a major study in the New Zealand context and is likely to have both clinical and political implications given the current climate of juvenile criminal law reform. The signal message or theme of this Report is that MST as trialled in New Zealand has not been effective in reducing Youth Offending or in having a positive impact upon any other outcome assessed. This is in stark contrast to most, but not all, overseas research. This paper is a critique of the RYOP study as represented in the RYOP Report. For reasons of brevity a familiarity with the aims of RYOP, MST and the evaluation Report itself will be assumed.

## **2. Treatment integrity**

The evaluation found multiple indications of poor treatment integrity across sites. MST is a complex and difficult intervention to implement and many aspects of the intervention were performed in a less than optimal way. In a number of prior studies, variations in the success of MST programmes have been found to be associated with the degree of adherence to the principles of MST by competent clinicians (e.g. Henggeler et al. 1997). Even well-established interventions or treatments may fail if core protocols are omitted or attempted without the appropriate skills or training. To cite just a few examples from the RYOP Report:

- “MST adherence is poor for the average RYOP caseworker” (p.39)<sup>2</sup>
- “MST-NZ acknowledged the poor quality of Caseworkers’ paperwork and by implication their analytical thinking” (p. 39)
- “A mix of structural and other interrelated factors ... undermined the effectiveness of clinical supervision in the first two years” (p.39)
- “Some caseworkers were ambivalent about clinical supervision and the requirement to adhere to a prescribed model” (p.39)
- “The absence of clinical authority meant that Caseworkers questioned the credibility of clinical decisions and were often resistant to clinical advice” (p.44)

---

<sup>1</sup> Grace, R., McLean, A. & Warren, J. (2006) *An Evaluation of the Reducing Youth Offending Programme* University of Canterbury (Canterprise) and Centre for Research, Evaluation and Social Assessment, Unpublished report.

<sup>2</sup> All page numbers given in this commentary relate to the page(s) of the report by Grace et al.

- “The outcomes included Caseworkers being reluctant to take clinical advice from the clinical supervisor and caseworkers and the manager questioning clinical decisions” (p.46)
- “...some caseworkers openly displayed resistance to clinical supervision. Others were either reluctant or unable to incorporate analytical thinking about a case and the related paperwork” (p. 54)

Other problems included:

- With respect to TAMs (Therapeutic Adherence Measures – a measure of the degree of caseworker adherence to the MST model completed monthly by families) the authors note that “...TAM scores suggest that adherence is poor for the average caseworker and that scores have not improved ... as the therapy teams mature” (p.58).
- The authors conducted various analyses of TAMs scores; mean scores on none of the six domains measured by the TAMs reached the target minimum level required for adequate adherence to the treatment model; scores did not improve over time, nor did they differ across sites.
- The authors also analysed the relationship between TAMs scores and reoffending in the RYOP sample, hypothesising that better TAMs scores may signal better adherence to the MST model and hence lower offending. Many multiple regression analyses were performed but only one cited, which showed a significant relationship between rate of conviction and one of the six TAMs domains.
- The duration of caseworker-family contact required by MST is stated as four hours per week during the early weeks of the intervention; caseworkers in the RYOP study report spending only up to one hour with each family per week (p.63).
- MST involves challenging the views and habits of families, which is reflected in the TAMs factor “therapist attempts to change interaction”. The authors remark that “caseworkers’ comments suggest they are more interested in being welcomed and liked than being challenging and effective” (p.64).
- The first Christchurch team supervisor was not experienced in MST, undermining the clinical authority of this role.
- Most MST caseworkers had social work or probation officer backgrounds which, in New Zealand, involve little clinical training or supervision; they were rarely psychology trained and rarely had post-graduate qualifications. By way of contrast the caseworkers in the original US studies conducted by the developers of MST were mostly postgraduate psychology students trained to work within a clinical model guided by structured supervision. Often these students had a year or more of direct post-qualification experience in clinical work, often with adolescents – the target group for MST (cf p.65). Furthermore, most MST therapist jobs in US jurisdictions currently require at least a Masters degree in psychology or social work and clinical experience. Any paucity of clinical experience or

qualifications is likely to be a major disadvantage given MST's reliance on techniques derived from Cognitive Behaviour Therapy (CBT) and family therapy. Interestingly, "MST NZ has also questioned the competency of some of the caseworkers" (p.69).

