



This is a repository copy of *Using compliance with probation supervision as an interim outcome measure in evaluating a probation initiative*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/100493/>

Version: Accepted Version

Article:

Sorsby, A., Shapland, J. and Robinson, G. (2017) Using compliance with probation supervision as an interim outcome measure in evaluating a probation initiative. *Criminology and Criminal Justice*, 17 (1). pp. 40-61. ISSN 1748-8958

<https://doi.org/10.1177/1748895816653992>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Using compliance with probation supervision as an interim outcome measure in evaluating a probation initiative

Angela Sorsby

University of Sheffield, UK

Joanna Shapland

University of Sheffield, UK

Gwen Robinson

University of Sheffield, UK

Abstract

This article addresses the issues involved in using compliance with probation supervision as an interim outcome measure in evaluation research. We address the complex nature of compliance and what it implies. Like much research on probation and criminal justice more generally, it was not possible to use random assignment to treatment and comparison groups in the case study we address, which evaluated the SEED training programme. We therefore compare two different data analysis methods to adjust for prior underlying differences between groups, namely regression adjustment of treatment covariates that are related to the outcome measure in the sample data and regression adjustment using propensity scores derived from a wide range of baseline variables. The propensity score method allows for control of a wider

range of baseline variables, including those which do not differ significantly between the two groups.

Key words

Probation, compliance, corrections, methodology, propensity scores

Introduction

The aim of this article is to present an account of our experience of using compliance with probation supervision for service users¹ subject to community orders as an interim outcome measure in evaluating a probation initiative: namely, the SEED (Skills for Effective Engagement and Development) training package developed by the National Offender Management Service (NOMS) in England & Wales (Rex & Hosking 2014). The long term aim of the initiative was to have an impact on behaviour in the form of reduced reoffending. However reconviction studies take time and cannot produce any results until a considerable period of time after the pilot period of an initiative has ended. In our evaluation compliance with probation supervision (i.e.

¹ We use the term 'service user' to mean those under supervision by probation staff, whether on licence or on a community order. At the time of the research, both probation officers and probation service officers were supervising service users in the three areas we researched, so we are calling them 'offender managers' or 'probation staff'. We also use 'probation supervision' to mean supervision of offenders subject to a community order, the latter being the only available community sanction in England & Wales since the implementation of the Criminal Justice Act 2003.

Corresponding author:

Angela Sorsby, University of Sheffield, School of Law, Bartolome House, Winter Street S3 7ND, UK; email a.sorsby@sheffield.ac.uk; telephone 0114 2226809.

whether or not the service user complied with the probation conditions or was breached for not complying) was used to provide a more immediate indication of whether the initiative was having an impact on compliance.

In addition to providing a somewhat earlier assessment of impact, determining whether the initiative was having an impact on compliance with the actual supervision was of interest in its own right. There is relatively little academic research on compliance with probation supervision (Ugwudike and Raynor, 2013) and still less on how probation initiatives may affect compliance with supervision. There is a small but growing body of research on what service users most appreciate within probation supervision as well as what aspects of supervision service users and supervisors feel have the greatest impact on compliance. Reviewing the available literature, Shapland et al. (2012) identified that for service users themselves the most valuable aspects of supervision appear to be: developing a relationship with their supervisor; having a supervisor that listens but also tries to steer them towards desistance through motivating them and encouraging them to solve problems; and the provision of practical help and support. Ugwudike (2010), investigating the views of service users and probation officers on the most effective ways of encouraging compliance, emphasises the importance of developing a good positive working relationship and 'the therapeutic and practical benefits of interacting with their supervising officer'

(2010:333) as incentives for compliance. In a study by Hucklesby (2013) investigating compliance with electronically monitored curfew orders, 'offenders reported that their interactions with monitoring officers had impacted on their thoughts about compliance – positively if they felt they had been treated well and negatively if they felt poorly treated' (2013:148).

Relationship building, engagement, active listening and problem solving were key elements of SEED training. On the basis of the above research on what service users find most helpful about supervision and what service users and practitioners report to be important in encouraging compliance, these elements of SEED training would be expected to improve compliance with probation supervision. However, it is also possible, although it was not designed to do so, that SEED training could lead to more intensive supervision, with longer more frequent supervision sessions for example. Furthermore, with its focus on reducing reoffending, elements within SEED, such as challenging pro-criminal attitudes and the use of cognitive behavioural techniques to try to make service users take more responsibility for their offences, could be uncomfortable for service users and could potentially make supervision more demanding. If the supervision is perceived as more intensive and demanding by service users this could potentially have a negative impact on compliance with supervision. As with any evaluation we need to ensure there are no unintended

negative consequences and to establish what impact SEED training actually had on compliance with supervision because probation initiatives should seek to maximise both short term compliance with supervision and longer term desistance from crime (Bottoms, 2001). Improving compliance with supervision is important for a number of reasons which we consider further below. The main focus of this article is methodological. We set out the challenges that we faced in using compliance with supervision as an outcome measure. Because of the lack of previous empirical work on the correlates of compliance (demographics, offence characteristics etc.), taking account of prior underlying differences was challenging. We present two different data analysis methods to adjust for prior underlying differences between groups, namely regression adjustment of treatment covariates that are related to the outcome measure in the sample data and regression adjustment using propensity scores derived from a wide range of baseline variables. We illustrate these two methods with some of the findings from the evaluation.

