Evaluation of the Citizenship Evidence-based Probation Supervision Programme using a

Stepped Wedge Cluster Randomised Controlled Trial

Abstract

This study evaluated an RNR evidence-based offender supervision programme, Citizenship, using a randomised controlled trial (RCT). Citizenship has a cognitive-behavioural basis and focuses on education, increasing motivation to change, and community integration. The RCT was a stepped wedge cluster randomised design which has rarely been used in criminal justice and overcomes some ethical objections to RCT implementation. Participants were all medium- and high-risk offenders commencing probation supervision (N=1,091) in any one of six office units during the one year roll-out of the programme. Overall there was a non-significant 20% effect of Citizenship in reducing reconvictions. However, controlling for risk the hazard with higher risk offenders was 34% lower than for the control group. Results therefore support RNR based probation supervision.

Keywords   Cognitive-behavioural · Community corrections · Risk-need-responsivity · Stepped wedge cluster randomised controlled trial · Structured probation supervision

Much has been learned in the last two decades about what is effective in reducing reoffending. Meta-analyses of large numbers of offender intervention studies have highlighted the key elements associated with achieving the desired outcome (Andrews, Zinger et al., 1990; Antonowicz & Ross, 1994; Lipsey, 1992). This evidence has demonstrated that when programmes are cognitive behavioural and are implemented with integrity, they can significantly reduce reconvictions (Andrews, Bonta, & Hoge, 1990; Lipsey, Landenberger, & Wilson, 2007;

McGuire, 1995). Specifically, community group-work programmes have been found to reduce recidivism by an average of 35% when they are designed and delivered according to Risk-Need-Responsivity (RNR) principles (Andrews & Bonta, 2010; Andrews, Zinger et al., 1990). RNR requires that the intensity of the intervention matches the level of risk of reoffending ('Risk' principle), the content targets the offender's criminogenic needs ('Need' principle), and the delivery style is suited to the learning capabilities of the individual offender ('Responsivity' principle).

Although the effectiveness of applying RNR principles to offending behaviour groupwork programmes is recognised, it is only in recent years that attempts have been made to apply them to individualised probation and parole supervision (Bonta et al., 2011; Paparozzi & Gendreau, 2005; Pearson, McDougall, Kanaan, Bowles, & Torgerson, 2011; Taxman, 2008; Taxman, & Thanner, 2006; Thanner & Taxman, 2003). This paper describes the evaluation of a structured probation supervision programme, Citizenship (Bruce & Hollin, 2009), based on evidence-based RNR principles. The evaluation applied a stepped wedge cluster randomised controlled trial (RCT) research design (Hussey & Hughes, 2007), rarely used in criminal justice research, to evaluate a relatively new approach to probation supervision of offenders.

Background

Little is known about the effectiveness of traditional probation supervision (Bonta, Rugge, Scott, Bourgon, & Yessine, 2008) in part due to a lack of definition of what is meant by the term (Taxman, 2002), and lack of confidence in studies claiming positive outcomes for probation where research designs are weak (McDougall, Perry, & Farrington, 2006). Taxman (2002) concluded that, from the range of different approaches adopted, probation supervision appeared to be atheoretical. McDougall et al. (2006) noted that generally community supervision was allocated to offenders with a lower risk of reoffending, and this factor often contributed to an illusion of positive outcomes. Where statistical adjustments were made to control for differences

in risk between community and custodial sentences, variations in reconviction rates were shown to be reduced. The lack of strong evidence for the effectiveness of traditional probation was supported in the review by Bonta et al. (2008) of 15 probation studies. This found a low level of reduction in reoffending associated with community supervision (two percentage points) and no decrease in violent recidivism.

More recently studies have examined whether the application of RNR principles to community supervision could improve its effectiveness, whilst at the same time identifying the key environmental factors needed to support this approach. Paparozzi and Gendreau (2005) found that RNR principles, applied in a balanced law enforcement/social casework approach to supervision with high-risk offenders, resulted in a 10% to 30% reduction in recidivism. Latessa and Lowenkamp (2006) found that the application of the same principles to community supervision contributed to greater effectiveness with high-risk offenders, who benefited from the link between more intensive supervision and treatment. Bonta et al. (2011) in a RCT, also demonstrated that offenders supervised by probation officers trained in RNR had a 40% reduction in recidivism, compared to a 25% reduction among offenders supervised by non-RNR trained probation officers. Therefore a pro-social environment, characterised by links between intensive supervision and treatment, and including training of staff in RNR, has been found to be a contributory component in reducing re-offending.

These illustrations of improved impact on recidivism resulted in the development of purpose-designed RNR supervision programmes such as Proactive Community Supervision (PCS) (Taxman, 2008), and Citizenship (Bruce & Hollin, 2009). The PCS programme has a cognitive-behavioural basis and is structured to identify risk and target the criminogenic needs of higher risk offenders. It focuses on typologies and case plans, and places emphasis on offender engagement. Additionally PCS provides a pro-social and supportive organisational environment, including partnerships with community organisations that can provide ancillary services. In the evaluation study which controlled for length of time on supervision and prior

history, offenders were found to be significantly less likely to be re-arrested following PCS (30% re-arrests) compared to traditional supervision (42% re-arrests) (Taxman, 2008). A similar supervision programme, Citizenship, is discussed below and its evaluation is the subject of this paper.

The Citizenship Programme

The Citizenship community supervision programme is described in detail in Bruce and Hollin (2009). Citizenship is cognitive-behavioural, based on RNR principles (Andrews & Bonta, 2010), and targets medium- to high-risk offenders for more intensive supervision and treatment as in other evidence-based community interventions (Paparozzi & Gendreau, 2005; Taxman, 2008; Taxman & Thanner, 2006). The model of change is based on enhancing motivation for treatment (Czuchry, Sia, & Dansereau, 2006; Ward, Day, Howells, & Birgden, 2004), and modelling pro-social behaviour (Trotter, 1996), combined with community integration (Latessa & Lowenkamp, 2006; National Institute of Corrections, 2005). An underlying premise is that if offenders can be integrated into mainstream community agencies during their period of probation supervision, then there is a support framework which can continue after statutory supervision ceases (Bruce & Hollin, 2009).

The format of Citizenship, underpinned by a programme and a management manual, requires all offenders starting community supervision orders or licences to complete a compulsory 'Induction' module. This involves seven individually-delivered weekly sessions designed to: discuss legal and practical requirements of the order; identify the function of the individual's offending (two sessions); develop an awareness of the balance of decisional factors in offending; increase awareness of the impact on victims; improve problem-solving skills; and jointly plan for future interventions (goal setting). During this Induction, and following risk screening using a Risk/Needs measure (OASys, see below), level of risk, areas of need, and learning style are identified in collaboration with the offender. This approach is similar to the

shared responsibility for change between offender and offender manager included in PCS

(Taxman, 2008). After Induction the subsequent menu of additional modules covers a range of

topics found to be related to re-offending: alcohol misuse; drugs misuse; criminal lifestyle and

associates; poor relationships; and/or problems in emotional well-being. The module(s) most

appropriate to address risks and needs of the individual offender are identified after the final

session of Induction. Each module uses exercises to encourage the offender to apply problem-

solving skills and self-monitor behaviour outside of the probation office. Accredited offending

behaviour programmes[1] and/or support from external agencies may be included in this part of

the Citizenship programme, further integrating treatment with supervision. A review at the end

of Citizenship consolidates the work conducted in the earlier modules and plans for the future

using a relapse prevention framework.

Previous evaluation of Citizenship. Citizenship was originally evaluated in the probation

area where it was first developed and implemented, County Durham. This probation supervision

programme became the main means of offender management in the County Durham probation

area, and included all risk levels of offender, with the intensity of supervision being related to

the level of risk. The evaluation (Pearson et al., 2011) was on an 'intention to treat' basis,

comparing time to reconviction of a total cohort of all offenders in the probation area receiving

Citizenship over a two-year window ($n = 3,819$), with those of a retrospective total cohort of all

offenders in the same probation area receiving traditional probation supervision over a prior

period of 12 months ($n = 2,110$). The analysis controlled for level of risk, age at sentence, age at

first conviction, number of prior convictions, gender, and changes in the national rate of

reconvictions over the evaluation period.

