

# **Economic Evaluation of Norfolk Arts-Based Interventions with Young Offenders**

## **Phase 1 - Initial Scoping Study**

by

Roger Bowles and Rimawan Pradipto

**Centre for Criminal Justice Economics and Psychology,  
University of York**

**Final draft,  
14 November 06**

**Acknowledgement:** The authors thank Norfolk Youth Offending Team, Connexions Norfolk and the Norfolk Arts Partnership for commissioning this initial scoping study, and for supplying the data on which this report is based. They would also like to thank Sue Massey, Kevin Gosling, Jeni Evans and Fraser Bowe for helpful discussions.

# Contents

Contents .....	2
Executive Summary .....	3
1. Introduction and background .....	6
2. Provision of arts-based activities for young offenders in Norfolk.....	10
2.1 External Programmes .....	10
2.2. In-house mentoring/course provision .....	10
2.3. Recommendations on recording participation in arts programmes .....	11
3. Hypotheses, data sources and methods .....	12
3.1. Hypotheses .....	12
3.2. Data sources .....	13
3.3. Methodology .....	13
3.4. Choice of outcome measures .....	15
4. Findings.....	16
4.1. Intervention Group.....	16
4.2. Non-intervention Group.....	16
4.3. Characteristics of the two groups.....	16
5. Selection of Offenders for Intervention .....	20
5.1 Comparison of means for intervention and non-intervention groups .....	20
5.2 Logistic Regression Analysis of programme referalls.....	21
5.3 Propensity Score Matching (PSM) .....	22
6. Estimating the impact of arts-based intervention.....	24
6.1 Are intervention group participants less likely to be reconvicted?.....	24
6.2. Are intervention group participants reconvicted fewer times or for less serious offences? .....	26
6.3 Is intervention effective?.....	27
6.4. Is there a reduction in the number of convictions? .....	28
7. Cost benefit analysis .....	32
8. Conclusions and scope for further development.....	33
8.1. Scope for an RCT.....	33
8.2. Development of other outcome measures.....	34
Appendices.....	37
Annex 1 Data collection procedure .....	37
Annex 2 Characteristics of the samples for analysis .....	39
Annex 3 Results from binary logistic regression analysis of referral .....	44
Annex 4 Propensity Score Matching Analysis .....	46
Annex 5: Cross Tabulation Analysis .....	50
Annex 6: Logistic Regression model of reconviction likelihood .....	52
Annex 7: Comparison of Means between Intervention and .....	54
Comparison Groups .....	54
Annex 8 ANCOVA analysis.....	55

## **Executive summary**

This report summarizes the initial scoping stage of a proposed three-year research programme. The main objective of the research will be to estimate whether, and by how much, re-offending by children and young people in Norfolk was reduced following their involvement in arts-based programmes on referral from Norfolk Youth Offending Team (YOT). The central hypothesis the research will test is that participation in such arts programmes influences some offenders in ways that reduce their propensity to re-offend and thereby reduce the probability of reconviction for these offenders. The research will attempt to quantify the cost benefits of any reduction in re-offending that can be demonstrated.

The aims of this initial scoping study were to:

- test whether the existing data in Norfolk YOT's YOIS<sup>1</sup> database was already extensive enough to allow statistically-valid conclusions to be drawn;
- make recommendations to ensure that future data recording is compatible with the proposed research.

As we recommend below, the ideal research design for tackling such a challenge is a randomised controlled trial (RCT). This initial scoping study has, necessarily, been conducted retrospectively. As a result our approach is based on the next best alternative, namely 'quasi-experimental' methodology and in particular the use of propensity score matching (PSM).

The intervention group in this scoping study comprises 48 young offenders in the Norfolk YOT who were referred to arts-based programmes over the period 2003-2005. Finding a suitable comparison group, with characteristics closely matched to the intervention group, entailed matching individually the characteristics of each participant with all the rest of the 3,000+ offenders listed in the Norfolk YOIS database who did not participate in arts programmes in the same period.

---

<sup>1</sup> Youth Offending Information System

In terms of the first aim of this scoping study – to test whether the existing data would permit statistically-valid conclusions to be drawn about the impact of arts-based interventions on reoffending – the following conclusions were drawn:

- There is insufficient evidence from the existing data to infer that participation in arts-based interventions reduced the likelihood of reconviction.
- There is not enough evidence from the existing data to support the hypothesis that the programmes reduced the number of offences committed by the participants within 24 months post intervention.
- There is insufficient evidence from the existing data to infer that programme participants committed less serious offences in the post intervention period than non-participants.
- Although a higher proportion of participants were reconvicted, those who were tended to commit fewer offences than non-participants.
- Reconvicted participants, on average, tended to commit less serious crime than non-participants.
- 

The results are consistent with previous studies in this field, which have struggled to provide solid evidence that such programmes were effective in reducing offending behaviour, due to the limitations of the recorded data. In the case of the Norfolk data analysed for this scoping study, these limitations included the following:

- The intervention group is small in size, limiting the scope for exploring variation between offenders within the group.
- The intervention group has an unusually high conviction rate prior to intervention as well as post intervention, consistent with it having been selected from the group of young offenders who are most difficult to work with.
- The referral methods used to select young offenders for the programmes are not entirely clear and this makes it more difficult to choose a comparison group fairly.
- Information about participation in interventions has not been recorded systematically in a way that can readily support monitoring and evaluation activities.

In terms of the second aim of this scoping study – to make recommendations on future data recording – the following measures are recommended:

1. Record arts-based work with young offenders in YOIS, distinguishing between in-house delivery by YOT staff and participation on external programmes.
2. Record the time spent by YOT staff on arts-based work.
3. Identify clearly the objectives of arts-based programmes and the kind of impact on young offenders it is hoped they will have.
4. In light of programme objectives identify clearly the criteria for referral to the programmes and the profile of young offenders thought most likely to benefit.
5. Develop indicators of impact at offender level that are both consistent with the stated objectives of programmes and are also observable following completion of programmes.
6. Design an RCT to explore the effectiveness of arts-based interventions

# 1. Introduction and background

This report summarizes the initial scoping stage of a proposed three-year research programme. The main objective of the research will be to estimate whether, and by how much, re-offending by children and young people in Norfolk was reduced following their involvement in arts-based programmes on referral from Norfolk Youth Offending Team (YOT). The programmes have been running in different centres for a number of years, although the number of participants has been relatively small.

The central hypothesis the research will test is that participation in such arts programmes influences some offenders in ways that reduce their propensity to re-offend and thereby reduce the probability of reconviction for these offenders. The research will attempt to quantify the cost benefits of any reduction in re-offending that can be demonstrated.

The aims of this initial scoping study were to:

- test whether the existing data in Norfolk YOT's YOIS<sup>2</sup> database was already extensive enough to allow statistically-valid conclusions to be drawn;
- make recommendations to ensure that future data recording is compatible with the proposed research.

The study applies state of the art methodology for evaluating the robustness of the existing data from Norfolk. The number of young offenders completing arts-based interventions in Norfolk to date, however, is too small for inferences to be drawn with any statistical confidence. Further work, recorded more robustly, is therefore recommended.

The organisation of the report is as follows. Section 2 summarises the provision of arts-based programmes in Norfolk to date. Section 3 summarises the methodology to be used for the scoping study, the hypotheses to be tested and data availability. Section 4 summarises the findings from analysis of the existing data about the impact of arts-based interventions in Norfolk. Section 5 reviews the strength of selection bias effects in the

---

<sup>2</sup> Youth Offending Information System

intervention group sample. Section 6 summarises findings about whether the existing data can support inferences about impact and Section 7 summarizes the potential for applying cost effectiveness and cost benefit methodologies to future data. Section 8 outlines the main conclusions our recommendations in relation both to data collection and a future research strategy.

Before moving onto the more technical part of the project we review some of the findings from the literature on the impact of arts-based programmes aimed at young offenders and young people at risk of offending. A number of such programmes have been initiated in the UK since 2000. Some of the programmes, such as Splash 2000 and Splash Extra 2002, have targeted youths at risk who live in deprived areas in England and Wales. These programmes aim to divert young people from offending by engaging them in various arts activities during the school summer holiday period, which traditionally has been believed to be a time of increased youth offending. Other programmes, such as Blagg, have targeted specifically young offenders in community settings using drama techniques to challenge offending behaviour.

The Splash programme targeted young people at risk aged 13-17 and was run by the YJB for the first time in summer 2000. In summer 2002, this programme was extended to targeted young people at risk aged 9-17 and was run by YJB and Arts Council England. Arts Council England (2003) reported the effectiveness of Splash in reducing offending behaviour. They claimed that in year 2001 the number of recorded offences in the Basic Command Unit<sup>3</sup> areas where the initiatives had been initiated fell significantly by comparison with BCUs without the initiatives. By contrast Loxley et al (2002) reporting on findings from 6 out of 105 schemes, found that although there was a reduction in the number of offences in summer 2000 compared with the previous year, the reduction was not significant.

The Blagg initiative is more comparable with Norfolk's arts-based interventions than the earlier programmes. Both programmes have targeted young offenders in community settings and aimed, among other things, to reduce offending behaviour. An evaluation of Blagg, in the Central Manchester and Bury YOTs, shows that 30%

---

<sup>3</sup> Basic Command Units are local areas defined for administrative and operational purposes by the police. The Norfolk Police Force Area is split into 3 BCUs, namely Central, Eastern and Western.

of 32 participants were reconvicted within a year of completing the programme. A control group comprising 71 young offenders was selected from other YOTs matched on offences, gender, court orders, and date of intervention. Of this group 39% were reconvicted. This result seems encouraging however the difference between the proportions for the two groups was not sufficiently large, given the comparatively sample size, as to be statistically significant.

It should be noted that the referral systems used in the programmes above were not systematic in the way they would be in a randomised control trial study. This gives rise to types I and II errors<sup>4</sup> in the referral systems, as Loxley et al (2002) reported that only 16% of participants in Splash 2000 were known to be at risk by the YOTs. A similar problem occurred with the evaluation of Blagg. The referral method for participants was not consistent. Most participants (71%) in the Central Manchester group had committed robbery, whereas the Bury participants had committed offences of a more general kind. Attempts have been made to match the characteristics of participants in both intervention and control groups. However, the matching procedures may not represent reliably the actual methods used to refer young offenders to the programme. It would not be surprising, therefore, to find unobserved heterogeneity<sup>5</sup> between young offenders in the two groups which may affect the estimates of the effectiveness of intervention found in the studies.

Hughes (2005) argues that these programmes raised awareness on the part of the young offenders of the criminal justice system and changed attitudes to offending, among other things. In the evaluation of Blagg, for instance, the participants were asked after programme completion to complete a short questionnaire measuring their attitudes to offending (Hughes, 2003). Although the results seem promising, they cannot be used to measure the effectiveness of the programme since participants' attitude had not been measured at the baseline prior to start of the programme and no comparison has been made with changes of attitude on the part of the control group. Ideally, this method can be applied in the future by measuring the attitudes of young offenders in both groups both before and after completion of the programme (or of the

---

<sup>4</sup> In this context this means that some young people who might have expected to be referred were excluded whilst others were included who might have expected to be excluded.

<sup>5</sup> This is a technical term covering the possibility that individuals treated as similar in fact differ systematically on some characteristic about which no information has been recorded.

respective court orders when the intervention has been conducted), using standardised psychometric measures such as BIS, Crime Pics, and PICTS to identify changes in cognitive skills.

Our review of published data from elsewhere in the country thus indicates that the existing Norfolk data is not alone in being problematic as a body of evidence suitable for statistical analysis. If implemented, the further research we recommend in Norfolk would be a useful addition to wider efforts to assess the impact of arts-based interventions in youth justice work.