We shall first address what is compliance with probation supervision and what is known about it, before setting out briefly the nature of the probation initiative and SEED training, then turn to the use of the different methods.

What is compliance and can we measure it?

Compliance, within the context of community sanctions, is a complex dynamic process which has a number of dimensions. A commonly used framework for conceptualising compliance is that proposed by Bottoms (2001) which identifies four main types or mechanisms of compliance namely: instrumental compliance which is based on self-interest; normative compliance which is based on moral obligations; constraint-based compliance involving physical restrictions or requirements; and habit-based compliance which occurs unthinkingly through the development of habits or routines. These mechanisms may operate singly or in combination and the salience of each for any individual may vary over time. Bottoms (2001) also distinguishes between 'short-term requirement compliance' which is 'compliance with the specific legal requirements of the community penalty' (2001:89) and longer-term legal 'compliance with the criminal law' (2001:89) which is essentially desistance from offending. Bottoms argues that those involved with the delivery of community penalties should be trying to maximise both.

Robinson and McNeill (2008), McNeill and Robinson (2013) and Robinson (2013) have further developed theories around short term requirement compliance. Noting that 'it is possible to think about *degrees or dimensions* of short term requirement compliance (Robinson and McNeill, 2008:433) they distinguish between '*formal* and *substantive* compliance'. (2008:434)

‘The former denotes behaviour which technically meets minimum behavioural requirements, such as attending appointments (or work placements) at designated times. The latter implies rather more: namely, the active engagement and cooperation of the offender with the requirements of his or her order and its broader objectives’. (Robinson, 2013:28)

Substantive compliance implies attitudinal acceptance of the community sanction and a willingness to engage with it.

In comparison with substantive compliance, formal compliance is relatively amenable to measurement (Robinson & McNeill 2008). It is possible to assess formal compliance quantitatively using administrative data; although as we shall see, assessing the extent of even formal compliance for a particular group of individuals is not entirely straightforward. Substantive compliance, on the other hand, is concerned with the quality of engagement and is therefore not amenable to quantitative assessment and certainly cannot be assessed by means of administrative data.

The available compliance data

The only compliance data available to us were from administrative data – data which were collected by the probation service for their own case management purposes.

The only reliable data on compliance were whether the order was completed or official

action was taken. This is what Robinson & McNeill (2008) have referred to as *formal compliance*.

More precisely, after consultation with NOMS, it was decided that the only reliable and suitable available data in relation to *community orders* were: whether or not breach proceedings had been initiated, the number of breach proceedings initiated within the order, the length of time between commencement of an order and initiation of breach proceedings, and the outcome classification on termination of the order itself. The last specifies whether the order was completed satisfactorily, or whether the order was revoked due to non-compliance or further offending. Breach proceedings may be initiated if there is reoffending, but also if the conditions of supervision are not fulfilled (e.g. if the person being supervised does not attend appointments). Other potentially available data, such as the number of unacceptable absences during supervision, or what happened to breach proceedings at court, were considered not to be sufficiently reliable since they could be influenced too much by the offender manager's own judgement in relation to discretionary decisions on what is 'acceptable', as well as by court processes which can be affected by what is being done about other offences or sentences.

The SEED initiative

The evaluation was of an initiative known as the Skills for Effective Engagement and Development (SEED) project which was implemented by the National Offender Management Service (NOMS). The aim of the SEED project was to provide training and continuous professional development for probation staff in relation to skills which could be used in supervising offenders, particularly in one-to-one meetings. The SEED training package was influenced by the Strategic Training Initiative in Community Supervision (STICS) project in Canada (Bourgon et al., 2008; 2010) and the aims of the broader NOMS Offender Engagement Programme (Rex 2012). SEED training, like STICS, was designed in accordance with Risk-Need-Responsivity (RNR) principles.

Practitioners were trained in what have been termed Core Correctional Practices, the use of which has been linked to reduced offending (Dowden and Andrews 2004). SEED included additional training for offender managers in relationship building, pro-social modelling, motivational interviewing, risk-need-responsivity, cognitive behavioural techniques and structuring of one-to-one supervision (Rex & Hosking 2014). The aims of SEED and the broader Offender Engagement Programme were to promote more effective engagement to reduce reoffending based on the hypothesis that the relationship between the service user and practitioner can be a powerful means of changing behaviour (Rex, 2012). Part of the impetus for SEED and of the Offender Engagement Programme was a realisation that in the preceding years there had been

too much of an emphasis on outputs and bureaucratic processes at the expense of a focus on engagement, so that many practitioners felt they had not been sufficiently equipped with skills for engagement or supported in using those skills. SEED training consisted of an initial session of three days, followed by three one day and one half-day sessions of follow-up training delivered every three months, so that the total package took just over a year. Between each training session, the SEED-trained groups met in their group to discuss particular cases they had been working on and their manager undertook observation of supervision sessions, feeding back to the offender manager on the supervision. This was therefore very much a group-based activity. Our evaluation of the training package was within three Probation Trusts. Further details of the initiative can be found in Sorsby et al. (2013).