Results showed that after two years 50% of offenders had reconvicted in the comparison

group, compared to 41% in the Citizenship group, with a significant difference between the

survival curves. The hazard of reconviction under Citizenship at any one time was 0.69 that of

the comparison group, which represented a 31% reduction in reconvictions in the Citizenship offenders (Pearson et al., 2011). However when risk levels were controlled for, contrary to other RNR research, a greater reduction in the hazard was found in low- and medium-risk categories than in the high-risk group. This discrepancy was attributed to high-risk offenders being less suited to referral to community integration agencies, due to the nature of their offences. Low-risk offenders on the other hand, in keeping with RNR principles, received a low level of intervention appropriate to their risk level with a short analysis of their offending and training in problem solving skills needed to manage future offence-related behaviour, followed by referral to an external community agency if necessary. Differences in reduction in re-offending between low level experimental and comparison groups were greater than the national reduction. A study focussing on Citizenship with higher risk offenders was thought to be warranted.

The Current Evaluation

The introduction of Citizenship to a new probation area, Teesside, offered the opportunity for research evidence to inform the selection of the target group and research design. Teesside opted to deliver the intervention with medium- and high-risk offenders only, consistent with RNR principles. Among the options for research design was the possibility of a RCT. Historically RCTs have been rarely used in the UK criminal justice system (Farrington, 2003; McDougall et al., 2006). Farrington noted that since 1960 only 14 UK criminal justice RCTs had been conducted, and very few have been performed since Farrington's review (McDougall et al., 2006). Objections to RCTs generally relate to depriving individuals of the benefits of a potentially effective intervention, practical problems of randomising in a busy organisation, and statistical problems of obtaining a large enough sample of suitable participants to randomise. In the present evaluation, ethical concerns were raised about withholding a treatment that had already been shown to be effective; and practical problems of contamination if individual offender managers were allocated both experimental and control group cases. The organisation

was however receptive to the concept of robust evaluation, and agreed to a stepped wedge cluster randomised design, which minimised the ethical and practical difficulties that had been presented. Using this design all offices would ultimately access Citizenship, with entire probation offices of target cases being randomly allocated, so reducing the potential for contamination likely with individual randomisation.

A stepped wedge cluster RCT (Hussey & Hughes, 2007) is a type of cross-over design that allows for clusters of participants to begin as no-intervention controls after which individual clusters are randomly selected in sequence at pre-planned time points to cross-over permanently from the control group to the intervention group. Thus at the end of the trial all clusters will have eventually crossed over to the intervention group. The design allows for an experimental comparison between participants in clusters receiving the intervention, and clusters receiving 'treatment as usual' awaiting cross-over to the intervention group. This approach is practically beneficial because no operational units are permanently consigned to a control group for the duration of the study, and is statistically beneficial in enabling an experimental design that may otherwise be resisted. In a systematic review of stepped wedge designs Brown and Lilford (2006) advocated that such methods were most suitable where the intervention was likely to do more good than harm, and where evidence of a beneficial effect was evident from a previous setting, which was the case with the present study.

The stepped wedge design has appeal in that, to some extent, it addresses ethical concerns in RCTs about control group participants who do not receive a treatment that might be of benefit. In the stepped wedge design all clusters eventually receive the treatment, although for some clusters the treatment will be delayed. In addition it has been demonstrated that a stepped wedge design makes implementation more manageable for practitioners than with a large-scale simultaneous introduction of the intervention, and hence can have financial advantages (Pearson, Torgerson, McDougall, & Bowles, 2010). Despite the advantages of the stepped wedge cluster design in addressing practical and ethical concerns, it has rarely been used in social sciences. A

systematic review of the application of stepped wedge trials internationally up to January 2010 revealed only twenty-five instances of its use, of which twelve were in medicine and just one, the present study, in a criminal justice setting (Mdege, Man, Taylor, & Torgerson, 2011). This may be due to a lack of awareness of the method by researchers and policy-makers in criminal justice, but the review's authors acknowledge that the stepped wedge design presents challenges due to practical and statistical complexity (Mdege et al., 2011).

The research literature would suggest that a lesser effect from the Citizenship intervention would be likely when measured by a RCT than by a quasi-experimental method. In an influential review of research designs in criminal justice, Weisburd, Lum, and Petrosino (2001) concluded that the weaker the design in terms of internal validity, the more likely it was that the study would report a positive result in favour of the treatment. In a partial replication, Welsh, Peel, Farrington, Elffers, and Braga (2011) found similar results with an overall correlation between research design and outcome which was also moderate and negative. Hence, in Citizenship it was hypothesised that reconvictions would be reduced and/or delayed in the Citizenship group compared to the control group, but that the RCT evaluation of Citizenship would show a lower level of effectiveness than the earlier quasi-experimental evaluation. This was particularly likely since the Teesside area typically has a higher level of crime.[2] In addition, the target group was higher risk and was therefore likely to be less compliant. It was anticipated that those offenders who received support from an external community agency, in addition to receiving supervision from their offender manager, would show benefit from this contact and reconvict less.

Method

Participants

Participants were offenders, aged 18 or over, under the supervision of the Teesside probation area. The experimental and control groups ($N = 1,091$), comprised all eligible

medium- and high-risk offenders sentenced to a community rehabilitation order (receiving probation supervision) or released from prison on licence (e.g., parole supervision) in the probation area, between February 1, 2007 and March 31, 2008. Participants were assigned to Citizenship according to their supervision office, which was randomly allocated to the stepped wedge (see Research Design below). Participant characteristics in each office are shown in Table 1 below for Citizenship and control group offenders.

The participant sample was mainly male (88%), of white ethnicity (97%), with a mean age of approximately 28.5 years. The predominance of the white ethnic category in the local offender population is typical of the wider non-offending population in North-East England, which, at 95% versus 88% nationally, has the highest proportion of people of White British ethnic origin in England and Wales (Office for National Statistics, 2010).

[Table 1 about here]

Risk and Need Assessment Tool

All offenders under prison or probation supervision in England and Wales have their demographic and offending data recorded on the Offender Assessment System (OASys: Home Office, 2002) database. OASys is a structured data protocol that assesses the likelihood of reconviction and identifies relevant areas of need. OASys risk assessments are based on a combination of static historical measures such as number of previous convictions and age at first offence, similar to those used in the Offender Group Reconviction Scale (OGRS: Copas & Marshall, 1998), and dynamic changeable information based on 10 variables similar to those employed by the revised Level of Service Inventory (LSI-R: Andrews & Bonta, 1995), such as education/employment, lifestyle and associates, emotional problems, and pro-criminal attitudes. An OASys risk score is calculated by adding together sub-scores, based on file and interview information, from these two static and dynamic sources. This allows for

identification of medium- and high-risk offenders. OASys has been validated (Howard, Clark, & Garnham, 2006) and was found to be a good predictor of risk relative to other existing instruments. Using the Area Under the Curve (AUC) of the receiver operating characteristic (Hanley & McNeil, 1982), OASys scored .76 in predicting the 24 month reconvictions of 757 assessed offenders (Howard et al., 2006). This level of accuracy was slightly better than the AUC of .73 observed with the LSI-R (Andrews & Bonta, 1995). This performance, however, was below that achieved by the purely static OGRS (Copas & Marshall, 1998) which produced an AUC of .80 on the same data (Howard et al., 2006). The OGRS measure has demonstrated its predictive validity with a wide range of offender populations (Coid et al., 2009; Gray, Laing, & Noaks, 2002; Lloyd, Mair, & Hough, 1994). The Coid et al. (2009) study used the revised version of OGRS (OGRS-2: Taylor, 1999) and the same measure was collected on each offender in the present study.