### **3. Comparison of RYOP reoffending with a statistical model of reoffending**

The centrepiece of the RYOP study is a comparison of the rates of reoffending which occurred in the sample of RYOP participants (n=223) with a statistical model of reoffending behaviour generated from an analysis of 4307 young persons who received a Youth Justice intake to CYF in 2002.

The lay reader may be misled by statistical wizardry into believing that the RYOP study constitutes a powerful methodology for assessing the efficacy of the RYOP programme; it is not. The approach used is manifestly weaker than other methodologies. The "gold standard" in clinical research is the randomised controlled trial (RCT), where participants are randomly assigned to a treatment group (who receive the therapy to be evaluated) and a control group (who do not receive the therapy) and a comparison of outcomes performed on this basis. The next best option is to match study participants and members of a comparison group individually on key variables of theoretical relevance to the domain under investigation (age, sex, SES, IQ etc.). The "successful" MST studies upon which the reputation of MST was built were all randomized controlled trials (Borduin et al, 1995; Henggler, Melton & Smith 1992). The RYOP study is neither a randomised controlled trial nor a matched group design; it is at best a quasi-experimental non-equivalent group design – a weaker methodology which must be interpreted accordingly.

Using information from the main CYF sample, the authors created regression models M1, M2 and M3 (which involved the variables XBeta1, XBeta2 and XBeta3) in order to predict reoffending outcomes YJPost, YJPostORConvicted (YJPOC) and Serious Prosecution respectively. All models were found to be statistically significant. However, these models all have indications that they capitalise on chance, lack generality, and were not tested on an independent sample; they are thus weak models against which to test obtained offending levels (from the RYOP sample).

Capitalising on chance is a risk in regression models where there are a wealth of predictor variables which can be entered into the model. Even if purely random data sets are entered into a regression model, sooner or later (purely by chance) the overall model will be statistically significant. It must be noted that the authors had access to NIA and CYRAS databases and each of these code large numbers of variables for each young person. Tell tale signs of this are:

- a) that the variables in the model appear arbitrary and have little theoretical relevance to the predicted outcome, and

- b) that the model lacks generality – e.g. if used with a different sample it fails to predict, or has to be modified with a slightly different set of variables to predict.

In this context it must be noted that M1 is not general: it uses different algorithms for Christchurch, Auckland and Elsewhere (p. 127). Thus, recidivism is related to previous sexual abuse in Christchurch but not Auckland, and sex is relevant Elsewhere but not in Christchurch. These are arbitrary differences which jar with theoretical conceptions of delinquency. Note that all of the models share this parochial bias as they all use XBeta1 from M1. M2 and M3 involve predictor variables related to prior intelligence and prior “Occurrences” from the NIA database; it is not clear what these are, but they do not resemble any known driver of conduct problems or delinquency – their presence is either adventitious, or because they correlate with an unknown other variable, or even due to the predicted outcome being itself semi-arbitrary (i.e. is YJPOC a good index of actual offending?)

A significant problem with the authors’ approach is that they did not test their models on an independent sample. Instead, the authors tested the models on a sample they called the “validation sample”, but this was itself a sub-sample of the original 4307 CYF cases that were used to develop the model; hence the models were tested on essentially the same sample on which they were developed (cf. p.140). The close correspondence between the actual and predicted survival curves only serves to show that the authors had done a very good job of creating a representative subsample. The models are essentially unverified on any truly independent sample and must be interpreted as such. Their utility for extrapolation to any other sample population is thus severely limited.