If SEED training were successful in its aims, it would be hoped that those being supervised would be more engaged in their supervision and that supervision would be more tailored to the criminogenic needs and progress towards desistance of the service user. This would suggest potentially increased willingness to stay within supervision by the service user and so increased compliance with the requirements of the order, as well as reduced reoffending.

Why use formal compliance as an outcome measure?

The scope of administrative data in assessing compliance in all its complexity is admittedly limited. Administrative data provide information about whether the service user has attended and completed supervision (formal compliance) not whether that person has actively engaged with supervision (substantive compliance); while it is the latter form of compliance (which is not amenable to measurement) that is seen as key to promoting longer term desistance from offending (Robinson and McNeill, 2008). It is indeed substantive compliance which SEED training seeks to promote through its focus on the supervisory relationship and offender engagement.

However, while the quality of engagement is considered key in promoting desistance and lies at the core of SEED training, quantitatively measureable compliance in the form of attending supervision is also important, as without attendance it would be difficult to secure active engagement, or for the service user to learn from SEED-inspired work, or access help in solving their problems. Formal compliance may not be a sufficient condition for substantive compliance but formal compliance may provide a foundation for the development of substantive compliance (Robinson and McNeill, 2008). We can draw an analogy with attendance in the field of education; attendance at school may not be sufficient to ensure learning but failure to attend is highly likely to lead to a poor educational outcome. If one believes supervision can be helpful in

aiding desistance in service users, managing to keep those people in supervision, without breaching them or recalling them to prison, should aid desistance.

There is a paucity of empirical research specifically examining the connection between short-term formal compliance with orders and long term desistance from crime. The available literature on 'What Works' in relation to offender supervision suggests that those who fail to complete programmes and other interventions do worse in terms of reconvictions than those who complete the programme and also worse than those who do not commence the programme or are assigned to comparison groups (see Harper and Chitty, 2005). This, as Hucklesby (2013:140) points out, supports 'a common-sense notion that offenders who comply with the requirements of the order are more likely to be the ones who will desist in the future'. However, as Hucklesby also notes this does not mean that there is a causal link between the two; nor that compliance is a sufficient or necessary condition for desistance.

If an intervention, such as SEED training which seeks to improve the quality of probation supervision, were to have a positive impact on attendance and completion it seems reasonable to posit that a likely mechanism through which this effect could be achieved would be through increasing the service user's motivation to comply. It seems reasonable to assume that improvement in attendance and completion may be

a consequence of service users being more willing to comply, because they feel they are gaining something from supervision; within the context of SEED, a likely benefit includes helping offenders to solve their own problems. In-depth interviews with service users in a study by Ugwudike (2010) support this view. Ugwudike found that the therapeutic benefits of interacting and discussing problems with a probation supervisor were frequently cited as an incentive that can motivate instrumental compliance. Hence, although all we can measure is the effect of the intervention on formal compliance, a likely mechanism underlying any effect is instrumental: i.e. that the service users are buying into the supervision more because they can see its benefits. If improved compliance with supervision is achieved through increased motivation to comply this is more likely to develop into normative compliance and potentially desistance than for example achieving compliance for (negative) instrumental reasons through stricter enforcement (see Robinson and McNeill 2008: 439). Furthermore, with its focus on relationship building, SEED would seem likely to foster a sense of obligation to comply through attachment to the supervisor and perceived fair treatment, also cultivating normative compliance (see Ugwudike 2010:338).

It should however be noted that there is an alternative mechanism through which an apparent improvement in attendance and completion could be achieved. If SEED

training gives offender managers increased faith in probation supervision as a means of effecting change, they could potentially be less likely to 'give up' on those who are finding it more difficult to adjust their lifestyles to a structured regime of supervisions, and so be also less likely to breach or recall more 'difficult' service users. This is a limitation of using compliance data to assess the impact of an initiative. Unlike reconviction, offender managers are involved in breach and recall decisions and, thus, in the construction of compliance (Robinson, 2013). For this and other reasons, although in the short term assessing the impact of an initiative on formal compliance is of interest, it has to be considered supplementary to rather than a substitute for a longer term reconviction study. Amongst other things, only a reconviction study can indicate whether an initiative has had an impact on longer term desistance from crime.