## 2. Provision of arts-based activities for young offenders in Norfolk

To date, provision of arts-based interventions in Norfolk has been made through two principal channels.

### 2.1 External programmes

Most frequently provision has been in the form of participation in programmes run by external agencies, based at a variety of locations in the county. Children and young people can be referred by YOT case managers to these programmes in the event that a programme is available and it is judged appropriate to the needs of the young offender. The programmes identified, with numbers of participants, are as follows:

Table 1: Participation in arts-based programmes in Norfolk YOT

<b>Programme</b>	<b>Location</b>	<b>Number of participants</b>
<b>Sea Breeze/Sea Change</b>	Great Yarmouth	7
<b>Norwich City College</b>	Norwich	10
<b>Castle</b>	Norwich	23
<b>Broadland</b>	Norwich	2
<b>Arts Centre/Up Front</b>	Kings Lynn	22
<b>Garage</b>	Norwich	23
<b>Total</b>		<b>87</b>

Some of the participants identified in Table 1 have been involved in more than one arts-based programme. We identified a total of 68 different young offenders participating in one or more such interventions.

### 2.2. In-house mentoring/course provision

In addition to externally-provided courses and programmes there are some in-house courses and mentoring activities that have an arts component but are not counted separately as arts-based interventions at present. If arts-based interventions are to be expanded in the future we would recommend that steps be taken to identify the number of YOT and/or County Council staff involved in this type of provision, with case loads, proportion of time etc.

### 2.3. Recommendations on recording participation in arts programmes

It is not possible at present to identify directly from Norfolk's YOIS data system those young offenders who have been involved in arts-based programmes. It is possible to compile such a list from information about participation on individual programmes, but only on an ad hoc basis. We would recommend that a method be developed for recording such participation systematically on YOIS in order to assist future monitoring of the relative effectiveness of arts-based and other interventions.

**Recommendation 1:** Record arts-based work with young offenders in YOIS, distinguishing between in-house delivery by YOT staff and participation on external programmes.

In light of our comments in section 2.2 above we make the further suggestion that:

**Recommendation 2:** Record the time spent by YOT staff on arts-based work.

The criteria for triggering referral to programmes are not very clear at present, at least to us. There is evidence, discussed further below, that referral is not occurring randomly. From informal discussion with Norfolk YOT staff our understanding is that it tends to be young people at greater risk of re-offending who are more likely to be referred to arts-based programmes. Our working hypothesis is that there are certain kinds of deficits, perhaps identified by the profile of (pre-intervention) scores on ASSET, which will tend to prompt referral. But we think it would be helpful to articulate more explicitly both the programme objectives and the referral criteria to be employed in identifying participants. This would help ensure a match between the objectives of the interventions and young offender needs as identified and recorded on the YOIS system and thereby an effective targeting of arts-based interventions.

**Recommendation 3:** Identify clearly the objectives of arts-based programmes and the kind of impact on young offenders it is hoped they will have.

**Recommendation 4:** In light of programme objectives identify clearly the criteria for referral to the programmes.

### **3. Hypotheses, data sources and methods**

The content of arts-based programmes, their objectives and the selection of offenders for referral to them are all matters that lie outside the scope of the present study. For present purposes we assume simply that there is a sub-group of offenders who have been selected for the intervention and that there are other offenders with similar characteristics who have not participated in the intervention primarily because no suitable programme was available at the time. We have also made the assumption, possibly unwarranted, that the various programmes covered share similar objectives and are delivered in a similar way.

#### **3.1. Hypotheses**

The central hypothesis we attempt to test from the existing data is that participation in arts programmes will influence some offenders in ways that reduce their propensity to re-offend and thereby reduce the probability of reconviction for these offenders. This hypothesis follows the standard format used when testing for the effectiveness of crime reduction programmes in general.

We note that the literature on the effectiveness of arts-based interventions contains relatively little evidence to support the hypothesis stated in this rather bald form. Much of the ‘evidence’ is largely anecdotal and robust statistical analysis of the impact of programmes is hard to find<sup>6</sup>. In fairness the designers of arts programmes will often argue that their aim is to encourage the most-difficult-to-engage young offenders to participate actively. Success in achieving such aims should not be under-estimated. But measuring success on such criteria requires that the aims be explicitly stated and that monitoring data be gathered from which achievement can be judged. Empirical analysis of this kind is needed at the programme design stage. The purpose of the present study is to look beyond the achievements of programmes as judged by these ‘internal’ measures and to establish whether there is evidence of a wider impact on reconviction outcomes for young offenders.

---

<sup>6</sup> According to speeches by Rod Morgan and Christopher Frayling at the National Conference ‘The Arts and Youth Justice: Arts and Young People at Risk of Offending’ (Tate Modern, 18 September 2006) high priority is being accorded by the Arts Council and the YJB to the production of such evidence.

### 3.2. Data sources

The principal source of data for the study was the Norfolk YOIS database of offenders. From this database we have taken various fields of data that characterise offenders and their convictions history. The data were taken in anonymised format from the YOIS system but with a case reference number in order to be able to trace any individual back to the database if required. The fields included basic demographic data such as date of birth, gender and ethnicity and offending data including number and gravity of convictions prior to intervention and number and gravity of convictions following intervention.

Data from lists of participants on individual arts-based programmes were matched against the main database so that we could distinguish those young people who had participated in an arts-based intervention and those who had not. The data, including an indicator of whether they had participated, were transferred into SPSS<sup>7</sup> for further analysis.

### 3.3. Methodology

The standard methodology underpinning studies of ‘what works’ in reducing offending by young people is similar to the methodology used in other areas of policy experimentation such as medicine or education<sup>8</sup>. The central question such studies pose is: does the re-offending of those participating in the intervention fall relative to the re-offending of otherwise similar individuals who do not participate? The methodology assumes that the intervention itself has objectives that have been developed from a model of offending and, ideally, of the kinds of offenders believed most likely to benefit from it. The intervention, it is assumed, is delivered in a standardised way and measures for each participant are taken both prior to its beginning and following its completion. From this a measure of the extent of ‘treatment change’ can be derived using a scale that captures benefits based on the project’s objectives<sup>9</sup>.

---

<sup>7</sup> SPSS is a widely-used software package for statistical analysis

<sup>8</sup> The best known source for the approach in the criminal justice field is the study by Sherman and others in the US: Sherman et al (1997, 2002). For discussion in the UK context see a review such as Perry et al (2006).

<sup>9</sup> Note that while the ‘classic’ outcome measure is the reconviction rate there are other measures that might be used to capture intervention objectives including a score on a section of ASSET, OASYS or a psychometric test or a variant on the reconviction measure such as the number of convictions within two years or the time elapsing before a further conviction.

In an ideal world the ideal research design for tackling this question is a randomised controlled trial (RCT)<sup>10</sup>. This identifies the offenders meeting a particular criterion or set of criteria and allocates all such offenders via a blinded, random process to either an ‘intervention group’ or a ‘control group’. The changes in outcome scores for the two groups are compared and an inference drawn as to whether there is a difference between the two. If a difference is found then the study makes an estimate of the ‘effect size’ or ‘treatment effect’. Since offenders have been allocated randomly to the treatment and control groups the possible sources of bias, such as self-selection effects or the ‘cherry picking’ of offenders thought most likely to benefit from intervention, are minimised.

In practice it is rare for interventions in the criminal justice sector to be analysed with this kind of methodology. In the context of arts-based interventions it seems often to be assumed that the benefits from participation are highly subjective and offender-specific. This is taken to imply first that a variety of qualitative measures need to be developed and customised for small client groups and secondly that any effort to aggregate across individuals and programmes is doomed to misrepresent objectives and to miss the point.

Other frequently-cited objections include the ethical issues entailed in withholding intervention from an offender. When there is a presumption that such intervention will be beneficial it may be difficult to allocate some offenders to a ‘no intervention’ alternative. Another common barrier is that practice managers usually wish to prioritise treatment for certain groups. A final inhibition is that tracking experiments via evaluation studies conducted in real time is costly.

Our approach in this initial scoping study is based on the next best alternative namely ‘quasi-experimental’ methodology. This makes use of data gathered from an uncontrolled procedure and seeks to make corrections for selection effects and other sorts of bias. We apply two different techniques that have been developed for such settings. One involves looking at the impact of intervention in a model where the ‘other factors’ that might influence the impact of treatment are allowed for explicitly<sup>11</sup>. The other involves a ‘propensity score matching’ approach in which a comparison group is constructed in a

---

<sup>10</sup> For reviews of methodology by Home Office researchers see Harper & Chitty (2004).

<sup>11</sup> In more technical terms this is known as a ‘covariance’ or ‘ANCOVA’ model. It involves estimating a regression model in which the score after intervention is regressed on the score prior to intervention, a set of offender characteristics such as age and gender and an ‘intervention status’ variable indicating whether the individual participated in a programme: see Brace et al (2003).

way that replicates as closely as possible the characteristics of the intervention group<sup>12</sup>. The changes in outcome for the two groups are then compared to see whether they are in the expected direction and whether they are significantly different.

### 3.4. Choice of outcome measures

The impact of arts-based programmes can be summarised in various ways. The standard measure used for analysing the impact of offender-based criminal justice interventions, and the one used below, is the impact on reconvictions within 24 months of programme or intervention completion. Even this however can be measured in at least three different ways.

The three dimensions covered by the existing YOIS database are:

- whether or not an offender has been convicted in the 24 months following completion of an intervention
- the number of offences for which the offender has been convicted during this 24 month period and
- the gravity of the most serious offence for which the offender was convicted during the 24 month follow-up period.

---

<sup>12</sup> See for example Blundell & Costa Diaz (2000) or Garrett (2006).

## 4. Findings

The first step was to split the database containing details on all young offenders into two sub-groups depending on whether they had participated in an arts-based intervention. We look first at the characteristics of these two groups.

### 4.1. Intervention group

We have data on 68 offenders identified as having been through external programmes with a substantial arts component. The procedure used for compiling this intervention group dataset is set out in Annex 1. After the exclusion of offenders with no offences recorded on the system prior to intervention and of those finishing a programme in 2006 or beyond, a total of 48 offenders remain. This set comprises our '**arts-based intervention group**', hereafter referred to as **ABIG**. As will become clear from the next paragraph, offenders in the intervention group comprise only a very small proportion (1.5%) of all the young offenders dealt with during the period of 2003-2005. From a statistical perspective the key weakness of the intervention group is that it is small in size. A sample of 48 limits the scope for exploring variation between offenders (and programmes) within this intervention group. The characteristics of the intervention group are summarised in Table A2.1 in Annex 2.

### 4.2. Non-intervention group

This dataset covers all offenders completing court orders or interventions during calendar years 2003, 2004 and 2005 recorded on the Norfolk YOIS data system. It excludes offenders in the ABIG and those completing programmes in 2006 for reasons listed above. After these eliminations the final number of observations in the non-intervention or baseline '**comparison group**', hereafter referred to as **COMG**, is 3,188. The characteristics of the intervention group are summarised in Table A2.2 in Annex 2.

### 4.3. Characteristics of the two groups

The first step is to make direct comparisons between the intervention and non-intervention groups in order to establish whether they have similar characteristics or differ from each other. If it can be established that, in effect, the young offenders who have been referred for arts interventions have been selected at random then it is comparatively straightforward to estimate the impact of the intervention. If the referred

group are distinctive in some way, and show signs of having been ‘specially chosen’ for referral, then life becomes more complicated.

Table 2 shows the characteristics and offending behaviour of offenders in both intervention and comparison groups. The majority of offenders in both groups are white and male. The average age at start of intervention of offenders in both groups is similar (around 15 years old), although the proportion of girls in ABIG is higher (33%) than in COMG (21%).