Eligibility Criteria

Offenders are managed by probation offender managers within a 'Tier' system, on the basis of identified risk of harm to others and likelihood of reoffending. Case management at Tiers 1 and 2 provides only monitoring and signposting, while at Tiers 3 and 4 it also integrates therapeutic and inter-agency elements. Offenders were targeted for Citizenship if they were tiered at level 3 or 4. These Tiers include offenders in the medium- to high- reconviction risk range, as well as offenders whose index crimes are so serious as to require by statute intensive supervision and treatment (e.g., domestic violence risk offenders). This explains the higher numbers in Tiers 3 and 4 (Table 1) than in OGRS and OASys medium- and high-risk categories. It is recognised that offenders generally, particularly as their OASys scores rise, will be assessed as having multiple and further entrenched criminogenic needs (Howard et al., 2006). The offender manager is tasked with selecting and sequencing the relevant Citizenship modules to meet those crime-related needs.

Research Design

Using the stepped wedge cluster RCT design (Hussey & Hughes, 2007), entire probation offices were randomly allocated[3] into the wedges at two-monthly intervals as shown in Figure 1.[4] When an office was selected for cross-over, only newly sentenced or newly released offenders commencing supervision received the intervention. The follow-up period ranged from a minimum of six months for the final office allocated, to 18 months, for earlier allocated offices.

Offender managers were initially trained in delivering Citizenship, and received 'top-up' sessions prior to entering the intervention period. All trained offender managers had their own programme manual to assist delivery. Implementation and data collection was managed by a project team at Operational Director level, and monitored throughout the implementation period in each office. The project team 'spot-check' monitoring reported that use of the programme with eligible offenders was generally good, with approximately 68% take-up across office steps (Portues, 2008).

[Figure 1 about here]

The outcome measures of the study were numbers of offenders convicted and time to reconviction, including for technical violation of the order or licence (i.e., breach of an order requiring a return to court for re-sentencing or breach of licence conditions requiring recall to custody). Reconviction data were obtained from the government department responsible for area re-offending reports.[5] Referrals of offenders to external community agencies were identified from electronic probation case records and contact logs.

Participant Progress Throughout the Trial

Participants' progress throughout the trial is illustrated by the CONSORT diagram (Consolidated Standards of Reporting Trials) in Figure 2.

[Figure 2 about here]

Ethical Considerations

For each offender, during the first session of Citizenship supervision the relevant court order or post-release licence was fully explained, together with the purpose and process of Citizenship, including monitoring data for the purposes of evaluation, and sharing information with other agencies ('limited confidentiality'). The offender then signed a legally enforceable supervision contract to this effect.

Statistical Analysis

The statistical analysis included all eligible offenders (Tiers 3 and 4) under supervision from February 1, 2007, to March 31, 2008, on an 'intention to treat' basis. Since offenders are on caseloads for different lengths of time, the study necessarily had variable follow-up intervals. Probation office units were divided for analysis purposes into two clusters per probation office, allocated simultaneously, i.e., 12 clusters in total. Office units had been randomly allocated at two monthly intervals, and the observation period covered 7 time points (i.e., 14 months). The proportion surviving at any given time (i.e., not being reconvicted or committing a technical violation) was calculated using the Kaplan-Meier survival estimate (Kaplan & Meier, 1958). A Kaplan-Meier survival curve was used to summarise the complex data graphically (e.g., Bowles & Florackis, 2007). Survival curves represent cumulative rates of reconviction over time taking into account the different periods of observation.

Discrete-time survival analysis was therefore used to explore the impact of Citizenship on time to reoffending, controlling for the following prognostic factors: OASys score, OGRS score,

and Tier, all presented individually. For each model investigated, age at sentencing, age at first conviction, number of prior convictions, gender, and agency contact were controlled for. Interaction between each of these scores and the intervention variable, Citizenship group, was examined.

To model individual level reconviction responses, a complementary log-log link function was used for the discrete-time hazard. This link was chosen because a proportional hazards model would follow for the continuous time points. The trial design was also taken into account (see Hussey & Hughes, 2007). Therefore, the following model was fitted to the data:

$$\ln\{-\ln(1-h_{mijk})\} = \Sigma\alpha_m d_m + C_j + S_k + X_{jk}\theta + \Sigma\beta_n Z_n \tag{1}$$

In model (1) h_{mijk} is the discrete-time hazard of reconviction corresponding to time point m, for individual offender i, in cluster j, and at step k. α_m is a fixed effect corresponding to duration d_m where m = 1 ..... 11 are one monthly intervals and m = 12 is the interval corresponding to months 12-20. Months 12-20 were lumped together because not many cases re-offended after 12 months. Therefore, the durations d_m for m = 1 .... 12 cover the whole period of time at risk. C_j is a random effect for cluster j such that $C_j \sim N(0, \tau^2)$. S_k is a fixed effect corresponding to step k (k in 2 … K, S_1=0 for identifiability). X_{jk} is an indicator of the treatment mode in cluster j at step k (1=intervention; 0=control), i.e., it takes the value of '1' when cluster j is in the intervention group at step k. θ is the treatment effect, and β_n is the effect of the prognostic variable Z_n for n = 1 …. This parametrisation makes no assumption about the shape of the baseline hazard/survival functions (see Rabe-Hesketh & Skrondal, 2008).

The hazard at a given time is the rate at which events (e.g., reconviction) happen. The Hazard Ratio (HR) gives an estimate of the difference in the hazard rates between the intervention and control groups.[6]

The earlier quasi-experimental analysis of Citizenship had needed to take into account a reduction in national reconvictions during the whole research period to ensure a reduction in the experimental group was not related to an overall reduction in national offending rates. This was not necessary in the present study, since the data from the experimental group and the control group were contemporaneous. However it was of interest to note that the actual national reconviction rate had not shown any statistically significant change during the time period (Ministry of Justice, 2010).

Results

Figure 3 shows the Kaplan-Meier curve for time to reconviction in the experimental and control groups. Although initially the proportion not yet reconvicted at any point in time was higher in the Citizenship group when compared to the control group, at the median of the survival time (201 days) the two lines began to converge, and subsequently the control group showed a lower probability of failure. This may reflect the dissipation of an initial effect of the programme in delaying reconvictions, or the lower proportion of offenders subsequently receiving Citizenship.

[Figure 3 about here]

Risk of Reconviction

Using model (1), separate models were created controlling for OGRS, OASys, and Tier, and results from these were similar. In the first instance, no interaction term was used between the intervention group and the risk level of each measure. Over all 'intention to treat' offenders, there was a 20% reduction in the proportion reconvicted under Citizenship compared to the control group. The effect in each model was not statistically significant: OGRS: Exp(b) = 0.80,

p =.19, 95% Confidence Interval (CI): 0.58 – 1.11; OASys: Exp(b) = 0.79, *p* =.17, 95% CI: 0.57-1.10; Tier: Exp(b) = 0.81, *p* =.20, 95% CI: 0.58-1.12.

[Table 2 about here]

To check whether the effect of risk of reconviction on the hazard differed by treatment group, an interaction term was introduced between group and risk level in each of the three risk models. In each case the model with the interaction term was compared to a model without the specified interaction. As shown in Table 2 the hazard for medium-risk offenders under Citizenship in the OGRS model was 4% lower than the control group (Exp(b) = 0.96, 95% CI: 0.60-1.52), while that for high-risk offenders was 34% lower (Exp(b) = 0.66, 95% CI: 0.46-0.97). Although the effect for high-risk offenders is statistically significant, caution should be exercised, as the upper confidence level is very close to 1.00. For the high-risk categories within OASys and Tier, there was a 24% and 26% reduction respectively comparing Citizenship to the control group, and for medium-risk 20% and 15%, though the effects were not statistically significant. This may reflect the dynamic nature of risk assessment in OASys and Tier as opposed to OGRS which measures static risk. Age and number of prior convictions had a small significant effect on reconviction in all three models, but again the limits of the confidence intervals were close to 1.00.

External Agency Contact

The effect on the hazard of the offender attending a meeting with an external support agency was examined by adding the variable 'agency contact' to the risk models without interaction terms between risk score and the treatment variable. This did not strengthen the model in detecting a difference between the intervention and the control group, compared to the values produced in the absence of controlling for agency contact. For example, the effect of

agency contact in the model which controlled for OGRS, age at sentencing, age at first

conviction, number of prior convictions, gender and design variables was non-significant,

Exp(b) = 1.02, p =.84, 95% CI: 0.84-1.24. Similar estimates for agency contact were observed

when this variable was added to the models that controlled for the other risk measures (OASys

and Tier).