Improving completion rates for community supervision without recourse to enforcement action could also be considered to be a worthwhile aim for other reasons. Imprisonment for failing to comply with a community order adds to the already high prison population (Ministry of Justice, 2010). Tough enforcement strategies, intended to increase the perceived credibility of the probation service and community sanctions, have resulted in high breach rates which have actually tended to damage the credibility of community sanctions in the eyes of the courts and the general public (Robinson and Ugwudike, 2012; Ugwudike and Raynor, 2013a). Extreme tolerance in

relation to enforcement may also damage the credibility of the probation service and community sanctions (Robinson, 2013). Measures which improve compliance, through better engagement with those being supervised, without recourse to enforcement could potentially improve the credibility of probation and community sanctions in the perceptions of the courts and general public, thereby helping to make community supervision more attractive to the courts with the potential to reduce prison numbers. Attempting to secure compliance through tough enforcement policies may also be perceived as unfair by service users. A number of studies have indicated that unfair treatment may have a negative effect on compliance, due to a perceived lack of legitimacy, potentially resulting in reduced feelings of obligation to comply (Bottoms, 2001; McNeill and Robinson, 2013). Improved compliance through increased engagement on the other hand would seem more likely to translate into increased feelings of obligation to comply and potential longer term normative compliance.

A further reason for using compliance at least as an intermediate outcome measure in evaluations, is the very practical reason that it has the potential to provide a more immediate, if only partial, indication of effectiveness compared to a reconviction study, providing at least some indication of whether an initiative is having an effect while the resources for delivering it are still in place. Although a reconviction study provides a much better measure of longer term compliance in the form of desistance,

reconviction studies inevitably take time and cannot produce any results until a considerable period of time after the pilot period of an initiative has ended. Over time, in the absence of any indication about whether or not an initiative may be effective, there may be a loss of momentum and the resources for delivering the initiative, such as the availability of trainers, may no longer be in place.

Finally, a study of the impact of an initiative on short-term formal compliance, combined with a longer term reconviction study can provide much needed information on the relationship between short-term formal compliance and long term desistance from crime.

Issues in relation to compliance data

Those seeking to study compliance need, however, to be aware of some issues surrounding the availability of data and their nature. These may provide some explanation for the lack of research on compliance – and may assist future researchers. At the outset of the SEED project and its evaluation the evaluators were promised that compliance data would be obtained by querying one uniform database covering all the Probation Trusts, which would be installed and made operational before we needed to collect compliance data. However this did not occur within the required time. Data

therefore needed to be obtained from the IT departments of the three separate Trusts using three separate computer systems.

The databases, like most criminal justice databases, were designed for case management purposes and for providing the statistics used by the Ministry of Justice. They were not designed for research purposes and they were not designed to easily provide data on specific groups of people, supervised by specific offender managers, such as those who had undergone SEED training, for a specific time period, such as those commencing supervision during the year following SEED training.

Unfortunately, although the case log on the computer systems maintained a permanent record of all that had happened in the case, fields upon which one could base a query, such as the offender manager's name and the date a person was released from prison, were updated and overwritten whenever a change took place. Hence, if someone was recalled to prison, the date on which they were originally released prior to recall was removed and left blank until the person was released again, when the new date was inserted. Therefore, if one queries the databases to extract people sentenced to community orders or released from prison and commencing supervision with named offender managers between certain dates, people for whom

any of the relevant information had subsequently been amended (e.g. recalled to prison) would not be extracted.

It was therefore necessary to obtain regular data extracts and build up a list of service users who should be included in the project, together with the relevant data on them from these regular extracts, as opposed to obtaining these data at the end of the project, by which time vital information, such as the date on which a person was released from prison on licence would have been lost if that person had been recalled to prison. As is often the case with criminal justice research, the queries used to extract data also extracted some cases outside the parameters of the study because they were sentenced or released outside the time frame, were still in custody, involved no supervision with an offender manager or were with offender managers outside the project. Hence the evaluators needed to carefully examine each case to make sure the person should be included in the analysis.

Some service users appeared in the data more than once with different sentence or release dates. In such cases we included only the case with the earliest sentence or release date and excluded the others. This was necessary so that we did not violate the assumption of independence made by most statistical tests.

Issues in relation to the evaluation of impact

Evaluation of the impact of an initiative requires comparison of those who have taken part in the initiative (a treatment group) with a comparison group of people who have not taken part in the initiative. In an ideal world participants would be randomly assigned to treatment and comparison groups so that, on balance, those in the treatment group do not differ in any systematic way from those in the comparison group.

In evaluating SEED random assignment was not possible, because it involved training teams of offender managers together (and with their managers). In order for random assignment to ensure there were no systematic differences between the teams, random assignment of probation teams to treatment or comparison groups would require an excessive number of probation teams to be involved in the study. A substantial element of SEED training involved colleagues learning from one another, particularly in peer group learning sessions after training where teams meet up to discuss a case, but also in the course of the training. In addition, having trained and comparison offender managers in the same probation office, may lead to contamination of the comparison group as SEED trained offender managers would be likely to discuss the training, pass on practice recommendations from the training and

even pass on exercises that can be used with service users which formed part of the training to colleagues that had not taken part in the training. Random assignment of service users to treatment and comparison offender managers, based in different offices, was not practical for a variety of reasons including that it would be unethical to expect service users to attend an office in a different geographical location. Matching participants was also not possible since the trained offender managers needed to be based in the same team.