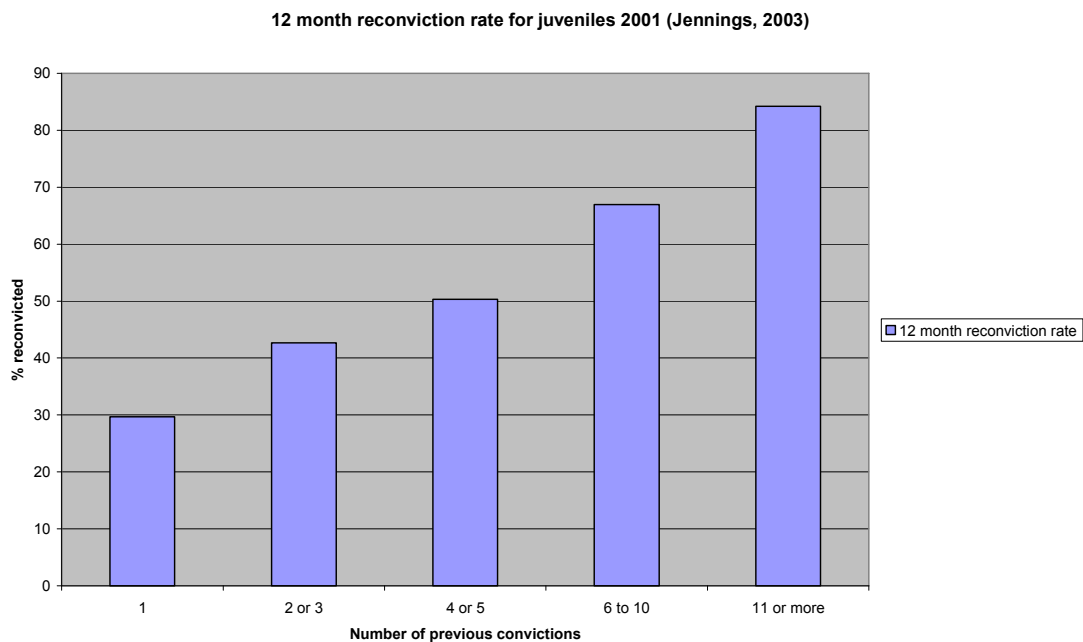
**Table 2: Characteristics of offenders in intervention and comparison groups**

		<b>Arts-based intervention group</b>	<b>Non-ABI offender group</b>
Number of offenders		<b>48</b>	<b>3,188</b>
% white		<b>94%</b>	<b>96%</b>
% female		<b>33%</b>	<b>21%</b>
Average age at start of intervention		<b>15.25 yrs</b>	<b>15.31 yrs</b>
Average age at the end of intervention		<b>15.75 yrs</b>	<b>15.62 yrs</b>
Average number of offences 12 months prior	Mean	<b>5.40</b>	<b>3.76</b>
	Standard deviation	4.48	4.21
Average maximum gravity score of offences 12 months prior	Mean	<b>4.29</b>	<b>3.67</b>
	Standard deviation	1.37	1.32
Average ASSET score (at end of intervention)	Mean	<b>18.00</b>	<b>13.64</b>
	Standard deviation	7.57	9.66
% of offenders reconvicted within 24 months		<b>65%</b>	<b>51%</b>
Average number of offences 24 months post (annualised rate)	Mean	<b>1.34</b>	<b>1.28</b>
	Standard deviation	2.11	2.40
Average maximum gravity score of offences 24 months post	Mean	<b>2.33</b>	<b>1.89</b>
	Standard deviation	2.02	2.08

Offenders in the intervention group, on average, tend to have committed more offences within the 12 months prior to intervention (5.40 offences per offender) than their counterparts in the comparison group (3.76 offences per offender). Evidence from studies of juvenile reconvictions, such as Jennings (2003), suggests that the reconviction rate is significantly higher for those offenders who have committed 6 to 10 offences previously (>65%) than for those who have 4 or 5 previous convictions (about 50%). A similar order of difference characterises the Norfolk data (Table 2) and the national picture from Jennings (2003) summarised in Fig. 1.

The average maximum offence gravity is also greater for offenders in the intervention group (average gravity score of 4.29) than for those in the comparison group (average gravity score of 3.67)<sup>13</sup>.

Figure 1: National data on juvenile reconviction rates



As a first approximation it would be possible to estimate the ‘impact’ of arts-based interventions by looking at the change in the mean number of convictions following a programme for those who participated in the programme relative to those who did not participate. This ‘difference in differences’ approach subtracts the mean number of offences pre-intervention from the mean number of offences post-intervention. It produces the finding reported in Table 3. This shows that the conviction rate (mean number of convictions per annum) falls for both the intervention and comparison groups but that the fall is greater for the arts-based intervention group than the comparison group. The size of this difference gives one possible measure of the impact of the arts interventions.

<sup>13</sup> Offence gravity is ranked on the ACPO scale which ranges from 1 – 8, running from the least to the most serious offence category.

Table 3: Difference in differences approach

	Mean number of offences per annum		
	Pre-intervention	Post-intervention	Change in mean
Intervention group	5.40	1.34	-4.06
Comparison group	3.76	1.28	-2.48
Difference in mean change between the two groups			-1.58

Unfortunately this finding (that the offending rate falls by 1.58 percentage points further for the ABIG group than for COMG) cannot be taken at face value. The greater number and gravity of offences committed by offenders in ABIG as compared with COMG prior to intervention suggests that arts interventions are being systematically targeted at a more difficult sub-group of young offenders. The implication of this is that there may be a ‘selection bias’ in the intervention group and that it cannot be safely inferred at this stage that there is a greater fall in the conviction rate for the intervention group. There is also the possibility that there may be ‘regression to the mean’ effects contaminating the findings. The initially higher conviction rates for the intervention group leave more scope for reduction than those of the non-intervention group so part of any reduction shown by the intervention group may be attributable to a spontaneous reduction that would have occurred even without intervention. We take up these matters in section 5 below.

We note in passing from Table 2 that the possibility of the intervention group being different in character from the non-intervention group is consistent also with the observation that the average ASSET score post-intervention is considerably higher for the intervention group than for the comparison group. However this score is only taken post-intervention and would only really be of much value if its values were known pre-intervention as well as post-intervention, since it is differences in the score prior to intervention that are critical in establishing how the groups differ before any arts-based work begins. In addition, the total ASSET score itself is not regarded in any event as a legitimate measure of the degree to which a young person represents a re-offending risk.

## 5. Selection of offenders for intervention

The comparison of offender characteristics as between the ABIG and COMG indicates that the groups differ particularly in relation to the frequency and gravity of offending prior to intervention. A closer look at the distribution of the number of offences enables a clearer sense of the difference in the structure of the offending risk across the two groups to be established.

From the histograms for the two groups presented as Figs A2.1 and A2.2 in Annex 2 it can be seen that there is a high proportion (more than 50%) of the ABIG who had been convicted of 3 or fewer offences during the 12 months prior to intervention. Likewise a significant minority of the COMG offenders had committed 10 or more offences but had not been referred for arts-based interventions. So although, on average, the ABIG group represent a greater reconviction risk they are by no means recruited solely from the group most at risk. They are neither typical of Norfolk young offenders nor are they pathological.

### 5.1 Comparison of means for intervention and non-intervention groups

The first step is to establish more formally whether the differences between the two groups are sufficiently large to rule out the possibility that the groups are quite similar really and that the apparent differences in characteristics are the result of chance.

From the analysis of the sample means presented in Table A2.3 in Annex 2 we derive the following inferences:

- Young offenders referred for arts-based interventions differ significantly<sup>14</sup> from other young offenders in terms of their offending profile, having typically:
  - a greater number of previous convictions
  - a history of convictions of greater gravity.
- There is no significant difference in the average age of the two groups
- There is some sign that there is a higher propensity to refer females for intervention than males, but the difference is not significant at the 95% level.

---

<sup>14</sup> The number of convictions and the offence gravity scores are both greater at the 99% significance level.

These findings taken together are significant and imply that the characteristics of the offenders participating in arts programmes are different from those of non-participants. The obvious question that this raises is how we might make allowance for the ‘selection bias’ in estimating the impact of these programmes.

## 5.2 Logistic regression analysis of programme referalls

In order to explore more thoroughly which variables give the best indication of the likelihood of referral onto arts-based programmes we fitted a binary logistic regression model as described in Annex 3. The findings from this model are as follows:

- Age and ethnicity play no discernible role
- The number of offences prior to intervention has a weakly positive influence
- Gender and the maximum gravity of offences prior to intervention both play a significant role, with girls and those committing more serious offences having a very significantly higher probability of being referred.

We note that this model gives a slightly different picture of the key differences between the intervention and comparison groups than the one that emerged from Table 2. This occurs because we are now allowing for several variables to play a role simultaneously. Correlations between the ‘explanatory variables’ are such that some become less significant when used in combination with others than they are when looked at in isolation. Further investigation shows that the model accounts for between 0.5% and 3.8% of the variance of the referral process. This is very low, and indicates that the model does not adequately fit the data (see Annex 3).

This finding that the characteristics of the ABIG differ from those of COMG can help reconcile some apparently conflicting findings from Table 2 above that we have not as yet commented on. Within 24 months post intervention, 65% of offenders in the intervention group had been reconvicted, while only 51% of offenders in the comparison group had been. In addition, the average number of annualised re-convictions post intervention for offenders in the intervention and the comparison groups is 1.34 and 1.28 respectively. On the face of it this might seem to suggest that arts-based interventions are making things worse and only increasing the re-offending risk. But such

an inference cannot be drawn because, as we have already established, the offending and other characteristics of participants in the two groups differ significantly.

There are various ways of dealing with selection effects of this kind. A popular method is to ignore them and to estimate the impact of policy on the assumption that such effects can be safely ignored. For reasons outlined at length elsewhere (Blundell & Costa Dias, 2000 and Bryson et al, 2002) this can produce highly misleading results.

A widely-canvassed approach, but one that could not be used for the present initial scoping study, is to avoid selection bias effects by using an experimental approach (as outlined in section 3 above) and assigning offenders randomly between an arts-based intervention and a 'normal treatment' pathway. We return to this research design in section 8.

### **5.3 Propensity score matching (PSM)**

The next best, quasi-experimental, approach is 'propensity score matching' (or PSM) that identifies a (smaller) comparison group matched as closely as possible to the intervention group. The outcomes for this group may represent a more reliable benchmark against which to measure the impact of intervention: Garrett (2006). We consider three quite closely-related methods of generating such a comparison group. These methods are summarised in Annex 4. The characteristics of the samples generated by the three methods are summarised in Table 4.

Table 4 Characteristics of comparison group samples generated by PSM

	Arts-based intervention group	Comparison groups			
		All non-ABI	PSM-1	PSM-2	PSM-3
Number of offenders	48	3,188	581	540	48
% female	33	21	21	22	33
Average age at start	15.25	15.31	14.96	14.94	15.31
Mean number of offences prior	5.40	3.76	2.50	2.61	5.40
Mean maximum gravity score prior	4.29	3.67	3.48	3.48	4.29
Average ASSET score at end	18.00	13.64	11.46	11.90	18.06
% reconvicted within 24 months	65	51	50	51	50
Mean number of offences post	1.34	1.28	0.94	0.96	1.70
Mean maximum gravity score post	2.33	1.89	1.74	1.78	2.00

The first two columns of data document the characteristics of the intervention group and the group of all non-intervention offenders as outlined above. The remaining three columns refer to two samples derived by propensity score matching models (PSM-1 and PSM-2) and a third sample (PSM-3) derived by random selection from sub-groups of non-intervention offenders matched with intervention group offenders.