Impact on Technical Violations

A total of 456 reconvictions were recorded out of which there were 31 technical violations

in total with 13 observed under Citizenship. This sample resulted in 1,735 time intervals. Three

independent models (OGRS, OASys and Tier) examined the difference between the

experimental and control groups on time to a technical violation resulting in breach. The results

for the three models were similar. For example with OGRS in the model, the effect of

Citizenship supervision on the hazard of technical violation was positive but non significant

(Exp(b) = 0.84, p =.77, 95% CI: 0.24-2.81).

Discussion

Traditional probation supervision is not systematically delivered, tends to be atheoretical,

and does not show detectable effects on recidivism (Bonta et al., 2008; Drake, 2011). The

current study presented an evaluation of the implementation of an RNR supervision programme,

using a robust RCT design. Results showed a 20% overall reduction in reconvictions in the

Citizenship group when compared with the control group, but this effect was not statistically

significant. When interaction with OGRS risk level was taken into account, the effect of

Citizenship on high-risk offenders was to reduce reconvictions by 34% compared with the

control group and this was statistically significant, although it should be noted that the upper

confidence level was very close to 1.00 so the result is borderline. The effect of age and number

of previous convictions also reached borderline significance. The effect on high-risk offenders

of the current probation supervision programme is in line with the results from Paparozzi and Gendreau (2005) and Latessa and Lowenkamp (2006), who found that the application of RNR principles had most impact on reoffending among high-risk offenders. The results agree with findings from the meta-analysis of community supervision (Drake, 2011), which demonstrated that supervision focused on RNR achieves a greater reduction in recidivism than intensive supervision without treatment of criminogenic needs. This is particularly relevant for the high-risk offenders in Citizenship where intensity of supervision with treatment specifically related to risk, reduced recidivism.

The PCS study (Taxman, 2008) was included in Drake's (2011) meta-analysis. The Citizenship programme had many similarities to PCS, focusing on motivation and a social learning approach in relation to RNR, and indeed achieved a similar reduction in proven reconvictions compared with arrests in the PCS evaluation. The programmes differed in that PCS focused on typologies in supervision plans, while Citizenship carefully identified risks and individual criminogenic needs to inform sentence plans.

The current RCT results give modest encouragement that, even under the rigorous 'intention to treat' analysis, the structured RNR supervision programme, combined with motivation and social learning techniques, appears to have had an impact on the offending of high-risk offenders, as measured by OGRS. We could not find any evidence that integration with external community agencies impacted on reduced reconvictions, and this tends to confirm the conclusion from the County Durham quasi-experimental study in which external agency contact was beneficial to lower risk offenders, but not to high-risk offenders.

The difference in results between the present study (20% overall reduction in reconvictions) and the previous evaluation of Citizenship (31% reduction in reconvictions), was anticipated since it supports conclusions from reviews suggesting that stronger research designs are less likely to obtain statistically significant positive outcomes (Weisburd et al., 2001; Welsh et al., 2011). Although the methodological differences were likely to produce different results

(i.e., a RCT in Teesside and a quasi-experimental study in County Durham), there were however other more practical differences between the studies that have had an impact. It is therefore appropriate to examine the results in terms of the problems in real-life studies, where there are numerous threats to implementation (Lipsey et al., 2007; Welsh et al., 2011), including potential practitioner resistance to implementing a RCT (Farrington, Gottfredson, Sherman, & Welsh, 2002; Kennedy & Grant, 1997; Schulz, 1995). For example in the current study 'treatment received' was lower than anticipated, across both medium- and high-risk offenders. Although project team monitoring suggested a 68% up-take during implementation, final analysis of the data revealed that only 38% of the 'intention to treat' sample actually received treatment, or there were inaccuracies in the recording of cases. This compared with 75% implementation in the previous evaluation, which demonstrates that a higher level of 'treatment received' should have been organisationally possible. As can be seen from the survival curve in the present study (Figure 3), early differences in time to reconviction between the intervention and control groups were encouraging. It was evident that the stepped wedge design facilitated the introduction of the Citizenship programme when compared to a full-scale implementation in a neighbouring area, which initially suffered implementation failure (Pearson et al., 2010). This is a common problem and it has been observed in other studies that programme integrity cannot be assumed beyond the implementation phase (Bonta et al., 2011).

An analysis of the data from the 38% who received treatment has not been presented as these do not represent a random sample. It was observed however that in those 38% there were positive effects on reconvictions compared with those who did not receive Citizenship, and that the number of reconvictions reduced as the number of sessions of Citizenship received increased. One must however question whether the reduction in re-offending in the Citizenship group was related to the characteristics of the particular offenders rather than to the supervision programme, i.e., the offenders were more motivated to take advantage of the programme. Notwithstanding, a positive effect with higher risk offenders under the RCT given the small

proportion overall that received treatment, suggests that a more reliable effect might have been achieved with greater commitment to the 'intention to treat' group.

In methodological terms, it is evident from this study that a stepped wedge cluster RCT design offers the opportunity to overcome some of the real-life practical and ethical objections to the application of RCTs. The design can be of value in situations where a large-scale implementation is impractical, and in the present study setting it was found to assist the initial implementation of the programme through staged introduction (Pearson et al., 2010). The under-use of the design in criminal justice may be due to a lack of awareness among researchers, as well as the perceived practical and statistical complexity associated with the design. A high level of statistical expertise was required to conduct the analysis in the current study. The study has however demonstrated that application of a stepped wedge clustered RCT is possible in a criminal justice setting. Based on the negative experience in a neighbouring area that attempted simultaneous area-wide implementation, reported in Pearson et al. (2010), it seems unlikely that implementing the programme more rapidly, with less time between steps, would have improved the rate of treatment received. Moreover this would have given less time for treatment to be delivered within each step. In the present authors' opinion, loss of statistical power due to a low treatment received rate, was unrelated to the RCT methodology and could have occurred with any research design requiring rigorous attention to programme integrity. Brown and Lilford (2006) and Mdege et al. (2011) advocate that stepped wedge designs are suitable where the intervention is likely to do more good than harm, and where evidence of a beneficial effect is evident from a previous setting, as in the present study. However, having reviewed all studies, they acknowledge that the practical and statistical complexity of the design requires careful planning and monitoring to ensure robust evaluation, a view which has been supported here.

Policy and Practice Implications

Discrepancies between early programme monitoring and later data collection emphasise the need for close monitoring throughout the initial implementation and follow-up phases. In the absence of a contractual obligation for the agency to ensure the programme was implemented thoroughly, this was allowed to drift. This emphasises the importance of policy support in programme implementation. It highlights also the importance of adopting an 'intention to treat' statistical analysis. Despite the best efforts of the researchers and managers, programme implementation was not as proposed, and we must measure what happens in reality, not what we would hope for in a large scale implementation.

A related learning point from this study is that the possibility of practitioner resistance should be anticipated and addressed carefully before the programme commences, and during training. Bonta et al.'s (2011) study illustrated that the quality of staff training in RNR can have a significant impact on reconviction outcomes. In their study one group of staff were given three days training in RNR skills, followed by monthly follow-up meetings and clinical supervision, and a control group received a half day briefing session discussing the goals of the research and project requirements. Results showed that the group that had received thorough skills training achieved a significant reduction in reoffending in the offenders they supervised, compared to the group that received only project briefings. This confirms that thorough training and clinical supervision follow-up are essential to the initial and continued performance of offender managers in maintaining effectiveness. The training received by the Teesside group was more in line with the Bonta et al. 'briefing' group training, being of similar length and intensity, although it did concentrate on how Citizenship should be delivered, and focused on supervision practice. Although County Durham received the same length and intensity of training, there was more sustained clinical follow-up in supervision sessions, and in monitoring the use of Citizenship. The PCS programme also appeared to concentrate more on staff training than the Teesside Citizenship implementation, which may have had an impact. The need for monitoring to ensure continuing performance standards, illustrates that the move to evidence-based structured

probation supervision will require a radical change in working style for many probation offender managers, which may be resisted. Bourgon, Bonta, Rugge, Scott, and Yessine (2010) confirm this, saying that that importance of probation officer 'buy-in' to the theoretical view of the intervention is often under-estimated, and must be addressed in training. It is necessary therefore to learn from this experience about the nature and amount of training and supervision that may be required, even with already trained and experienced practitioners.