Each of the three probation Trusts had two SEED trained teams and the study also included one or two comparison teams in each Trust. Practical issues for NOMS dictated that the Trusts involved in the initiative were those that had volunteered to take part. In order to ensure a sufficient number of service users for analysis purposes, the Trusts used in the external evaluation were the three largest of the Trusts that volunteered. Practical issues also dictated that the Trusts chose which were to be the two trained teams and also the comparison teams, which to minimise contamination were located some distance from the trained teams. Hence neither NOMS nor the evaluators were in charge of the selection, making this similar to a prospective longitudinal natural experiment, because the service users to be supervised were those sentenced by the courts and released from prison during that period, hence service users did not self-select.

Service users included in the evaluation in the 'SEED trained' group were those who *commenced* one-to-one supervision in the community (i.e. not in custody) on a community order within a one year period of the offender managers in the trained group completing their initial three days SEED training, provided the order was not terminated within less than one month.² The equivalent service users in the 'comparison' groups were those commencing their supervision within a one year period after the relevant SEED trained group had done their initial training. We obtained compliance data for 1,161 service users on community orders³.

As there could be no matching or random assignment of participants we could not assume that the service users supervised by SEED trained and comparison offender managers were similar on background variables. We therefore first needed to assess just how comparable the two groups were.

In order to achieve this we obtained Offender Assessment System (OASys) data. OASys is the risk assessment system used by NOMS, the probation service and the prison

² Orders terminated within one month tended to involve service users who never commenced supervision (for example, because they were in prison for another offence, or never turned up to start the order).

³ We also obtained data on 325 service users being supervised on licence, after having been released from prison. As, however, these numbers were too small to use in the analyses reported in this article, they are not referred to further.

service.⁴ Further details about OASys can be found in Moore (2015). We compared the trained and comparison groups on a large number of OASys variables obtained from initial assessments of service users at the start of the order. The OASys variables included were those which were likely to reflect some external evidence rather than being based too heavily on offender managers' opinions. The analysed data included variables on: age, gender, ethnicity, offence type, sentence type, offence motivation, number of previous convictions, accommodation, employment and work skills, domestic violence, relationship status, lifestyle, substance misuse, emotional well-being, thinking and behaviour, predicted likelihood of reoffending, assessed risk of harm and criminogenic needs.

In fact, the SEED trained groups proved to have significantly different scores to the comparison group on a number of variables. For community orders, they were significantly more likely to score highly on the OASys Violence Predictor (OVP), to have committed domestic violence, to be assessed as posing a greater risk of causing serious harm to children in the community, to have relationships as a criminogenic need, to have had police contact at an earlier age, to be at risk of self-harm, to require a specialist report, to have a higher score on thinking and behaviour as a criminogenic

⁴ OASys variables include offender manager views on the criminogenic needs of the service user, as well as offence, criminal record and demographic variables.

need, and to have problems in understanding other people's views. One could summarise this as service users in the trained group being more risky in terms of causing harm and in terms of likelihood of reoffending for violent crime, but not in showing a significant property crime profile. They might be seen as being a 'harder' group to supervise in terms of risk of harm, but not necessarily in terms of reoffending (since the comparison groups tended to be higher in terms of property crime and drug use). Note that these differences could not have been foreseen before the evaluation began, because the service users entering supervision were not allocated by NOMS, the Trusts or the evaluators.

Research analysis strategies to take account of group differences

Variables on which there are differences between service users supervised by the trained and comparison offender managers are potential confounders. A variable should be considered a confounder if it differs between the two groups and is also associated with the outcome measure (compliance with supervision): in other words if it is a risk factor.

There is a lack of prior empirical research on which factors relate to supervision compliance, making it impossible to specifically identify known risk factors on the basis of prior research. The literature on compliance has concentrated on what regime

changes (such as National Standards) do in relation to the overall numbers of service users complying (e.g. Robinson and McNeill, 2008; Hedderman and Hough, 2004), not on how different kinds of service users or those convicted for different offences perform in relation to compliance with supervision. There is substantial literature on which background factors affect reconviction (see Brunton-Smith and Hopkins, 2014) but it is important to note that there can be no presumption that factors related to supervision compliance are the same as those for reconviction.

One of the very few studies of compliance with probation supervision in England and Wales is that by Gyateng et al. (2010) conducted in London. In that study the most important factors predicting breach were the number of requirements on the current sentence and being sentenced to a drug-related requirement. Unfortunately, we did not have data on these elements for our Probation Trusts. Other significant variables in the Gyateng et al. (2010) research were age; having a drug need; a previous history of breach; the borough in which an offender was supervised; and the length and type of disposal. We did have data on the first two of these and they were included in our analysis. Unfortunately, history of breach, although included in the OASys data that we received, consisted mostly of missing values and length of disposal was not available to us. The specific geographic area in our study was of course conflated with

the selection of areas for training. Type of disposal was not relevant as we were not looking at different disposals.