No very clear picture emerges from comparisons between the intervention group and these four somewhat different comparison samples. The two samples constructed using propensity score matching techniques proper (PSM-1 and PSM-2) do not seem to be a good match for the intervention group. The principal reason for this is likely to be that the characteristics of young offenders that incline YOT workers to refer them for arts-based interventions are not picked up by the variables used in the logistic regression model: see Annex 3 for further discussion.

At a purely impressionistic level it seems that it is not the volume or seriousness of past offending that prompts referral so much as an offender being difficult to engage in

conventional education and training. If this is the case then any model relying on gender and offending history as explanatory variables will predict referral rather poorly. Improvement would require the inclusion of more qualitative data such as scores on particular sub-scales in ASSET that come closer to capturing the relevant characteristics. The other thing likely to be inhibiting the predictive power of the PSM models is that the intervention group is very small relative to the non-intervention group and this makes it more difficult for the model to discriminate between the factors contributing to referral. One of the more notable characteristics of the final comparison group (PSM-3) is that as well as matching the age and offending history characteristics of the intervention group closely, it is the one comparison group with a greater mean number of convictions post intervention. The proportion of the intervention group reconvicted is higher at 65% compared with 50% but the mean number of offences for the intervention group is lower at 1.34 than for PSM-3 for which it is 1.70.

## 6. Estimating the impact of arts-based intervention

The critical issue being addressed in this scoping study is whether there is evidence from the existing data that offending by the arts-based intervention group was lower following participation in an arts programme than it would have been had they not participated. It is easy to characterise, but more difficult to operationalize, a measure of intervention impact based on such a counterfactual.

### 6.1 Are intervention group participants less likely to be reconvicted?

One approach to estimating impact is to compare the proportions of offenders in the intervention and comparison groups committing offences within 24 months post intervention. This can be done via cross tabulation as in Table 5.

Table 5: Cross tabulation of reconviction outcomes: intervention group versus all non-intervention

		Comparison or intervention group?		Total
		Comparison group	Intervention group	
Whether re-offended 24 months post intervention	No	1,571	17	1,588
	Yes	1,617	31	1,648

<b>Total</b>	3,188	48	3,236
--------------	-------	----	-------

A chi-square analysis based on the data in Table 4 indicates that there is some difference between the proportion of offenders in each sample who are reconvicted, but the difference is not quite significant at the 5% level:  $\chi^2 = 3.636$ ,  $df = 1$ ,  $p = 0.057$ : [see Table A5.1 in Annex 5]. The result indicates a positive correlation between participation in the intervention group and reconviction:  $\rho = 0.034$ ,  $p = .057$ , two tailed [see Table A5.2 in Annex 5]. Of 3,188 offenders in the non-intervention group, 1,617 (50.63%) were reconvicted within 24 months post intervention. By contrast, of 48 offenders in the intervention group, 31 (64.58%) were reconvicted within the same interval.

Caution should be taken in interpreting this result as implying that participation in an arts-based intervention increases the likelihood of offending. The continuity correction or Yates's corrected chi-square<sup>15</sup> is weakly significant (i.e.  $\chi^2 = 3.102$ ,  $d.f. = 1$ ,  $p = 0.078$ ) implying that the sample may not be a good representation of the Norfolk young offender population. This is consistent with the 'selection bias' finding reported above.

A similar analysis was done by making comparisons between the intervention group and the three samples generated by the PSM procedures outlined in the previous section. The findings are all broadly similar, namely that there is a slight difference between the reconviction rates but that also there are differences in the characteristics of the intervention group and the various comparison groups. Tables A5.3 and A5.4 illustrate these findings for the comparison between the intervention group and PSM-3.

Further evidence that participation in the programme appears not to influence significantly the probability of reconviction can be found by estimating a logistic model of the likelihood of reconviction using the data on the intervention and non-intervention groups combined. From both models documented in Annex 6 it can be seen that the 'status' variable, which indicates whether or not an offender belonged to the intervention group, is insignificant. This suggests that involvement in the intervention group has limited marginal explanatory power and that the likelihood of being reconvicted is much more sensitive to other factors such as gender and number of previous convictions. The

---

<sup>15</sup> This is a statistical correction used in cases for either the sample is too small or the sample is not a good approximation to total population.

results, however, should be treated with caution since further tests show that neither model fits adequately with the data: see Annex 6.

## 6.2. Are intervention group participants reconvicted fewer times or for less serious offences?

Although offenders in the intervention group have a higher likelihood of reconviction this need not imply that they committed more offences on average than their counterparts in the comparison groups. We look next at the change in the number of offences pre- and post-intervention for both groups.

Table 6: Comparison of changes in the average number of offences and in offence gravity (ABIG vs COMG)

**Group statistics**

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. deviation</b>	<b>Std. error mean</b>
<b>Change in the number of offences (annualised 24 months post - 12 months pre)</b>	<b>Intervention group</b>	48	-4.052	4.649	0.671
	<b>Comparison group</b>	3,188	-2.484	4.201	0.074
<b>Change in offence gravity</b>	<b>Intervention group</b>	48	-1.958	2.143	0.309
	<b>Comparison group</b>	3,188	-1.785	2.267	0.040

Table 6 shows that offenders in both groups tend to commit fewer offences in the period post-intervention relative to the period pre-intervention. Nevertheless, on average, offenders in the intervention group show a greater reduction (1.34 offences) in the number of convictions than do their counterparts in the non-intervention group (1.2 offences). As can be seen from Table A7.2 in Annex 7 the impact on reconviction rates is statistically significant at the 95% level ( $t = -2.323$ ,  $df = 48.163$ ,  $p = 0.024$ ).

Offenders in both groups tend to be convicted for offences of lower gravity post-intervention relative to pre-intervention. The reduction of offence gravity is greater in the intervention group (1.96) than in the comparison group (1.79). Nevertheless, as shown in Table A7.2 in Annex 7, this difference is not statistically significant ( $t = -0.527$ ,  $df = 3,234$ ,  $p = 0.598$ ).

Of course we have argued above that selection bias in the existing data is such that these comparisons are not really valid. For purposes of completeness we look at whether the findings apply when the comparisons are based on samples generated by other matching methods. In Table 7 below it can be seen that the fall in the number of reconvictions is on average greater for the intervention group than for the sample PSM-3, the one that most closely matches the intervention group. But the reduction in offence gravity is not as great for the intervention group as for PSM-3.

Table 7: Comparison of the average changes in number of reconvictions and offence Gravity: Intervention group (ABIG) vs Comparison group PSM-3

	<b>Group</b>	<b>N</b>	<b>Mean</b>	<b>Std. deviation</b>	<b>Std. error mean</b>
<b>Change in the number of offences (annualised 24 months post - 12 months pre)</b>	ABIG	48	-4.052	4.649	0.671
	COMG PSM-3	48	-3.698	4.513	0.651
<b>Change in offence gravity</b>	ABIG	48	-1.958	2.143	0.309
	COMG PSM-3	48	-2.292	2.601	0.375

The results reported in Table A7.2 in Annex 7 demonstrate that there is no significant difference in the changes in either the number of convictions following intervention or the gravity of offences as between the intervention group and the PSM-3 group.

### 6.3 Is intervention effective?

An alternative approach is to look directly at the question of whether intervention has a significant effect on reconvictions. This involves estimating a 'covariance model' (or ANCOVA). In this analysis, the number of offences post intervention (or the offence

gravity post intervention) serves as dependent variable. The number of offences pre intervention (or offence gravity pre intervention) and ‘status’<sup>16</sup>, on the other hand, serve as explanatory variables. The novelty of the analysis is that the effectiveness of an intervention can be ascertained and at the same time the regression toward the mean effect can be minimised: see Vickers and Altman (2001) and Bland (2004).

As previously the analysis can be run as a comparison between the intervention group and the various samples created using matching methods. Table A8.1 in Annex 8 displays the results of applying this model to the large comparison sample COMG. The ‘status’ variable indicates whether an offender belonged to the intervention group. This variable is insignificant, indicating that we cannot find an ‘intervention effect’ that is statistically significant. A similar result is found in relation to offence gravity, namely that participation in a programme does not reduce offence gravity post intervention: see Table A.8.2 in Annex 8. As the set of results reported as Tables A8.3 and A.8.4 in Annex 8 indicates, we get a similar finding when making the comparison with the matched group PSM-3 for both number of offences and offence gravity post intervention.

#### 6.4. Is there a reduction in the number of convictions?

It is possible that, while the proportion of offenders reconvicted remains little changed, the number of offences committed by recidivists nevertheless falls. This possibility can be tested separately. The obvious way of checking it is to measure the total number of convictions committed by the intervention and comparison groups during the 12 month windows either side of the intervention interval. From Table 8 it can be seen that the number of convictions is lower post intervention for both groups, but that the proportionate reduction is greater for the intervention group than for the comparison group.

---

<sup>16</sup> The status variable is a binary indicator of whether the person participated in an arts-based intervention.

Table 8 Estimate of the crime reduction impact of arts-based programmes

	Intervention group	Comparison group
Number of offenders	48	3,188
Total offences 12 months prior intervention	259	11,990
Total offences 24 months post intervention (annualised)	65	4,072
% Change	-74.90%	-66.04%
Offences saved (in percentage points)		<b>8.86%</b>
Short term impact (number of offences saved within 12 months post intervention) (A)		<b>23</b>
Offence multiplier from the age of 18 - 40 years old (based on Offender Index cohort 1958) ^ (B)		<b>2.92</b>
Long term impact (estimated number of offences saved when offenders aged 18-40 years old) (C) = (A) *(B-1)		<b>44</b>
Total Impact (D) = (A) + (C)		<b>67</b>

^ see YJB (2005) and Bowles and Pradiptyo (2005)

We conducted further investigation of reconvictions, number of offences and offence gravity committed post intervention for each arts-based programme in Norfolk YOT. Although young offenders in the intervention group are more likely to be reconvicted those who were reconvicted had committed fewer offences (down from 5.40 pre to 4.16 post) than those in comparison groups (up from 3.76 pre to 5.04 post for COMG and from 5.40 pre to 6.79 post for COMG PSM-3): see Table 9. The results also suggest that only reconvicted participants show a reduction in the average number of offences. Apart from convicted participants in Castle Art, all reconvicted participants in other programmes showed a substantial reduction in the number of offences committed. This result suggests high variation in offending behaviour among participants between the programmes. Nevertheless the sample size within each individual intervention programme is small and so the result should be treated with great caution.

Table 9: Comparison of offending behaviour by arts-based programmes

Programmes		Pre-intervention			Post-intervention		
		Total offences	No of offenders	Average offences per offender	Total offences	Number of offenders who were reconvicted	Average offences/convicted offender
ABIG	Sea Change	48	6	8	20	3	6.67
	Kings Lynn	54	12	4.5	12	6	2.00
	The Garage	92	14	6.57	46	12	3.83
	City College	29	6	4.83	22	5	4.40
	Castle Art	34	9	3.78	28	4	7.00
	Broadland	2	1	2	1	1	1.00
<b>Total ABIG</b>		<b>259</b>	<b>48</b>	<b>5.40</b>	<b>129</b>	<b>31</b>	<b>4.16</b>
<b>Total COMG PSM-3</b>		<b>259</b>	<b>48</b>	<b>5.40</b>	<b>163</b>	<b>24</b>	<b>6.79</b>
<b>Total COMG</b>		<b>11,990</b>	<b>3,188</b>	<b>3.76</b>	<b>8143</b>	<b>1,617</b>	<b>5.04</b>

A similar result applies also for maximum offence gravity. On average the reconvicted offenders in the intervention group tend to commit less serious crime (average gravity down from 4.29 pre to 3.61 post) than those in the comparison groups (down from 3.76 pre to 3.72 post for COMG and from 4.29 pre to 4.00 for COMG PSM-3): see Table 10. Since the sample in the intervention group is relatively small, this result should be treated with caution.