There is a genuine risk that the models M1-3 yield idiosyncratic predictions which lack generality. Such actuarial models are often unstable and unreliable (unless very carefully standardised); for instance Kjelsberg (2005) used Cox’s regression to look at offending in conduct disordered adolescents and found quite different survival functions for different cohorts of youth from the 1960s to the 1980s. M1-3 are based on CYF and NIA data from 2002; if these models lack generality even within New Zealand (different algorithms for Auckland & Christchurch) how can we be sure that were the exercise repeated with 2003 or 2006 data the derived models would be different yet again? We cannot.

The authors calculate the *estimated baseline survival function (EBSF)* (p. 161) using the mean XBeta score of the RYOP group and compare this with the survival function obtained from the actual rate of offending for the group (Figures 11.6-11.14). This is problematic as the EBSF is essentially the survival curve for an individual who is perfectly average on all the covariates in the model – however no such individual may exist; how does one average male vs. female, Auckland vs. Christchurch and sexually abused/not sexually abused? Insofar as no individual or group in the sample is “average in everything”, the predicted survival function is a meaningless extrapolation. If the data was segregated into analyses for say, male, female, older, younger,

then some groups may have achieved better than the predicted survival functions.

Substitution of the average XBeta score into the EBSF equation to create predicted survival curves for the RYOP group assumes that the XBeta scores are a valid index of risk (as reoffending is likely to be proportional to risk). However, as we have seen, the validation of the models M1-3 and their XBeta scores was of debateable validity; if the XBeta scores systematically underestimate risk, they will generate a “predicted” survival curve of less severity (i.e. less steep) and give the undue impression that the actual survival functions were worse than predicted.

If on the other hand the XBeta scores are a valid or accurate indicator of risk, the RYOP group were of very high risk (for example their average XBetaM3 score fell at the 86<sup>th</sup> percentile compared to the CYF sample of 4307 [p. 154]). It is commonly established that greater risk connotes higher rates of re-offending. Given that models M1-3 were derived from this more normal sample, it is difficult to know if the “predicted” survival curve for the RYOP group will be accurate even when the RYOP average XBeta score is used as a covariate. This supposes that the difference between the average and extreme members of the group is simply a matter of quantity, not quality. In other words, offenders at the 86<sup>th</sup> percentile may be a different type from those at the 50<sup>th</sup> percentile, and models from one may not predict the other.

After showing that the RYOP participants reoffend at a greater rate than the rejected referrals the authors note that this might be due to the former having a greater level of risk as they were selected on the basis of a high score on the Youth Risk Survey. In a series of analyses they used XBeta scores to control for risk in the RYOP group. Compared with the group as a whole (usual services control group, rejected referral and RYOP participants), programme completers (those that did not drop out of MST) reoffended at a lower rate than average for the group, suggesting that MST reduced offending to some degree (not significantly) when the sample was controlled for risk. This was the case for both serious prosecutions and YJPOC. It is interesting that in the main analyses the survival functions are said to be poorer for RYOP on the basis of visual inspection despite no statistical test being provided, yet the present results are discounted.

#### **4. MST was only partially delivered when the key data were collected**

Although the mean follow-up time (time to reoffend) for the RYOP group was 132 days, it can be seen that the great majority of Young Persons reoffend within 40 days of programme entry (p. 158). The expected duration of a complete MST intervention is 5-6 months, yet the young persons are judged to have reoffended, and the intervention failed (arguably), when in most cases the intervention is less than 50% complete. This is problematic for two reasons:

- a) MST is judged and assessed on the basis of incomplete interventions, and

- b) there is a potential bias with survival data; a youth who is arrested early in treatment but whose offending rate decreases over 6 months or a year will be judged a “failure” by survival analysis but a “success” when offending is measured as a rate per unit time. The potential bias to outcome assessment should be obvious.