Research in real-life settings is recognised as being more challenging than in demonstration projects, and more difficult to manage (Lipsey, 1992). The present research has offered some important lessons: RCT designs are possible to implement in community probation settings; academic/practitioner collaborations, as proposed by Petersilia (2008), can work in adopting gold-standard research designs, but the impetus and clinical support must be maintained beyond the implementation phase. Drake's (2011) systematic review of RNR programming in community supervision has demonstrated the effectiveness of this approach in reducing reconvictions with medium- and high-risk offenders. The present evaluation of the Citizenship structured supervision programme, using a rigorous RCT design, provides further support and evidence of the effectiveness of RNR in reducing reconvictions with high-risk offenders.

Footnotes

1   Programmes which have been approved by the Correctional Services Accreditation Panel (CSAP) of England and Wales as being based on international research evidence of what is effective in reducing reoffending. Programmes are included in the offender's Citizenship pathway if they are a legal requirement in the sentence plan.

2   Reoffending rates in Teesside in 2008 were 15.4%, as measured in a three-month independent follow-up, while in County Durham they were 12.2% for the same year and length of follow-up (Ministry of Justice, 2009).

3   Randomisation of office units was carried out by DT using a coin toss procedure, witnessed by the research team. Neither those administering nor delivering the programme, nor the statistical analyst was present.

4   The two month interval was selected as it met the practical need to roll out the programme across the probation area within a single year.

5   Re-offending performance reports are provided quarterly by the Research, Development and Statistics department of the National Offender Management Service [NOMS]).

6   Technically based on the assumption of proportional hazards, survival in the experimental group equals survival in the comparison group to the power of the hazard ratio.

References

Andrews, D. A., & Bonta, J. (1995). *The Level of Service Inventory - Revised*. Toronto: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct,* 5th edn. New Providence, NJ: LexisNexis Matthew Bender.

Andrews, D. A., Bonta, J., & Hoge, I. (1990). Classification for effective rehabilitation:

Rediscovering psychology. *Criminal Justice and Behavior, 17,* 19-52.

doi:10.1177/0093854890017001004

Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does

correctional treatment work? A clinically relevant and psychologically informed meta-

analysis. *Criminology, 28,* 369-404. doi:10.1111/j.1745-9125.1990.tb01330.x

Antonowicz, D., & Ross, R. (1994). Essential components of successful rehabilitation programs

for offenders. *International Journal of Offender Therapy and Comparative Criminology*, *38*,

97-104. doi: 10.1177/0306624X9403800202

Bonta, J., Bourgon, G., Rugge, T., Scott, T-L., Yessine, A. K., Gutierrez, L., & Li, J. (2011). An

experimental demonstration of training probation officers in evidence-based community

supervision. *Criminal Justice and Behavior, 38*, 1127-1148. doi: 10.1177/0093854811420678

Bonta, J., Rugge, T., Scott, T-L., Bourgon, G., & Yessine, A. K. (2008). Exploring the black

box of community supervision. *Journal of Offender Rehabilitation, 47,* 248-270.

doi**:**10.1080/10509670802134085

Bourgon, G., Bonta, J., Rugge, T., Scott, T-L., & Yessine, A. K. (2010). The role of program

design, implementation and evaluation in evidence-based 'real world' community

supervision. *Federal Probation, 74,* 1, 2-15.

Bowles, R. A., & Florackis, C. (2007). Duration of the time to reconviction: Evidence from UK

prisoner discharge data. *Journal of Criminal Justice, 35*, 365-378.

doi:10.1016/j.jcrimjus.2007.05.002

Brown, C. A. & Lilford, R. J. (2006). The stepped wedge trial design: a systematic review.

BMC Medical Research Methodology, 6, 54-63. doi: 10.1186/1471-2288-6-54.

Bruce, R., & Hollin, C. R. (2009). Developing 'Citizenship': A structured supervision

programme. *EuroVista: Probation and Community Justice, 1* (1), 24-31.

Coid, J., Yang, M., Ulrich, S., Zhang, T., Sizmur, S., Roberts, C., …, & Rogers, R. (2009). Gender differences in structured risk assessment: Comparing the accuracy of five instruments. *Journal of Consulting and Clinical Psychology, 77* (2), 337-348. doi:10.1037/a0015155

Copas, J., & Marshall, P. (1998). The offender group reconviction scale: A statistical reconviction score for use by probation officers. *Journal of Applied Statistics, 47*, 159-171.

Czuchry, M., Sia, T. L., & Dansereau, D. F. (2006). Improving early engagement and treatment readiness of probationers. *The Prison Journal, 86,* 56-74. doi:10.1177/0032885505283877

Drake, E. K. (2011). *'What Works' in community supervision: Interim report (Document No. 11-12-1201).* Olympia: Washington State Institute for Public Policy.

Farrington, D. P. (2003). British randomised experiments on crime and justine. *Annals of the American Academy of Political and Social Science, 589*, 150-167. doi:10.1177/0002716203254695

Farrington, D. P., Gottfredson, D. C., Sherman, L. W., & Welsh, B. C. (2002). The Maryland scientific methods scale. In L. W. Sherman, D. P. Farrington, B. C. Welsh, and D. L. MacKenzie (Eds.) *Evidence-based crime prevention*, (pp.13-21). London: Routledge.

Gray, N., Laing, J., & Noaks, L. (2002). *Criminal justice, mental health and the politics of risk.* London: Cavendish Publishing.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29-36.

Home Office. (2002). *Offender Assessment System: User Manual.* London: National Probation Service.

Howard, P., Clark, D., & Garnham, N. (2006). *An evaluation of the Offender Assessment System (OASys) in three pilots.* London: Home Office.

Hussey, M., & Hughes, J. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials, 28*, 182-191. doi:10.1016/j.cct.2006.05.007

Kaplan, E., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*, 457-481. doi:10.1080/01621459.1958.10501452

Kennedy, A., & Grant, A. (1997). Subversion of allocation in a randomised controlled trial. *Controlled Clinical Trials, 18* (Supplement 3), 77-88. doi:10.1016/S0197-2456(97)91044-8

Latessa, E., & Lowenkamp, C. (2006). What works in reducing recidivism? *University of St. Thomas Law Journal, 3* (3), 521-535.

Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. D. Cook, H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light,…F. Mosteller (Eds.), *Meta-analysis for explanation: A casebook* (pp. 83-127). New York, NY: Russell Sage Foundation.

Lipsey, M. W., Landenberger, N. A., & Wilson, S. (2007). Effects of cognitive-behavioural programs for criminal offenders. *Campbell Systematic Reviews 2007:6*. Campbell Collaboration. doi:10.4073/csr.2007.6

Lloyd, C., Mair, C., & Hough, M. (1994). Explaining reconviction rates: A critical analysis. *Home Office Research Study No. 136.* London: Her Majesty's Stationery Office.

McDougall, C., Perry, A. E., & Farrington, D. P. (2006). Overview of effectiveness of criminal justice interventions in the UK. In A. Perry, C. McDougall, and D. Farrington, *Reducing crime: The effectiveness of criminal justice interventions* (pp. 163-226). Chichester, UK: Wiley.

McGuire, J. (1995). *What Works: Reducing reoffending.* Chichester: Wiley.