Table 10: Comparison of offending behaviour by art-based programmes

		Pre intervention			Post Intervention		
		Total offence gravity	No. offenders who were reconvicted	Average offence gravity/ convicted offender	Total offence gravity	No. offenders who were reconvicted	Average offence gravity/ convicted offender
Art-based interventions	Sea Change	27	6	4.50	12	3	4.00
	Kings Lynn	49	12	4.08	23	6	3.83
	The Garage	68	14	4.86	45	12	3.75
	City College	23	6	3.83	18	5	3.60
	Castle Art	36	9	4.00	11	4	2.75
	Broadland	3	1	3.00	3	1	3.00
<b>Total ABIG</b>		206	48	4.29	112	31	3.61
<b>Total COMG PSM-3</b>		206	48	4.29	96	24	4.00
<b>Total COMG</b>		11,704	3188	3.67	6014	1,617	3.72

## **7. Cost benefit analysis**

The question of how evidence about effectiveness translates into an analysis of the rate of return on resources devoted to arts-based initiatives is, in essence, a comparatively straightforward matter. The methodology for expressing the value of crime reduction benefits in financial terms is well developed (Dhiri & Brand, 1999) and can rely on valuations derived from increasingly sophisticated research into the costs of crime: Brand & Price (2000); Duborg et al (2005).

At this stage, however, the scope for applying cost effectiveness or cost benefit analyses to the existing data from Norfolk – or, indeed, elsewhere - is very limited because of the poor quality of the evidence we have been able to assemble about effectiveness. There are various complications that would have to be overcome before the return on Norfolk arts-based interventions could be estimated with any confidence from the existing data. The various programmes we have generically listed as arts-based may differ significantly from one another in terms of both their impact on re-offending and also in their cost per participant and so may need to be investigated separately. Another concern would be that some of the resources allocated to arts programmes, such as their supervision by YOT workers, are common to other programmes. In addition the arts-based interventions are not implemented in isolation, so steps might be needed to separate out the effects of arts-based work from the effects attributable to other programmes running alongside.

A more far-reaching concern would be that the objectives of the programmes may extend beyond reducing reconvictions over a short interval of one or two years to delivering longer-term benefits in terms of improved lifetime opportunities for young offenders, better employment prospects and so on. Capturing and monetising such benefits would require more by way of longitudinal study and could not produce readily useable findings for some years.

## **8. Conclusions and scope for further development**

The principal conclusion we draw is that the existing Norfolk data does not provide statistically robust evidence that arts-based interventions have been effective in reducing offending. However the tests we have been able to construct are limited in critical ways. Important issues thus remain unresolved and the scope for, and the hunger for, assembling an evidence base on the effectiveness of arts interventions remains great – nationally as well as in Norfolk.

### **8.1. Scope for an RCT**

One way of improving the evidence base would be to design an RCT capable of discerning treatment effects if they are present. As a prelude to such a study the objectives of arts-based interventions would have to be clearly articulated, as would the criteria for referring young offenders to them. Once this had been achieved the flow of offenders to an arts intervention and to an alternative intervention (or placebo or non-intervention) could be randomised and offenders followed up for one or two years following the programme to explore their reconvictions<sup>17</sup>. Our view is that the practical barriers to such a project could be overcome. The failure of the scoping study to establish any clear presumption of the effectiveness of arts-based programmes, albeit on limited evidence as noted, weakens the strength of the ethical argument against withholding treatment from some young offenders.

Because of minimum sample size requirements and the limited number of arts programmes available it might be prudent for such a trial to be spread across a number of YOYS. This would reduce the length of time taken to recruit enough youngsters into the trial although it might be troublesome to coordinate. There is also the question of how similar the interventions would need to be.

Given the paucity of existing evidence, and given also the political commitment to assembling credible scientific evidence, as to programme effectiveness we see an RCT as the most promising line for future research.

---

<sup>17</sup> Of course it would be possible to use an outcome measure other than reconvictions but this would make it less easy to compare the effectiveness of arts based work with other types of intervention.

**Recommendation 5:** Design an RCT to explore the effectiveness of arts-based interventions

## 8.2. Development of other outcome measures

There are many grounds for arguing that reconviction rates are very limited as a means of capturing the impact of arts-based programmes. If such programmes are to be more widely implemented then there are strong grounds for seeking better means of characterising their effects. Just as we have argued that there is scope for being more explicit about referral criteria for selecting offenders for arts-based programmes so we contend that there is room to develop tools for monitoring the impact of programmes by building on existing data collection tools and data. Having identified deficits or whatever as referral criteria the obvious step is to use tools such as scores on sub-scales of ASSET (or scores on psychometric tests designed to measure various kinds of cognitive deficits) as a measure of the impact that a programme has had. These criteria might be programme-specific and they may be qualitative rather than quantitative. The key principles to apply in their development, in our view, would be:

- that they capture some component of behaviour with relevance to offending and
- that they involve a minimum of any new data collection.

**Recommendation 6:** Develop indicators of impact at offender level that are both consistent with the stated objectives of programmes and are also observable following completion of programmes.

## References

- Arts Council England (2003) *Addressing Youth Offending: Art Council England's Contribution to the Splash Extra Programme 2002* (Executive Summary), Arts Council England.
- Bland M (2004) 'Regression towards the mean or why was Terminator III such a disappointment?' *mimeo*, Department of Health Sciences, University of York, <http://www-users.york.ac.uk/~mb55/talks/regmean.htm>.
- Blundell R & Costa Dias M (2000) 'Evaluation methods for non-experimental data', *Fiscal Studies*, 21(4)
- Bowles, R and R Pradiptyo. (2005) *Young Adults in the Criminal Justice System: Cost and Benefit Considerations*, Report to Barrow Cadbury Foundation, Centre for Criminal Justice Economics and Psychology, University of York.
- Brace N, Kemp R & Snelgar R (2003) *SPSS for Psychologists: A Guide to Data Analysis using SPSS for Windows*, Second Edition, Basingstoke: Palgrave Macmillan
- Brand, S. and Price, R. (2000), *Economic and Social Costs of Crime*. London: Home Office
- Bryson, A., Dorsett, R. and Purdon, S. (2002) *The use of propensity score matching in the evaluation of active labour market policies*, Department for Work and Pensions Working Paper No.4. Available online at: <http://www.dwp.gov.uk/asd/asd5/wp-index.html>.
- Burtless G (1995) The Case for Randomized Field Trials in Economic Policy and Research, *Journal of Economic Perspectives*, 9(2), 63-84
- Dhiri S & Brand S (1999) *Analysis of Costs and Benefits: guidance for evaluators, Crime Reduction Programme – Guidance Note 1*, London, Home Office
- Dubourg R, Hamed J & Thorns J (2005) *The economic and social costs of crime against individuals and households 2003/04*, Home Office Online Report 30/05, London, Home Office
- Garret B (2006) *Propensity Score Methods* at [www.urban.org/toolkit/data-methods/propensity.cfm](http://www.urban.org/toolkit/data-methods/propensity.cfm)
- Gold M (1996), *Cost-Effectiveness in Health and Medicine*, New York, Oxford University Press.
- Harper G & Chitty C (2004) "The impact of corrections on re-offending", *Home Office Research Study 291*, London, Home Office, December 2004
- Hughes, J (2003) 'The Impact of Blagg on Challenging and Reducing Offending by Young People; an Evaluation of a Drama-Based Offending Behaviour Workshop', *mimeo*, Centre for Applied Theatre Research, University of Manchester.
- Hughes, J (2005) 'Doing the Arts Justice; A Review of Research Literature, Practice and Theory', *mimeo*, Centre for Applied Theatre Research, University of Manchester.

Jennings D (2003) One-year juvenile reconviction rates: first quarter of 2001 cohort, *Home Office Online Report* 18/03, London, Home Office

Loxley, C, L Curtin and R Brown (2002) 'Summer Splash Schemes 2000: Findings from Six Case Studies', *Crime Reduction Research Series Paper* 12. London: Home Office.

Sherman, L. W., Farrington, D.P.; Welsh. B.C.; MacKenzie, D.L. (2002). *Evidence-Based Crime Prevention*, Routledge.

Sherman, L. W. G., D.C.; MacKenzie, D.L. ; Eck, J.; Reuter, P.; Bushway, S.D.; (1997). *Preventing Crime: What Works, What Doesn't, What's Promising*. Office of the Justice Programs Research Report, Washington, DC: US Department of Justice: 1 - 356.

Torgerson C J & Torgerson D J (2003) 'Randomised Controlled Trials in Education: Lessons from Health'. *Oxford Review of Education*: 2003;29:67-80.

Vickers A j & Altman D G (2001) 'Analysing controlled trials with baseline and follow up measurements', *British Medical Journal*, 323: 1123-1124.

Welsh, B. C. and D. P. Farrington (2001). "Toward an Evidence-Based Approach to Preventing Crime." *Annals of the American Academy of Political and Social Science* **578**(0): 158-73.

Youth Justice Board (2005) Monitoring and Evaluating the Safer School Partnership Programme, London: YJB. On-line version available at:

<http://www.youth-justice-board.gov.uk/Publications/Scripts/prodView.asp?idProduct=269&eP=YJB>

# Appendices

## Annex 1 Data collection procedure

The procedure used for compiling the intervention group dataset was as follows:

- a list of external arts-based activities was compiled following conversations with Jeni Evans (Norfolk YOT) and others;
- Fraser Bowe (Norfolk YOT) searched the database for programme participants by searches based on a combination of {programme name & art};
- the resulting list of names was used to compile for each offender (identified by case number) the following information:
  - (a) Offences committed (number of offences and ACPO gravity measure for the most serious offence) during the 12 months prior to the beginning of the Court Order or intervention of which the arts-based work was a component,
  - (b) Offender demographics (searched via ‘interventions’) – date of birth, gender, ethnicity, plus further information from the ASSET completed at the end of the intervention,
  - (c) Offences committed (number of offences and ACPO gravity measure for the most serious offence number of previous offences, number of reconvictions including gravity) during the 24 months following completion of the intervention<sup>18</sup>
- Information on the number of offences prior to the intervention is patchy. Cases have been eliminated where the number of offences prior to the invention is equal to 0.
- The impact of the programme on the participant’s offending behaviour has been estimated only for those who finished the programme prior to 1 January 2006. This ensures that a period of at least 6 months has elapsed since completion of the intervention. Those scheduled to finish a programme in 2006 or beyond are not included in the analysis.

---

<sup>18</sup> Various adjustments have to be made here, since programme completion in some instances has not happened yet or has occurred only recently. For purposes of calculating the post-intervention offending rate we excluded cases where completion occurred no more recently than 6 months ago. In order to get an appropriate comparison with the pre-intervention offending rate the 2 year data were converted to an annual rate. In those cases where completion had occurred within the last 24 months but more than 6 months ago the number of offences was converted to an annual rate equivalent.

- The dataset covers all offenders completing court orders or interventions during calendar years 2003, 2004 and 2005. It excludes 2006 completions on the grounds that there was insufficient time since programme completion for any reliable information about reconvictions amongst this group to have accumulated.

The procedure used<sup>19</sup> for compiling the comparison group dataset was as follows:

- Offenders active on the system were identified for each of the financial years 2003, 2004 & 2005, and a separate data file was created for each year. The total number of cases for the comparison group was 4,519. Observations which satisfy the following criteria were also eliminated:
  - the end date of intervention is in year 2006 or beyond, or
  - the number of offences prior to intervention is equal to 0, or
  - it duplicates a case in the intervention group.