Confounding variables, variables on which there is a difference between the treatment and comparison groups and which are *also* related to compliance must be taken account of in conducting the analysis. Any variable on which there is a significant difference between the trained and comparison groups and which itself shows a relationship with the compliance measures could in fact be responsible for any observed difference between the groups or alternatively may mask any effect of treatment.

An issue arises around identifying what should be considered confounding variables. If there is a substantial body of prior research this can assist in identifying risk factors but in the absence of such research there could be a large number of potential risk factors. In such a situation the first step is to assess the comparability of the groups across a broad range of potential risk factors, which is why we assessed comparability across a wide range of available OASys data. The next step is to consider whether the variables on which there is a difference might be related to the outcome measure. The issue is how much of a difference does there need to be between the two groups on a variable

and how strong does the relationship between that variable and the outcome have to be for the variable to be considered to be having a confounding effect?

A traditional approach is to compare treatment and comparison groups on each of the baseline variables for which data are available using an appropriate statistical test (e.g. t-test, Mann-Whitney or chi-square) and consider as potential confounders all those variables on which there is a statistically significant difference ($p < 0.05$), then using statistical tests identify on the basis of the sample data which of these variables is related to the outcome measure. Any variable that is significantly related to both treatment and the outcome is considered a confounding variable, the effect of which needs to be adjusted for. Adjustment may be done by entering these variables first into a regression analysis and then entering group membership (treatment versus comparison). Provided none of the confounding variables are collinear with the provision of treatment, correlations (multicollinearity) between the confounding variables are not a problem as one is making no causal assertions about the confounding variables. Potential criticisms of the above method are that differences between groups on the baseline characteristics which are not statistically significant could still have an effect on the outcome measure, in addition differences on combinations of variables may be important.

An alternative procedure which avoids these criticisms is to calculate propensity scores and then use these in regression adjustment. The propensity score was defined by Rosenbaum and Rubin (1983: 41) as 'the conditional probability of assignment to a particular treatment given a vector of observed covariates'. Propensity scores are usually estimated by obtaining predicted probabilities from a logistic regression model with treatment status (treatment versus comparison) as the outcome measure and the observed baseline characteristics as predictors. Essentially, the propensity score reduces the background characteristics to a single dimension and those with similar scores should have similar distributions on the covariates used to calculate the score. The various uses of propensity score analysis are discussed in Austin (2011) and D'Agostino (1998).

An advantage of first calculating the propensity score and then using this in regression is that one can use a large number of variables in calculating the propensity score without being 'concerned with over-parametrizing' (D'Agostino, 1998: 2277). The real advantage of this method is that given a sufficient sample size all possible variables can be used, so differences between the groups that are not statistically significant and differences on combinations of variables are taken care of. Multicollinearity is not considered a problem in calculating propensity scores. General advice is that it is better to tend towards being over-inclusive, rather than risk leaving out a confounding

variable, and to include any covariate potentially related to the outcome or treatment (Faries et. al, 2010: 25). In short, if data are available, one needs to be convinced that a variable is not relevant in order to justify leaving it out. A disadvantage of including a large number of background variables is that if data are missing on any of the variables for a participant then the participant is lost from the analysis.

Once obtained the propensity score can be entered into regression analyses either as a raw score or stratified into categories (D'Agostino, 1998). As generally there is no reason to assume that the propensity score, which reduces the background characteristics to a single dimension, will be in a linear relationship with the outcome measure, it is usually more appropriate to divide the propensity score into categories and include the propensity score as a categorical variable (Pasta, 2000). People have frequently used quintiles in stratifying propensity scores but, with larger samples, it can be useful to have more categories (Faries et. al, 2010: 26). A subset of the variables used in calculating the propensity score can also be included in the regression (D'Agostino, 1998: 2277) to control for any residual imbalances between the treatment and comparison groups after adjusting for the propensity score.

Implementing the two strategies

We analysed compliance with community sanctions using both of the above strategies. The data were analysed using SPSS version 21. We carried out regression analyses in which we first entered each of the variables which showed a significant difference between the trained and comparison group (on bivariate tests) and also showed a significant relationship with the outcome measure (on bivariate tests) followed by the group membership (treatment versus comparison) variable. Hence in these analyses we included only variables on which there was a significant difference between the trained and comparison group and only variables that were significantly related within the sample data to the various compliance measures.

In using the second strategy propensity scores were estimated using logistic regression with treatment status (trained versus comparison) as the outcome and the demographic and OASys measures as the predictors. As any OASys variable could reasonably be expected to be related to compliance and, furthermore as in this part of the analysis we wished to take account of non-significant differences between groups and non-significant relationships with the outcome, as well as possible differences on combinations of variables which might affect the outcome, we included all demographic and OASys variables that were available to us which did not have an

excessive amount of missing data. The only potentially relevant variables that had to be excluded on this ground were breach history and re-sentencing for breach, where we had substantial amounts of missing data.

As outlined above the propensity score reduces the background characteristics to a single dimension and those with similar scores should have similar distributions on the covariates used to calculate the score. As relationships between the propensity score and the compliance measures did not appear to be linear, it was considered more appropriate to divide the score into groups and include it as a categorical variable. We divided the scores into ten strata as this seemed to achieve a better within strata balance on the background variables than five, although there was still some residual imbalance within some of the strata.