Mdege, N. D., Man, M-S., Taylor, C. A., & Torgerson, D. J. (2011). Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *Journal of Clinical Epidemiology, 64*, 936-948. doi:10.1016/j.jclinepi.2010.12.003

Ministry of Justice (2009). Local adult reoffending 1 October 2007 – 30 September 2008 England and Wales. *Ministry of Justice Statistics Bulletin.* London: Ministry of Justice. Available from http://www.justice.gov.uk/publications/docs/local-adult-reoffending-oct07-sept08.pdf

Ministry of Justice (2010). Reoffending of adults: Results from the 2008 cohort. *Ministry of Justice Statistics Bulletin.* London: Ministry of Justice. Available from http://www.justice.gov.uk/publications/docs/reoffending-adults-2008-cohort.pdf

National Institute of Corrections. (2005). Implementing evidence based practice in community corrections: The principles of effective interventions. Washington, DC: National Institute of Corrections.

Office for National Statistics (2010). Census 2001 – Ethnicity and religion in England and Wales. Retrieved February 11, 2010, from www.statistics.gov.uk/census2001/profiles/commentaries/north_east.asp#country

Paparozzi, M., & Gendreau, P. (2005). An intensive supervision program that worked: Service delivery, professional orientation, and organizational supportiveness. *The Prison Journal, 85* (4), 445-466. doi:10.1177/0032885505281529

Pearson, D. A. S., McDougall, C., Kanaan, M., Bowles, R. A., & Torgerson, D. J. (2011). Reducing criminal recidivism: Evaluation of a structured process for probation supervision. *Journal of Experimental Criminology, 7,* 73-102. doi:10.1007/s11292-010-9115-3

Pearson, D. A. S., Torgerson, D. J., McDougall, C., & Bowles, R. A. (2010). Parable of two agencies, one of which randomizes. *Annals of the American Academy of Political and Social Science, 628* (1), 11-29. doi:10.1177/0002716209351500

Petersilia, J. (2008). Influencing public policy: An embedded criminologist reflects on California prison reform. *Journal of Experimental Criminology, 4*, 335-356. doi:10.1007/s11292-008-9060-6

Portues, R. (2008). *Teesside Citizenship implementation monitoring*. Unpublished internal

 document, Durham Tees Valley Probation Trust.

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata,*

 2nd edn. USA: Stata Press.

Schulz, K. F. (1995). Subverting randomisation in controlled trials. *Journal of the American*

 Medical Association, 274, 1456-1458. doi:10.1001/jama.1995.03530180050029

Taxman, F. (2002). Supervision - Exploring the dimensions of effectiveness. *Federal*

 Probation, 66 (2), 14-27.

Taxman, F. (2008). No illusions: Offender and organizational change in Maryland's proactive

 community supervision efforts. *Criminology and Public Policy, 7* (2), 275-302.

 doi:10.1111/j.1745-9133.2008.00508.x

Taxman, F., & Thanner, M. H. (2006). Risk, need and responsivity (RNR): It all depends.

 Crime & Delinquency, 52 (1), 28-51. doi:10.1177/0011128705281754

Taylor, R. (1999). Predicting reconvictions for sexual and violent offences using the revised

 offender group reconvictions scale. *Home Office Research Findings No. 104.* London: Home

 Office.

Thanner, M. H., & Taxman, F. (2003). Responsivity: The value of providing intensive services

 to high-risk offenders. *Journal of Substance Abuse Treatment, 24* (2), 137-147.

 doi:10.1016/S0740-5472(02)00352-5

Trotter, C. (1996). The impact of different supervision practices in community corrections:

 Cause for optimism. *Australian & New Zealand Journal of Criminology, 29* (1), 1-19. doi:

 10.1177/000486589602900103

Ward, T., Day, A., Howells, K., & Birgden, A. (2004). The multifactor offender readiness

 model. *Aggression and Violent Behavior, 9* (6), 645-673. doi:10.1016%2Fj.avb.2003.08.001

Weisburd, D., Lum, C., & Petrosino, A. (2001). Does research design affect study outcomes in Criminal Justice. *Annals of the American Academy of Political and Social Science, 578,* 50--70. doi:10.1177/0002716201578001004

Welsh, B., Peel, M., Farrington, D., Elffers, H., & Braga, A. (2011). Research design influence on study outcomes in crime and justice: A partial replication with public area surveillance. *Journal of Experimental Criminology, 7* (2), 183-198. doi:10.1007/s11292-010-9117-1

Dominic Pearson, PhD, is a chartered forensic psychologist and senior lecturer at the University of Portsmouth. He is a licensed practitioner, working for many years with the UK probation services. He is interested in offender risk assessment and the evaluation and development of programmes to reduce reoffending.

Cynthia McDougall, PhD, is a professor of psychology at the University of York and co-director of York Criminal Justice Economics. She has worked as a consultant to the probation services and has promoted the implementation of randomised controlled trials, both in the probation service and in HM Prison Service.

David Torgerson, PhD, is director of the York Trials Unit. He has a methodological interest in randomised trials in both the social sciences and health sciences. He is co-author of *Designing Randomised Trials in Health, Education, and the Social Sciences* (Palgrave Macmillan 2008).

Mona Kanaan, PhD, is a lecturer in Health Statistics at the University of York, UK. She has worked on the analysis of randomised trials, including the Stepped Wedge Design, both in the health sciences and social sciences. She has also worked on the analysis of infectious diseases and spatial statistics.

Roger Bowles, PhD, is a professor of economics and co-director of York Criminal Justice Economics. He has worked on a wide range of areas, including corruption, divorce, removal of illegal gain, and the economics of pretrial detention. He has also published widely on policy-related research, applying methods of economic appraisal and evaluation to criminal justice interventions and projects.

Tables

Table 1

Participant Characteristics by Office and Citizenship Group

	Office											
	1. Southbank		2. Redcar		3. Hartlepool		4. Middlesbrough (a)		5. Middlesbrough (b)		6. Stockton	
Group	Control	Citizenship	Control	Citizenship	Control	Citizenship	Control	Citizenship	Control	Citizenship	Control	Citizenship
N	4	47	61	81	119	110	99	60	159	56	254	41
Gender:												
Male, *n* (%)	4 (100.00)	41 (87.23)	59 (96.72)	74 (91.36)	95 (79.83)	100 (90.91)	87 (87.88)	49 (81.67)	142 (89.31)	46 (82.14)	229 (90.16)	37 (90.24)
Ethnicity:												
White, *n* (%)	3 (75.00)	46 (97.87)	61 (100.00)	80 (98.77)	118 (99.16)	110 (100.00)	92 (92.93)	58 (96.67)	151 (94.97)	53 (94.64)	244 (96.06)	41 (100.00)
Age, *M* (*SD*)	26.0 (8.16)	29.38 (7.96)	27.64 (7.25)	27.67 (8.08)	28.63 (8.00)	28.58 (7.94)	26.76 (9.28)	29.05 (6.92)	28.36 (7.90)	29.00 (8.58)	28.35 (8.31)	30.39 (7.90)
Number of Prior Convictions, *M* (*SD*)	14.75 (6.50)	17.04 (13.25)	15.57 (11.92)	13.86 (10.77)	19.26 (16.90)	18.80 (15.68)	14.68 (10.39)	21.43 (17.27)	17.78 (14.98)	16.50 (14.75)	15.30 (12.79)	17.37 (15.91)
Age at first Conviction, *M* (*SD*)	13.75 (2.87)	15.89 (4.76)	16.25 (4.19)	17.29 (6.79)	16.26 (4.54)	15.96 (5.06)	15.72 (3.17)	16.40 (4.15)	16.05 (3.58)	16.36 (4.36)	17.02 (4.96)	16.59 (4.35)
OGRS score, *M* (*SD*)	81.50 (9.33)	70.40 (22.12)	72.49 (24.77)	69.66 (24.49)	71.41 (25.53)	72.69 (25.28)	74.88 (24.01)	75.89 (22.70)	72.59 (24.52)	69.67 (23.49)	68.60 (25.48)	65.85 (25.99)
OGRS score grouped, *n* (%)												
0-40	0	6	9	12	19	17	12	6	20	6	47	9