After these eliminations the final number of observations in the comparison group is 3,188.
- Data were collected for a similar set of variables on each offender as in the case of the intervention group, namely:
  - offender demographics including ASSET score at the end of the intervention,
  - number of offences prior to intervention (including gravity)
  - Number of reconvictions in the two years following (including gravity).

---

<sup>19</sup> The successful method was reached following a couple of unsuccessful experiments. The very large overall size of the YOIS data is such that extracting subsets of data stretches processor and memory powers with the result that it is both time-consuming and potentially non-feasible. A first extraction method ran for over an hour before failing. A second method took a further hour before failing. The data were finally generated by Fraser using a year-by-year search.

## Annex 2 Characteristics of the samples for analysis

**Table A2.1 Descriptive statistics for the arts-based intervention group (ABIG)**

**Descriptive statistics**

	N	Minimum	Maximum	Mean	Std. deviation
Ethnicity	48	1	3	1.10	.425
Gender	48	.00	1.00	.6667	.47639
Post Asset score link to court order/intervention	48	5.00	34.00	18.0000	7.56813
Number of offences pre intervention	48	1.00	17.00	5.3958	4.48040
Number of offences post intervention	48	.00	20.00	2.6875	4.21828
Offence gravity pre Intervention	48	2.00	6.00	4.2917	1.36769
Offence gravity post Intervention	48	.00	6.00	2.3333	2.02467
Whether reoffended 24 months post intervention	48	.00	1.00	.6458	.48332
Age at start of intervention	48	12.00	17.00	15.2500	1.42172
Number of offences 12 months post intervention	48	.00	10.00	1.3438	2.10914
Change in the number of offences 24 (24 months post - 12 months pre)	48	-15.00	14.00	-2.7083	5.66046
Change in the number of offences 12 (12 months post - 12 months pre)	48	-16.00	4.00	-4.0521	4.64854
Change in offence gravity	48	-6.00	3.00	-1.9583	2.14336
Valid N (listwise)	48				

**Table A2.2 Descriptive statistics for the comparison group (COMG)**

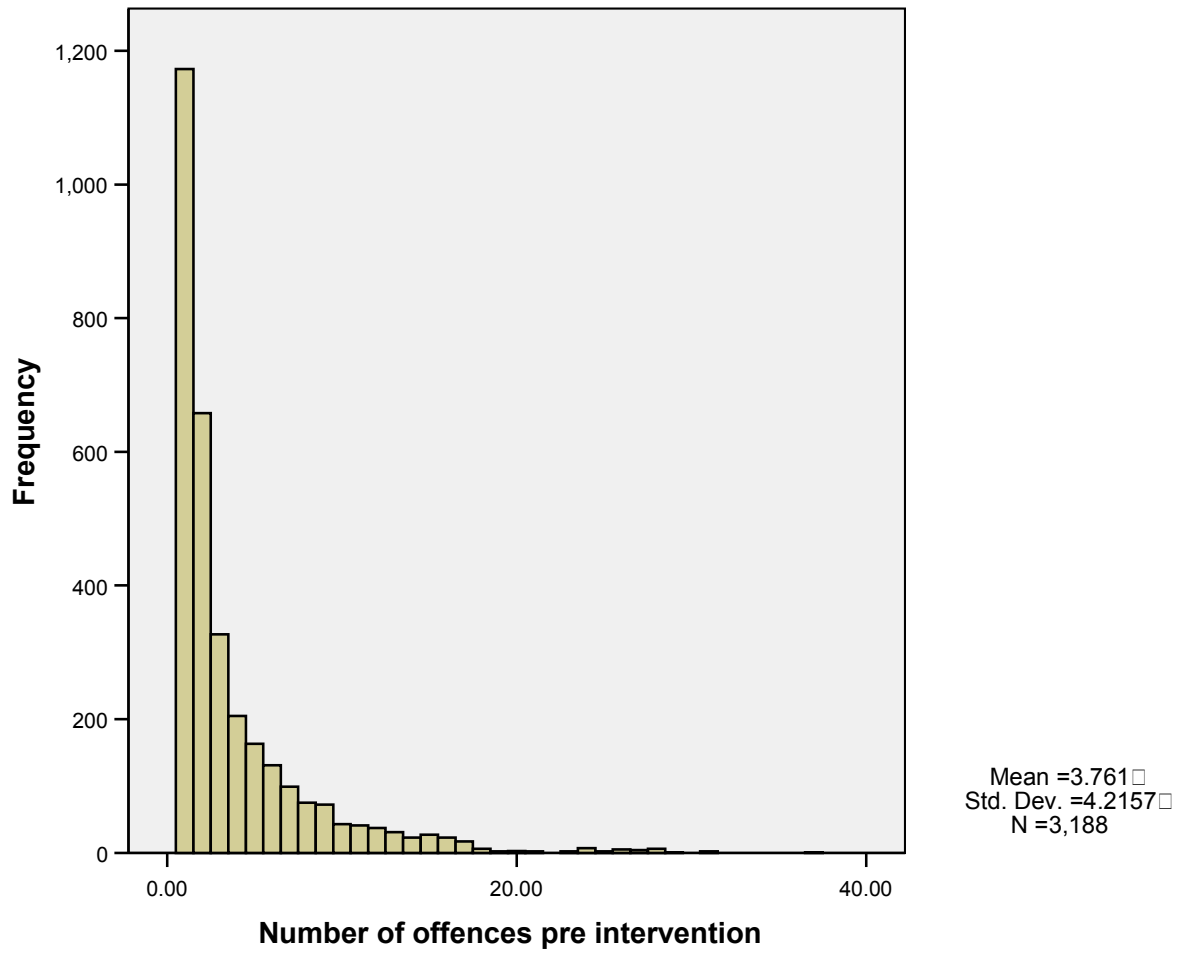
**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
Ethnicity	3188	1	8	1.14	.728
Gender	3188	.00	1.00	.7889	.40816
Post Asset score link to court order/intervention	3188	.00	48.00	13.6402	9.65509
Number of offences pre intervention	3188	1.00	37.00	3.7610	4.21570
Number of offences post intervention	3188	.00	50.00	2.5543	4.80670
Offence Gravity pre Intervention	3188	.00	8.00	3.6713	1.31505
Offence Gravity post Intervention	3188	.00	7.00	1.8864	2.08367
Whether reoffended 24 months post intervention	3188	.00	1.00	.5072	.50003
Age at start of intervention	3188	10.00	45.00	15.3093	2.03214
Number of offences 12 months post intervention	3188	.00	25.00	1.2771	2.40335
Change in the number of offences 24 (24 months post - 12 months pre)	3188	-31.00	40.00	-1.2067	5.39220
Change in the number of offences 12 (12 months post - 12 months pre)	3188	-31.00	15.50	-2.4838	4.20094
Change in offence gravity	3188	-8.00	6.00	-1.7848	2.26724
Valid N (listwise)	3188				

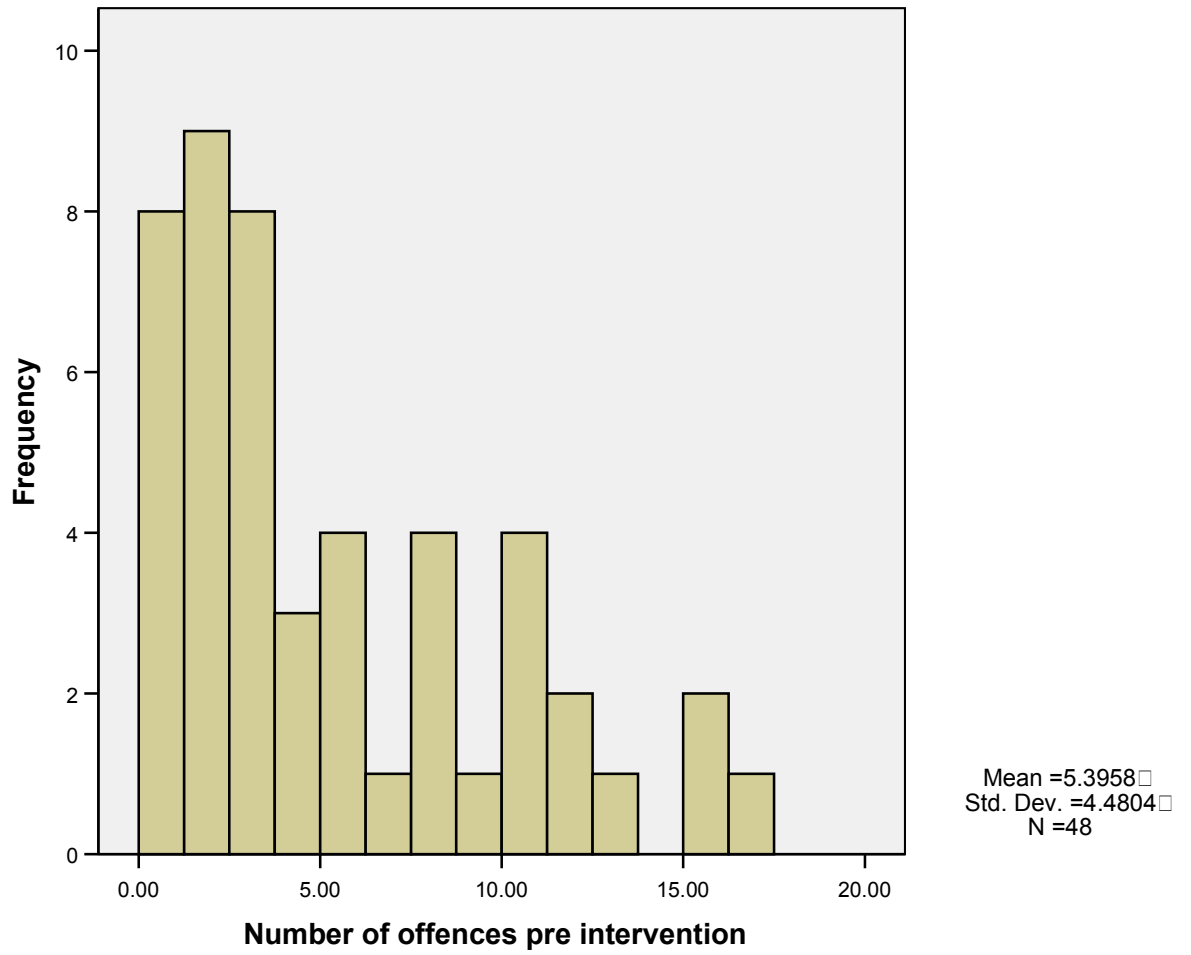
**Table A2.3 Comparison of characteristics of the intervention and comparison groups (ABIG vs COMG)**

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Number of offences pre intervention	Equal variances assumed	3.532	.060	-2.664	3234	.008	-1.63485	.61362	-2.83799	-.43172
	Equal variances not assumed			-2.511	48.261	.015	-1.63485	.65099	-2.94357	-.32614
Offence Gravity pre Intervention	Equal variances assumed	2.527	.112	-3.242	3234	.001	-.62040	.19135	-.99558	-.24522
	Equal variances not assumed			-3.121	48.317	.003	-.62040	.19878	-1.02000	-.22080
Gender	Equal variances assumed	10.630	.001	2.054	3234	.040	.12223	.05951	.00555	.23891
	Equal variances not assumed			1.768	48.045	.083	.12223	.06914	-.01678	.26124
Age at start of intervention	Equal variances assumed	.858	.354	.201	3234	.840	.05928	.29442	-.51798	.63655
	Equal variances not assumed			.285	49.935	.777	.05928	.20834	-.35919	.47776

**Fig. A2.1**      **Distribution of offences prior to intervention: comparison group (COMG)**



**Fig. A2.4**      **Distribution of offences prior to intervention:**  
**Arts-based Intervention Group (ABIG)**



## Annex 3 Results from binary logistic regression analysis of referral

**Notes:** The model is fitted using the Binary Logistic Regression Analysis procedure in SPSS. It is run on a dataset formed by adding together the ABIG and COMG observations and using whether the offender belonged to ABIG as the dependent variable. The covariates used as ‘explanatory variables’ and their significance levels are listed below. Note that we have omitted variables measured after intervention, particularly the ASSET score. The latter seems to be quite different for the two groups: it has a mean of 18.00 for ABIG compared with mean of 13.64 for COMG. It could therefore quite plausibly be expected to play a role. But since this is the measure **following intervention**, it cannot be used as part of a model to predict whether or not the offender would be referred in the first place, since it is information that is not available at the time the referral decision was made.