Results using the two strategies

Table 1 below shows the proportions of successful and unsuccessful terminations for cases supervised by SEED trained and non-SEED trained offender managers for each of the Trusts. Cases are only included if they were terminated during the timescale of the study ⁵ and the analysis excludes cases where the termination outcome was classified

⁵ Identical proportions of cases were terminated for both groups (89%) so, although the analysis excludes longer orders, particularly if they started late in the sampling period, as identical proportions were excluded for both groups this is unlikely to create any bias in terms of comparing the groups.

as neutral. Neutral termination outcomes are cases that have been transferred outside the relevant Probation Trust, revoked on application⁶ or terminated for other reasons.⁷ The termination reasons for unsuccessful terminations were ‘Revoked – failure to comply’, ‘Revoked – further offence’ or ‘Expired – breach listed’.

Table 1 Termination classification for terminated cases excluding neutral outcomes

Termination classification	Trust A		Trust B		Trust C		Overall	
	Comparison n=114	Trained n=198	Comparison n=176	Trained n=274	Comparison n=60	Trained n=109	Comparison n=350	Trained n=581
Successful	70.2%	72.2%	69.9%	72.3%	83.3%	86.2%	72.3%	74.9%
Unsuccessful	29.8%	27.8%	30.1%	27.7%	16.7%	13.8%	27.7%	25.1%

(a) Using regression analysis and controlling for confounding variables

Implementing the first strategy outlined above each of the variables which showed a significant difference between the trained and comparison groups and which also showed a significant relationship with the termination classification were entered in the first block of a logistic regression analysis. Group membership (trained or comparison) was entered as the second block. There was a significant relationship between termination classification and group membership ($p=0.014$) – and therefore a main effect of the SEED training on this measure. After controlling for differences

⁶ Cases revoked because there has been a further offence or for failure to comply are all classified as unsuccessful.

⁷ Not specified but only 13 cases in total.

between the two groups on the various OASys measures, the odds of a successful outcome for service users in the trained group were 1.6 times the odds of a successful outcome for service users in the comparison group. Detailed results from the regression analysis are provided in Table 2.

Table 2 Results of logistic regression employing the first strategy in relation to termination classification

	B	S.E.	Sig.	Exp(B)
Offence category			0.002	
Above or below cut-off point where relationships are considered to be a criminogenic need	0.326	0.206	0.113	1.386
Understands other people's views	0.526	0.211	0.013	1.693
Highest risk of serious harm in community across all categories	-0.311	0.204	0.127	0.733
Risk of serious harm to children in the community	-0.054	0.205	0.793	0.948
Age first police contact	0.017	0.014	0.244	1.017
Thinking and behaviour as a criminogenic need score	-0.376	0.072	0.000	0.686
OVP score	-0.038	0.007	0.000	0.963
Trained versus comparison group	0.472	0.191	0.014	1.603
Constant	3.411	0.561	0.000	30.291

$N=813$, $R^2=0.22$ (Cox & Snell), 0.321(Nagelkerke). Model $\chi^2(18)=204.07$, $p<0.001$.

In this and subsequent analyses the outcome measure is the termination classification for terminated cases (excluding neutral outcomes) and is coded 0=Unsuccessful 1=Successful
Trained versus comparison group is coded 0=Comparison 1=Trained. The reference category is the comparison group.

However, 'unsuccessful' termination included both those who reoffended and those who did not satisfactorily keep to their probation conditions. SEED training might be

predicted to affect both these aspects in different ways, so it is worth running the analyses separately for each group. This found that if we include in the analysis only cases with successful termination classifications and those that have been unsuccessful because of lack of compliance with probation conditions (rather than due to reoffending), there is a significant relationship between the outcome measure and group membership ($p=0.026$) –and so a main effect of SEED training on this measure.⁸ After controlling for differences between the two groups the odds of a successful outcome for service users in the trained group (i.e successful completion without the order being revoked for non-compliance with probation conditions (as opposed to reoffending)) were 1.7 times the odds of a successful outcome for those in the comparison group.

If we include in the analysis only cases with successful termination classifications and those that were unsuccessful because of further offending, there was also a significant relationship between the outcome measure and group membership ($p=0.023$), and so

⁸ The variables which were significantly related to termination in terms of lack of compliance with probation conditions were Offence category; Relationships as a criminogenic need score; Understands other people's views; Highest risk of serious harm in community across all categories; Risk of serious harm to children in the community; Age first police contact; Thinking and behaviour as a criminogenic need score; OVP score. When these variables were entered into a logistic regression, along with group membership (trained versus comparison) $N=709$, $R^2=0.163$ (Cox & Snell), 0.274(Nagelkerke). Model $\chi^2(18)=126.495$, $p<0.001$.

also a main effect of the SEED training on this measure.⁹ After controlling for differences between the two groups the odds of a successful outcome for service users in the trained group as compared to the order being revoked for further offending were 1.8 times the odds of a successful outcome for those in the comparison group.