	(0.00)	(12.77)	(14.75)	(14.81)	(15.97)	(15.45)	(12.12)	(10.00)	(12.58)	(10.71)	(18.50)	(21.95)
41-75	1	19	15	24	31	25	19	16	47	21	73	12
	(25.00)	(40.43)	(24.59)	(29.63)	(26.05)	(22.73)	(19.19)	(26.67)	(29.56)	(37.50)	(28.74)	(29.27)
76+	3	22	33	44	67	66	62	35	87	28	126	19
	(75.00)	(46.81)	(54.10)	(54.32)	(56.30)	(60.00)	(62.63)	(58.33)	(54.72)	(50.00)	(49.61)	(46.34)
OASys score, *M* (*SD*)												
	95.75	102.02	92.75	97.39	101.44	104.00	104.64	110.45	101.37	98.71	98.63	99.85
	(12.61)	(31.78)	(38.89)	(32.09)	(32.42)	(29.37)	(27.86)	(25.80)	(29.08)	(29.83)	(33.22)	(33.10)
OASys score grouped, *n* (%)												
0-49	0	4	9	9	11	6	5	2	11	4	27	4
	(0.00)	(8.51)	(14.75)	(11.11)	(9.24)	(5.45)	(5.05)	(3.33)	(6.92)	(7.14)	(10.63)	(9.76)
50-99	2	13	16	26	37	36	25	15	53	16	80	16
	(50.00)	(27.66)	(26.23)	(32.10)	(31.09)	(32.73)	(25.25)	(25.00)	(33.33)	(28.57)	(31.50)	(39.02)
100+	2	30	34	45	70	68	68	43	92	36	141	21
	(50.00)	(63.83)	(55.74)	(55.56)	(58.82)	(61.82)	(68.69)	(71.67)	(57.86)	(64.29)	(55.51)	(51.22)
Tier												
T3	3	36	46	55	76	67	61	35	114	45	174	23
	(75.00)	(76.60)	(75.41)	(67.90)	(63.87)	(60.91)	(61.62)	(58.33)	(71.70)	(80.36)	(68.50)	(56.10)
T4	1	11	15	26	43	43	38	25	45	11	80	18
	(25.00)	(23.40)	(24.59)	(32.10)	(36.13)	(39.09)	(38.38)	(41.67)	(28.30)	(19.64)	(31.50)	(43.90)
Offence Category, *n* (%)												
Violence	0	5	4	4	6	8	11	3	18	2	13	2
	(0.00)	(29.41)	(12.90)	(11.11)	(11.11)	(16.33)	(20.75)	(15.79)	(30.51)	(10.53)	(13.40)	(11.76)
Sexual	0	0	1	0	0	0	0	0	0	0	0	0
	(0.00)	(0.00)	(3.23)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Robbery	0	0	0	1	1	0	1	0	0	0	0	0
	(0.00)	(0.00)	(0.00)	(2.78)	(1.85)	(0.00)	(1.89)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Burglary/ Theft/Fraud	0	5	20	15	30	23	24	9	18	8	45	9
	(0.00)	(29.41)	(64.52)	(41.67)	(55.56)	(46.94)	(45.28)	(47.37)	(30.51)	(42.11)	(46.39)	(52.94)
Criminal Damage	2	2	0	2	1	3	2	1	2	1	6	2
	(100.00)	(11.76)	(0.00)	(5.56)	(1.85)	(6.12)	(3.77)	(5.26)	(3.39)	(5.26)	(6.19)	(11.76)

Drugs	0 (0.00)	0 (0.00)	1 (3.23)	3 (8.33)	1 (1.85)	3 (6.12)	1 (1.89)	0 (0.00)	3 (5.08)	0 (0.00)	6 (6.19)	0 (0.00)
Motoring	0 (0.00)	2 (11.76)	1 (3.23)	1 (2.78)	4 (7.41)	3 (6.12)	1 (1.89)	0 (0.00)	4 (6.78)	1 (5.26)	4 (4.12)	1 (5.88)
Other	0 (0.00)	3 (17.65)	4 (12.90)	10 (27.78)	11 (20.37)	9 (18.37)	13 (24.53)	6 (31.58)	14 (23.73)	7 (36.84)	23 (23.71)	3 (17.65)

Note. Values represent counts (percentages) unless otherwise stated, with percentages calculated taking into account missing data. M = mean; SD = standard deviation.

Table 2

Models Predicting Reconviction Controlling for Risk Scores (categories) Adjusting for Agency and other Covariates

	Model A1 OGRS		Model B1 OASys		Model C1 Tier	
	(N=895)		(N=985)		(N=1090)	
	Exp(b)	95% CI	Exp(b)	95% CI	Exp(b)	95% Ci
OGRS *Group						
OGRS(41-75): Cit vs Comp	0.96	0.60-1.52				
OGRS(76+): Cit vs Comp	0.66*	0.46-0.96				
Group: Comparison						
OGRS: 76+ vs 41-75	1.21	0.88-1.67				
Group: Citizenship						
OGRS: 76+ vs 41-75	0.84	0.56-1.27				
OASys *Group						
OASys(50-99): Cit vs Comp			0.80	0.51-1.25		
OASys(100+): Cit vs Comp			0.76	0.53-1.09		
Group: Comparison						
OASys: 100+ vs 50-99			1.07	0.81-1.41		
Group: Citizenship						
OASys: 100+ vs 50-99			1.02	0.70-1.48		
Tier *Group						
Tier 3: Cit vs Comp					0.85	0.60-1.23
Tier 4: Cit vs Comp					0.74	0.49-1.13

Group: Comparison						
Tier 4 vs Tier 3					1.31*	1.02-1.69
Group: Citizenship						
Tier 4 vs Tier 3					1.14	0.80-1.63
Age	0.97**	0.96-0.99	0.97*	0.96-0.99	0.97*	0.96-0.99
Age at First Conviction	1.02	0.99-1.04	1.02	0.99-1.05	1.02	0.99-1.04
Gender: Female vs Male	1.27	0.96-1.69	1.25	0.94-1.65	1.30	0.98-1.72
Number of Prior Convictions	1.01*	1.00-1.02	1.01*	1.00-1.02	1.01*	1.00-1.02
Agency: Yes vs No	1.02	0.84-1.24	1.02	0.84-1.23	1.04	0.86-1.26
Step 1	0.90	0.65-1.25	0.93	0.67-1.29	0.91	0.66-1.26
Step 2	1.13	0.80-1.62	1.15	0.81-1.63	1.16	0.82-1.64
Step 3	1.49	1.05-2.11	1.5	1.05-2.11	1.46	1.03-2.07
Step 4	0.64	0.41-1.00	0.65	0.41-1.01	0.64	0.41-1.00
Step 5	0.86	0.55-1.36	0.86	0.55-1.36	0.85	0.54-1.34
Step 6	1.28	0.76-2.14	1.27	0.76-2.14	1.24	0.74-2.09
Log-likelihood	-1563		-1583		-1583	
Intra-class correlation (ICC)[a]	0.02*	0.00-0.10	0.02*	0.00-0.10	0.02*	0.00-0.10

Note. [a] Residual ICCs.

*   = p<.05   **= p<.01   ***= p<.001

Figures

	step 0	step 1	step 2	step 3	step 4	step 5	step 6	
Stockton	Stockton	Stockton	Stockton	Stockton	Stockton	Stockton	**Stockton**	
M'bro (b)	M'bro (b)	M'bro (b)	M'bro (b)	M'bro (b)	M'bro (b)	**M'bro (b)**	M'bro (b)	FOLLOW UP DATA
M'bro (a)	M'bro (a)	M'bro (a)	M'bro (a)	M'bro (a)	**M'bro (a)**	M'bro (a)	M'bro (a)	
Hartlepool	Hartlepool	Hartlepool	Hartlepool	**Hartlepool**	Hartlepool	Hartlepool	Hartlepool	
Redcar	Redcar	Redcar	**Redcar**	Redcar	Redcar	Redcar	Redcar	
Southbank	Southbank	**Southbank**	Southbank	Southbank	Southbank	Southbank	Southbank	
	1/2/07-31/3/07	1/4/07-31/5/07	1/6/07-31/7/07	1/8/07-30/9/07	1/10/07-30/11/07	1/12/07-31/1/08	1/2/08-31/3/08	1/4/08-30/9/08

control group commencements

Citizenship group commencements
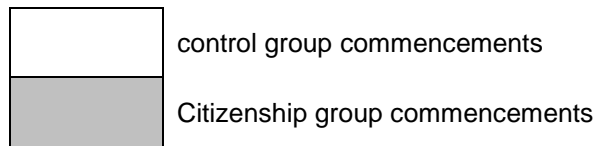
Note. M'bro – refers to Middlesbrough offices.