The criterion used to judge whether the covariates that are included are a significant part of the explanation is whether the value in the Sig column is less than 0.05 (for 5% significance level) and less than 0.01 (for 1% significance level). The two key variables meeting this latter criterion are grav\_pre (i.e. maximum offence gravity 12 months pre intervention) and gender.

**Table A3.1 Selection model estimates**

		Variables in the Equation					
Step		B	S.E.	Wald	df	Sig.	Exp(B)
1 <sup>a</sup>	ethnicit	-.098	.245	.161	1	.688	.906
	gender	-.765	.315	5.876	1	.015	.466
	pre_offe	.045	.028	2.508	1	.113	1.046
	grav_pre	.301	.115	6.840	1	.009	1.352
	age_start	-.052	.092	.319	1	.572	.949
	Constant	-4.134	1.449	8.140	1	.004	.016

a. Variable(s) entered on step 1: ethnicit, gender, pre\_offe, grav\_pre, age\_start.

A useful by-product that can be generated when estimating this model is a set of predictions of the likelihood with which each offender might have been referred for intervention. This set of predictions could be used as the basis for generating a new ‘comparison group’ with which to compare the ABIG. This would entail estimating the

predicted probability of ABIG offenders' referral via the 'predicted values probabilities' option as part of the logistic regression analysis. The COMG offender with the closest match to this probability would then be selected as a member of the new comparison group, which we will call PSMG since the technique is known as Propensity Score Matching<sup>20</sup>. This procedure would need to be repeated 48 times so as to compile 48 entries for PSMG. It would then be possible to make comparisons between ABIG and PSMG since the latter group would correspond much more closely to the intervention group than the original 'comparison' group COMG. We have refrained from applying this technique as yet since the number of offenders in the intervention group is small (48) and the model is not particularly successful in predicting which offenders will be referred to ABIG.

Table A3.2. Model Summary and Hosmer and Lemeshow Test of Logistic Regression

**Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	482.095(a)	.005	.038

a Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

**Hosmer and Lemeshow Test**

Step	Chi-square	df	Sig.
1	17.832	8	.023

The Cox and Snell and Nagelkerke R-Squares are the approximation of the goodness of fit for these models. The model above accounts for between 0.5% and 3.8% of the variance for young offenders being referred to arts-based interventions. These values are considered very low. Further investigation of the model using the Hosmer and Lemeshow test shows that we can reject the hypothesis that the model adequately fits the data, indicating that our model is not good enough to predict the likelihood of participation in the programmes.

<sup>20</sup> For further details of the Propensity Score Matching techniques see the description at the website of the Urban Institute authored by Garrett (2006).

## Annex 4 Propensity Score Matching Analysis

Three methods were used in efforts to produce a sample of non-intervention offenders that was comparable with the intervention group. We outline each method in turn and the characteristics of the sample it produced.

### **Method 1:**

The first step is to explore the factors (if any) associated with selection for treatment. A logistic regression analysis was conducted to predict the likelihood of offenders being included in the intervention group, as discussed in Annex 3. A by-product of this analysis is a prediction for each offender of the probability of their being referred to intervention, given their personal characteristics and given what is known about the selection criteria employed. A new comparison group (hereafter referred to as COMG PSM-1) has been constructed by selecting (all) offenders from the original comparison group who have a predicted probability of referral for arts-based intervention identical to that of an offender in the intervention group.

This method has limitations. There are cases in which, although there is a group of offenders sharing the same predicted probability of intervention as the relevant member of the intervention group, none of them has similar characteristics (e.g. gender, number of offences and maximum offence gravity pre-interventions) to those of the individual in the intervention group for whom matches are being sought.

In addition it is possible that for each offender in the intervention group there may be several offenders in the comparison group with identical probabilities of referral for intervention. Indeed what we find is that the comparison group based generated by this first method, COMG PSM-1, consists of 581 observations making it substantially larger than ABIG. Descriptive statistics for this COMG PSM-1 sample are given in Table A4.1. A summary is given also in Table 3 in the main text.

This procedure produces a comparison sample that fails to resemble the intervention group more closely than the original comparison group comprising all non intervention group offenders. In particular the mean values for the two key offending history variables (number and gravity of offences prior to intervention) lie further from the means for the intervention group sample than do the corresponding means for the whole

group of non-intervention offenders. This kind of finding is common in PSM analysis in cases where the logistic regression model has only limited predictive power and where the sample size is comparatively small.

**Table A4.1: Descriptive Statistics for COMG PSM-1**

	N	Minimum	Maximum	Mean	Std. Deviation
Ethnicity	581	1	4	1.02	.226
Gender	581	.00	1.00	.7935	.40517
Post Asset score link to court order/intervention	581	.00	40.00	11.4613	8.60036
Number of offences pre intervention	581	1.00	17.00	2.5026	2.60545
Number of offences post intervention	581	.00	28.00	1.8881	3.39059
Offence Gravity pre Intervention	581	2.00	6.00	3.5164	.96051
Offence Gravity post Intervention	581	.00	6.00	1.7401	1.94482
Whether reoffended 24 months post intervention	581	.00	1.00	.5043	.50041
Age at start of intervention	581	12.00	18.00	14.9570	1.20482
Number of offences 12 months post intervention	581	.00	14.00	.9441	1.69529
Change in the number of offences 24 (24 months post - 12 months pre)	581	-16.00	20.00	-.6145	3.88579
Change in the number of offences 12 (12 months post - 12 months pre)	581	-16.50	7.50	-1.5585	2.84075
Change in offence gravity	581	-6.00	3.00	-1.7762	2.10292

**Method 2:**

An alternative to method 1 was tried in order to avoid the limitations of COMP PSM-1. A second comparison group (COMG PSM-2 hereafter) was selected from offenders in the original comparison group by choosing all non-intervention offenders with a predicted probability of belonging to the intervention group which was identical to that of an intervention group offender. In addition to this, further selection has been conducted by matching gender, number of offences pre-intervention and also maximum offence gravity pre-intervention. As with the previous method, for each offender in the ABIG there are in many cases several offenders in COMP PSM-2 who share identical values for the predicted probability, gender, number of convictions and offence gravity prior to intervention. The implication of this is that the number of observations in COMP PSM-2 remains substantially larger than that in ABIG, although its total of 540

observations is slightly smaller than for the previous method. The characteristics of the comparison group and its descriptive statistics can be seen in Table A4.2.

**Table A4.2: Descriptive Statistics for COMG PSM-2**

	N	Minimum	Maximum	Mean	Std. Deviation
Ethnicity	540	1	3	1.01	.122
Gender	540	.00	1.00	.7778	.41613
Post Asset score link to court order/intervention	540	.00	40.00	11.8963	8.71803
Number of offences pre intervention	540	1.00	17.00	2.6111	2.67203
Number of offences post intervention	540	.00	28.00	1.9148	3.44748
Offence Gravity pre Intervention	540	2.00	6.00	3.4759	.98663
Offence Gravity post Intervention	540	.00	6.00	1.7759	1.95206
Whether re-offended 24 months post intervention	540	.00	1.00	.5130	.50030
Age at start of intervention	540	12.00	18.00	14.9407	1.24249
Age at the end of intervention	540	12.00	18.00	15.2148	1.33948
Number of offences 12 months post intervention	540	0	14	0.9574074	1.7237404
Change in the number of offences 24 (24 months post - 12 months pre)	540	-16.00	20.00	-.6963	3.93316
Change in offence gravity	540	-6.00	3.00	-1.7000	2.10106

### Method 3

Methods 1 and 2 both suffer from the shortcoming that they produce a ‘comparison group’ that does not appear to be a good match for the intervention group in relation to measures such as the number (or gravity) of convictions prior to intervention. The means of the average number of offences and offence gravity pre-intervention in COMG PSM-2 are substantially lower than those for COMG. This is inevitable since the minimum and maximum values of both variables in COMG PSM-2 are matched perfectly with those of ABIG, and the observations in COMG PSM-2 are substantially larger than those of in ABIG. Similar to offenders in COMG PSM-2, the offenders in ABIG had committed between 1 and 17 offences during the 12 months prior to intervention. By contrast the offenders in COMG had committed between 1 and 37 offences prior to intervention. The offence gravity in both ABIG and COMG PSM-2 lies

in the range 0 to 6. Offenders in COMG, however, had committed offences with gravity in the range 0 to 8.

This third method is not strictly a PSM method since it abandons the use of the predicted probabilities of intervention derived from a logistic regression model. Instead it uses a simple method designed to ensure that the means of the covariates (gender, number and gravity of convictions pre-intervention) match closely between the intervention and comparison groups. For each intervention group offender we identify all non-intervention offenders with the same gender and number and gravity of previous convictions. We then select an offender at random from this latter group to join the new comparison group PSM-3. The result is a sample of the same size as the intervention group (48), with the same gender balance and with similar means for both the number and gravity of previous convictions. In a more sophisticated variant it would be possible to generate a succession of such samples using the same procedure to improve the statistical properties of the comparison group.

**Table A4.3: Descriptive Statistics for COMG PSM-2**

	N	Minimum	Maximum	Sum	Mean	Std. Deviation
<b>Ethnicity</b>	48	1	3	53	1.10	0.43
<b>Gender</b>	48	0	1	32	0.67	0.48
<b>Number of offences pre intervention</b>	48	1	17	259	5.40	4.48
<b>Offence Gravity pre Intervention</b>	48	2	6	206	4.29	1.37
<b>Number of offences post intervention</b>	48	0	28	163	3.40	5.83
<b>Offence Gravity post Intervention</b>	48	0	6	96	2	2.25
<b>Whether re-offended 24 months post intervention</b>	48	0	1	24	0.5	0.51
<b>Age at start of intervention</b>	48	12	17	735	15.31	1.39
<b>Age at the end of intervention</b>	48	12	18	748	15.58	1.43
<b>Length of Intervention (weeks)</b>	48	0	54	811	16.90	13.72
<b>Number of offences 12 months post intervention</b>	48	0	14	81.5	1.70	2.92
<b>Change in the number of offences (annualised 24 months post - 12 months pre)</b>	48	-16	6	-177.5	-3.70	4.51
<b>Change in offence gravity</b>	48	-6	3	-110	-2.29	2.60

## Annex 5: Cross Tabulation Analysis

Table A5.1: Chi-Square Analysis of Reconviction Behaviour:  
intervention group (ABIG) versus all non-intervention (COMG)

### Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
<b>Pearson Chi-Square</b>	3.636(b)	1	0.057		
<b>Continuity Correction(a)</b>	3.102	1	0.078		
<b>Likelihood Ratio</b>	3.695	1	0.055		
<b>Fisher's Exact Test</b>				0.06	0.038
<b>Linear-by-Linear Association</b>	3.635	1	0.057		
<b>N of Valid Cases</b>	3236				
a Computed only for a 2x2 table					
b 0 cells (.0%) have expected count less than 5. The minimum expected count is 23.56.					