## 5. Psychometrics

The psychometric results were mixed, and only one significant result reported (RBPC Socialised Aggression: significant difference, RYOP more improved). Self Report Delinquency (SRD), scores halved pre-post (no control group comparison), however the authors comment that “It is most improbable that these reductions ... had anything to do with participation in RYOP” and go on to mention an ANOVA (analysis of variance) which was evidently non-significant (no interaction), but fail to give the ANOVA results.

The most striking aspect of this section is the low number of completed assessments; for some measures only 10 completed assessments were obtained (despite there being 223 RYOP participants). Either the authors placed little stock in psychometric assessments or they had considerable difficulty organising these. The SRD scale is particularly important as much offending is covert and arrests and prosecutions tend to under-estimate rates of anti-social behaviour.

## 6. MST Offence frequency data

Offence frequency data from MST NZ showed a general trend towards a reduction in the rate and severity of offences over the course of the study, from pre-programme to 24 month follow-up (pp. 186-7). Only a few young persons had progressed to the latter, but 52 had been followed up at 12 months. The authors are critical of these figures and downplay them. They note, erroneously, that the figures are based on self or informant report (when they are actually based on the caseworker’s report, where the caseworker is required to use and check all sources of information). There ensues a long discussion, the crux of which is the notion that any decrease in offending from pre-intake to post intake is a selection artefact because a YJIntake and a high risk score (hence a high rate of offending) are required for entry into the study – in their words, a “spike” in offending rates just before intake. The authors’ argument is that “improvements” are due to regression to the mean (RtM) although they never mention this term.<sup>3</sup> There are difficulties with this view:

- a) according to some of their own data (NIA records, page 191) there is no spike just prior to programme entry
- b) RtM may be responsible for a short term “improvement” between T1 and T2 if these are a few weeks apart, but it is not credible to put long term trends over 0 to 24 months, as detailed

---

<sup>3</sup> RtM refers to the purely statistical tendency for extreme scores or values to move towards the average over time, e.g. very tall or very short individuals will tend to have slightly shorter or taller offspring.

in the MST offending data, down to RtM – if this were true, intervention for delinquency would not be necessary! All offending would tail off by itself due to RtM. On the contrary, it is a fact that offending behaviour is very often chronic and often increases from early to late adolescence

- c) the authors express similar sentiments with regards to the halving of SRD scores in general delinquency; a similar criticism as above in (b) applies.

## **7. RYOP participants older than ideal**

RYOP participants were on average older than participants in the well known American trials. For instance Borduin et al (1995) involved young persons with an average age of 14.8 and the mean age for Henggeler et al. (1997) was 15.2. In the present study the average age was 17.1 at referral (p 86). Increasing age is likely to have a negative impact upon treatment implementation and efficacy as the young person ages away from family association and influence.

## **8. Summary**

The poor outcomes of the RYOP study as detailed in the report by Grace, McClean and Warren should be interpreted cautiously. The overall methodology is non-experimental and is not robust. The validity and reliability of the statistical models used as a control condition is uncertain at best. Psychometric data was collected fitfully and promising offence data collected by MSTNZ was discounted. That aside, there is evidence that the MST intervention was not implemented especially well, nor for an optimum period of time before crucial data was collected. Further, the RYOP participants were significantly older than in previous successful MST studies, with possible deleterious impacts upon the influence of a family-based intervention.

## References

Borduin, C., Mann, B., Cone, L., Henggeler, S., Fucci, B., Blaske, D. & Williams, R. (1995) Multisystemic treatment of serious juvenile offenders: Long-term prevention of criminality and violence. *Journal of Consulting & Clinical Psychology, 4*, 569-578

Henggeler, S., Melton, G. & Smith, L. (1992) Family preservation using multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting & Clinical Psychology, 6*, 953-961

Henggeler, S., Melton, G., Brondino, M., Scherer, D. & Hanley, J. (1997) Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology, 65*, 821-833.

Kjelsberg, E. (2005) Conduct disordered adolescents hospitalised 1963-1990: Secular trends in criminal activity. *European Journal of Child & Adolescent Psychiatry*, 14, 191-199.