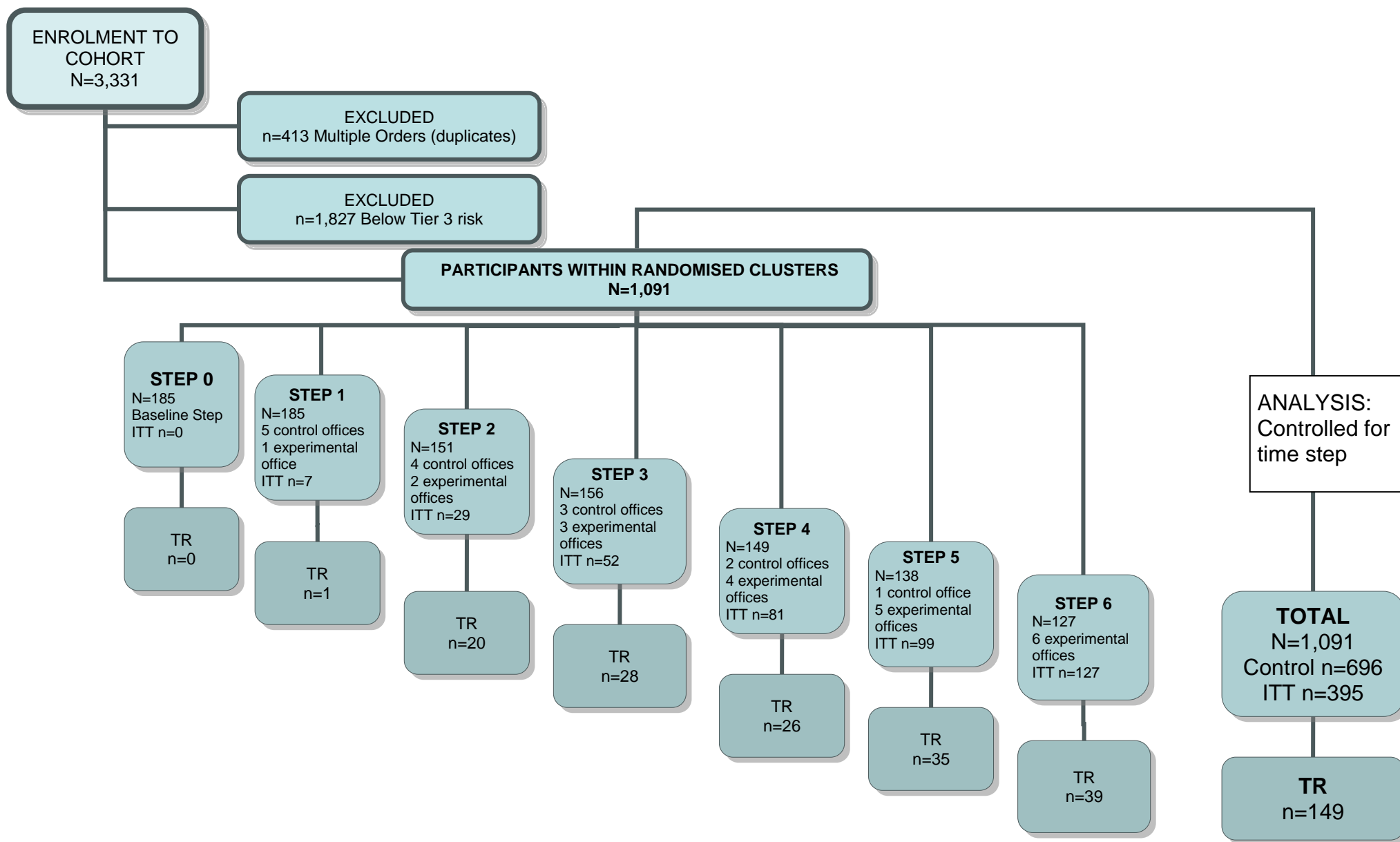*Figure 1.* Model of office implementation

Figure 2. CONSORT diagram showing participants within randomised stepped wedge clusters
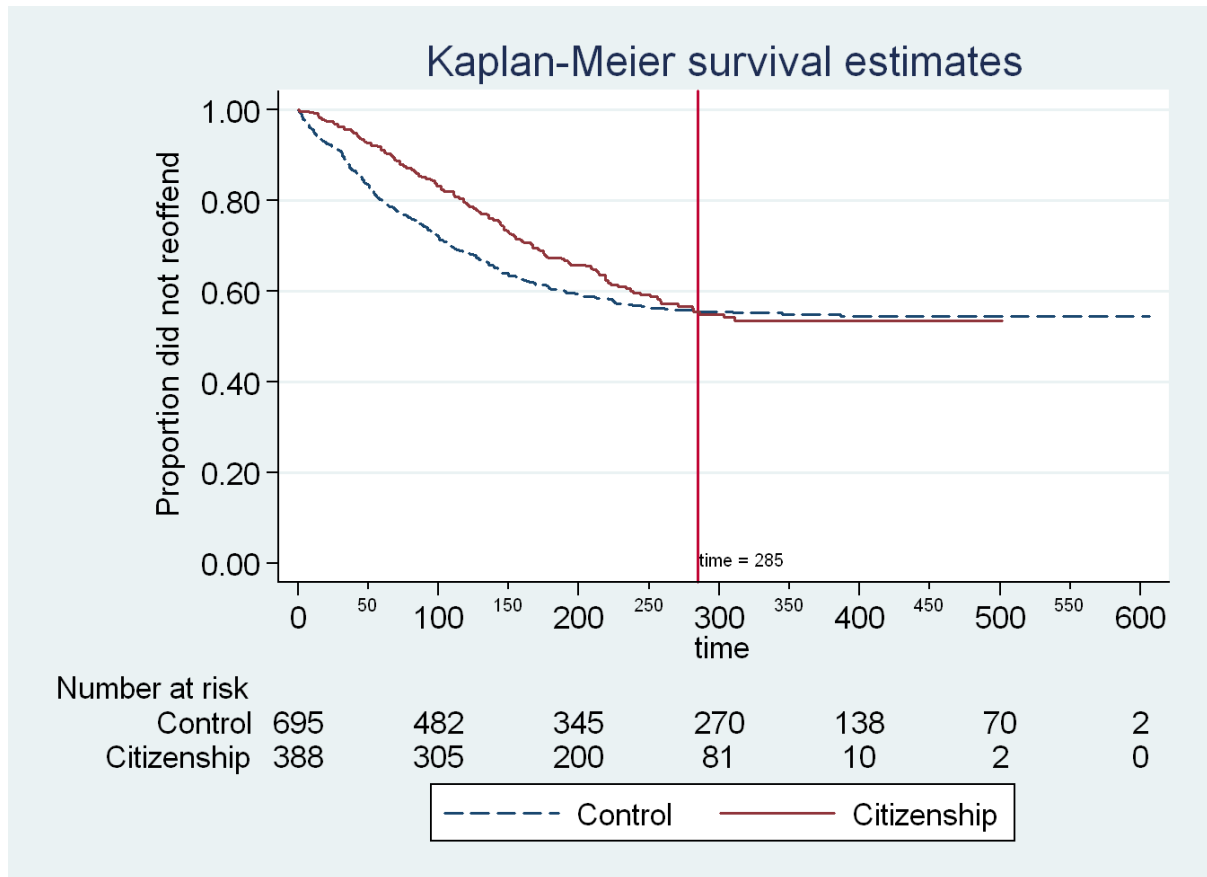
Figure 3. Kaplan-Meier survival curve for reconvictions by group