Table A5.2: Correlation Analysis of Reconviction  
intervention group (ABIG) versus all non-intervention (COMG)

### Symmetric Measures

		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
<b>Interval by Interval</b>	<b>Pearson's R</b>	0.034	0.017	1.907	.057(c)
<b>Ordinal by Ordinal</b>	<b>Spearman Correlation</b>	0.034	0.017	1.907	.057(c)
<b>N of Valid Cases</b>		3236			
a Not assuming the null hypothesis.					
b Using the asymptotic standard error assuming the null hypothesis.					
c Based on normal approximation.					

Table A5.3: Chi-Square Analysis of Reconviction Behaviour:  
intervention group (ABIG) versus COMG PSM-3

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
<b>Pearson Chi-Square</b>	2.086(b)	1	0.149		
<b>Continuity Correction(a)</b>	1.533	1	0.216		
<b>Likelihood Ratio</b>	2.094	1	0.148		
<b>Fisher's Exact Test</b>				0.216	0.108
<b>Linear-by-Linear Association</b>	2.064	1	0.151		
<b>N of Valid Cases</b>	96				
a Computed only for a 2x2 table					
b 0 cells (.0%) have expected count less than 5. The minimum expected count is 20.50.					

Table A5.4: Correlation analysis of Reconviction Behaviour:  
intervention group (ABIG) versus COMG PSM-3

Symmetric Measures					
		Value	Asymp. Std. Error(a)	Approx. T(b)	Approx. Sig.
<b>Interval by Interval</b>	<b>Pearson's R</b>	0.147	0.101	1.445	.152(c)
<b>Ordinal by Ordinal</b>	<b>Spearman Correlation</b>	0.147	0.101	1.445	.152(c)
<b>N of Valid Cases</b>		96			
a Not assuming the null hypothesis.					
b Using the asymptotic standard error assuming the null hypothesis.					
c Based on normal approximation.					

## Annex 6: Logistic Regression model of reconviction likelihood

The model estimates the likelihood of young offenders being reconvicted within 24 months post intervention. A binary choice variable indicating whether a young offender was reconvicted within 24 months post intervention serves as dependent variable. It takes a value 1 if an offender was convicted within 24 months post interventions, and 0 otherwise. The covariates used as explanatory variables are: gender, number of offences within 12 months pre intervention, maximum offence gravity pre intervention, age at the start of intervention, the length of court order, and status (i.e. 1 if an offender is in intervention group or 0 otherwise). The logistic regression model has been estimated for two data sets, namely ABIC vs COMG and ABIC vs COMG PSM-3. The results are as follows:

### Model 1: The Likelihood of Reconviction (ABIC vs COMG)

#### Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	gender(1)	-.295	.090	10.748	1	.001	.745
	pre_offe	.112	.012	93.691	1	.000	1.119
	grav_pre	.044	.031	1.995	1	.158	1.045
	age_start	-.200	.024	70.797	1	.000	.819
	length_intv	.014	.003	29.426	1	.000	1.014
	<b>status(1)</b>	<b>-.233</b>	<b>.319</b>	<b>.535</b>	<b>1</b>	<b>.465</b>	<b>.792</b>
	Constant	2.601	.489	28.297	1	.000	13.477

a Variable(s) entered on step 1: gender, pre\_offe, grav\_pre, age\_start, length\_intv, status.

#### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4219.043(a)	.079	.105

a Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

#### Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	44.045	8	.000

## Model 2: The Likelihood of Reconviction (ABIG vs COMG PSM-3)

### Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	gender(1)	-.607	.507	1.434	1	.231	.545
	pre_offe	.185	.069	7.264	1	.007	1.203
	grav_pre	.126	.194	.418	1	.518	1.134
	age_start	-.468	.191	5.985	1	.014	.626
	length_intv	.016	.013	1.525	1	.217	1.016
	<b>status(1)</b>	<b>-.526</b>	<b>.485</b>	<b>1.174</b>	<b>1</b>	<b>.279</b>	<b>.591</b>
	Constant	6.120	2.813	4.733	1	.030	454.734

a Variable(s) entered on step 1: gender, pre\_offe, grav\_pre, age\_start, length\_intv, status.

### Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	109.225(a)	.203	.273

a Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	17.042	8	.030

In both models, the results show that there is not sufficient evidence to suggest that involvement in an arts-based intervention reduced the likelihood of reconviction. In the first model, which uses all non-participants as a comparison group, the model explains only between 7.9% and 10.5% of the variance of the reconviction. In the second model, which uses COMG PSM-3 as a comparison group, the R squared values have increased to between 20.3% and 27.3% (see Cox & Snell and Nagelkerke R Squares). This shows that the second model gives a better explanation of the variance of the reconviction. Nevertheless, neither model adequately represents the data according to the Hosmer and Lemeshow Test.

## Annex 7: Comparison of Means between Intervention and Comparison Groups

**Table A7.1: t-Test of Independence between Intervention and Non-Intervention Groups (ABIG vs COMG)**

Variables		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Change in the number of offences (annualised 24 months post - 12 months pre)	*	4.473	0.035	-2.563	3234	0.010	-1.568	0.612	-2.768	-0.368
	**			<b>-2.323</b>	<b>48.163</b>	<b>0.024</b>	<b>-1.568</b>	<b>0.675</b>	<b>-2.925</b>	<b>-0.211</b>
Change in offence gravity	*	0.79	0.374	-0.527	3234	0.598	-0.174	0.329	-0.819	0.472
	**			-0.556	48.597	0.581	-0.174	0.312	-0.801	0.454

\*) Equal variances assumed

\*\*) Equal variances not assumed

**Table A7.2: t-Test of Independence between Intervention and Comparison Groups (ABIG vs COMG PSM-3)**

		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Change in the number of offences 12 (12 months post - 12 months pre)	^	0.522	0.472	-0.379	94	0.706	-0.354	0.935	-2.211	1.503
	^^			-0.379	93.918	0.706	-0.354	0.935	-2.211	1.503
Change in offence gravity	^	1.729	0.192	0.685	94	0.495	0.333	0.486	-0.633	1.299
	^^			0.685	90.689	0.495	0.333	0.486	-0.633	1.300

\*) Equal variances assumed

\*\*) Equal variances not assumed

## Annex 8 ANCOVA analysis

A8.1 Number of offences annualised 24 months post interventions (ABIG vs COMG)

A8.1.a Test for homogeneity in covariance (ABIG vs COMG)

### Tests of Between-Subjects Effects

Dependent Variable: Number of offences 12 months post intervention

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1565.933(a)	3	521.978	98.936	.000
Intercept	49.295	1	49.295	9.343	0.002
Status	1.715	1	1.715	0.325	0.569
pre_offe	52.819	1	52.819	10.011	0.002
status * pre_offe	<b>8.101</b>	<b>1</b>	<b>8.101</b>	<b>1.535</b>	<b>0.215</b>
Error	17051.758	3232	5.276		
Total	23904	3236			
Corrected Total	18617.691	3235			

a R Squared = .084 (Adjusted R Squared = .083)

**A.8.1.b Final ANCOVA for annualised 24 months post interventions (ABIG vs COMG)**

### Tests of Between-Subjects Effects

Dependent Variable: Number of offences 12 months post intervention

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1557.832(a)	2	778.916	147.612	.000
Intercept	54.99	1	54.99	10.421	0.001
pre_offe	1557.622	1	1557.622	295.184	.000
status	<b>1.93</b>	<b>1</b>	<b>1.93</b>	<b>0.366</b>	<b>0.545</b>
Error	17059.859	3233	5.277		
Total	23904	3236			
Corrected Total	18617.691	3235			

a R Squared = .084 (Adjusted R Squared = .083)

## A8.2 Offence Gravity (ABIG vs COMG)

### A8.2.a Test of homogeneity of covariance (ABIG vs COMG)

#### Tests of Between-Subjects Effects

Dependent Variable: Offence Gravity post Intervention

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	420.372(a)	3	140.124	33.254	0
Intercept	11.731	1	11.731	2.784	0.095
status	0.089	1	0.089	0.021	0.885
grav_pre	35.097	1	35.097	8.329	0.004
status * grav_pre	<b>0.843</b>	<b>1</b>	<b>0.843</b>	<b>0.2</b>	<b>0.655</b>
Error	13618.633	3232	4.214		
Total	25636	3236			
Corrected Total	14039.005	3235			

a R Squared = .030 (Adjusted R Squared = .029)

### A8.2.b Final ANCOVA for offence gravity (ABIG vs COMG)

#### Tests of Between-Subjects Effects

Dependent Variable: Offence Gravity post Intervention

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	419.529(a)	2	209.764	49.794	0
Intercept	131.301	1	131.301	31.168	0
grav_pre	410.085	1	410.085	97.346	0
status	<b>3.669</b>	<b>1</b>	<b>3.669</b>	<b>0.871</b>	<b>0.351</b>
Error	13619.476	3233	4.213		
Total	25636	3236			
Corrected Total	14039.005	3235			

a R Squared = .030 (Adjusted R Squared = .029)

### A8.3 Number of Offences annualised 24 months post intervention (ABIG vs COMG PSM-3)

#### A8.3.a Test of homogeneity of covariance (ABIG vs COMG PSM-3)

##### Tests of Between-Subjects Effects

Dependent Variable: Number of offences 12 months post intervention

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	47.353(a)	3	15.784	2.574	0.059
Intercept	23.172	1	23.172	3.779	0.055
status	1.23	1	1.23	0.201	0.655
pre_offe	36.159	1	36.159	5.897	0.017
status * pre_offe	<b>8.184</b>	<b>1</b>	<b>8.184</b>	<b>1.335</b>	<b>0.251</b>
Error	564.105	92	6.132		
Total	833.5	96			
Corrected Total	611.458	95			

a R Squared = .077 (Adjusted R Squared = .047)

#### A8.3.b Final ANCOVA for annualised 24 months post intervention (ABIG vs COMG PSM-3)

##### Tests of Between-Subjects Effects

Dependent Variable: Number of offences 12 months post intervention

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	39.169(a)	2	19.585	3.183	0.046
Intercept	23.172	1	23.172	3.766	0.055
pre_offe	36.159	1	36.159	5.876	0.017
status	<b>3.01</b>	<b>1</b>	<b>3.01</b>	<b>0.489</b>	<b>0.486</b>
Error	572.289	93	6.154		
Total	833.5	96			
Corrected Total	611.458	95			

a R Squared = .064 (Adjusted R Squared = .044)

#### A8.4 Offence Gravity (ABIG vs COMG PSM-3)

##### A8.4.a Test of homogeneity of covariance (ABIG vs COMG PSM-3)

###### Tests of Between-Subjects Effects

Dependent Variable: Offence Gravity post Intervention

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	14.740(a)	3	4.913	1.08	0.362
Intercept	14.223	1	14.223	3.126	0.08
status	2.392	1	2.392	0.526	0.47
status * grav_pre	<b>4.566</b>	<b>1</b>	<b>4.566</b>	<b>1.003</b>	<b>0.319</b>
grav_pre	7.508	1	7.508	1.65	0.202
Error	418.593	92	4.55		
Total	884	96			
Corrected Total	433.333	95			

a R Squared = .034 (Adjusted R Squared = .003)

##### A8.4.b Final Ancova for offence gravity post intervention (ABIG vs COMG PSM-3)

###### Tests of Between-Subjects Effects

Dependent Variable: Offence Gravity post Intervention

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	10.174(a)	2	5.087	1.118	0.331
Intercept	14.223	1	14.223	3.126	0.08
grav_pre	7.508	1	7.508	1.65	0.202
status	<b>2.667</b>	<b>1</b>	<b>2.667</b>	<b>0.586</b>	<b>0.446</b>
Error	423.159	93	4.55		
Total	884	96			
Corrected Total	433.333	95			

a R Squared = .023 (Adjusted R Squared = .002)