(b) Using propensity score analysis

Our second strategy was to use the propensity score method. The results of the logistic regression analysis with termination classification (successful or unsuccessful) as the outcome measure with the propensity score (in ten strata) and trained versus comparison as predictors indicate a significant difference between the groups ($p=0.045$). The odds ratio is 1.5. If one incorporates the known confounders (variables that differed significantly between the trained and comparison group and are significantly related to the outcome) into the regression analysis, to control for residual within strata imbalances on these variables, the difference between the groups is significant at $p=0.006$. The odds ratio is 1.9. Details of the logistic regressions can be found in Table 3.

⁹ The variables which were significantly related to termination in terms of lack of compliance with probation conditions were Offence Category; Relationship status; Understands other people's views; Highest risk of serious harm in community across all categories; Age first police contact; Thinking and behaviour as a criminogenic need score; OVP score. When these variables were entered into a logistic regression, along with group membership (trained versus comparison) $N=678$, $R^2=0.172$ (Cox & Snell), 0.310 (Nagelkerke). Model $\chi^2(18)=128.182$, $p<0.001$.

Table 3 Results of logistic regression employing the second strategy in relation to termination classification with propensity score in 10 strata alone

	B	S.E.	Sig.	Exp(B)
Propensity score in 10 strata			0.225	
Trained versus comparison group	0.379	0.189	0.045	1.461
Constant	0.908	0.253	0.000	2.479

N=727, R²=0.19 (Cox & Snell), 0.27 (Nagelkerke). Model $\chi^2(10)=13.690$, p=0.188.

Using regression adjustment to reduce residual imbalance on variables where there was a significant difference between trained and comparison groups

	B	S.E.	Sig.	Exp(B)
Propensity score in 10 strata			0.776	
Offence category			0.006	
Above or below cut-off point where relationships are considered to be a criminogenic need	0.375	0.221	0.090	1.456
Understands other people's views	0.685	0.238	0.004	1.984
Highest risk of serious harm in community across all categories	-0.394	0.226	0.081	0.674
Risk of serious harm to children in the community	0.107	0.234	0.647	1.113
Age first police contact	0.017	0.015	0.256	1.018
Thinking and behaviour as a criminogenic need score	-0.438	0.081	0.000	0.645
OVP score	-0.035	0.008	0.000	0.965
Trained versus comparison	0.620	0.224	0.006	1.860
Constant	3.598	0.706	0.000	36.543

N=727, R²=0.23 (Cox & Snell), 0.34 (Nagelkerke). Model $\chi^2(27)=193.175$, p<0.001.

Again it is worth considering separately those who had unsuccessful outcomes because they did not keep to their probation conditions and those who had unsuccessful outcomes because they reoffended. If we include in the analysis only cases with successful termination classifications and those that have been unsuccessful because

of lack of compliance with probation conditions the results of the logistic regression analysis with only the propensity score (in ten strata) and trained versus comparison as predictors did not indicate a significant difference between the groups ($p=0.144$). If one incorporates the known confounders (variables that differed significantly between the trained and comparison group and are significantly related to the outcome) into the regression analysis, to control for residual within strata imbalances on these variables, the difference between the groups is significant at $p=0.023$. The odds ratio is 1.9.¹⁰ . Almost identical results were obtained whether one used the propensity score categorised into five or ten strata in the regression analyses, or if the propensity score was entered into the regressions as a continuous variable.

If we include in the analysis only cases with successful termination classifications and those that have been unsuccessful because of further offending with only the propensity score (in ten strata) and trained versus comparison as predictors, this indicated a significant difference between the groups ($p=0.045$), with an odds ratio of 1.7. If one incorporates the known confounders (variables that differed significantly

¹⁰ The variables which were significantly related to termination in terms of lack of compliance with probation conditions were Offence category; Relationships as a criminogenic need score; Understands other people's views; Highest risk of serious harm in community across all categories; Risk of serious harm to children in the community; Age first police contact; Thinking and behaviour as a criminogenic need score; OVP score. When these variables were entered into a logistic regression, along with the propensity score in ten strata and group membership (trained versus comparison $N=641$, $R^2=0.180$ (Cox & Snell), 0.303 (Nagelkerke). Model $\chi^2(27)=127.380$, $p<0.001$.

between the trained and comparison group and are significantly related to the outcome) into the regression analysis, to control for residual within strata imbalances on these variables, the difference between the groups is significant at $p=0.006$. The odds ratio is 2.4.¹¹ Similar results were obtained whether one used the propensity score categorised into five or ten strata in the regression analyses, or if the propensity score was entered into the regressions as a continuous variable.

The above is about whether supervision resulted in a successful or unsuccessful outcome. We also had data on whether breach proceedings were initiated for the two groups and the numbers of breach proceedings initiated. Using either of the two analysis strategies there was no significant difference between the trained and comparison groups in terms of whether or not breach proceedings were initiated. Cox regression was used to account for individual participants having different periods of time at risk due to differences in the length of orders. There was also no significant difference between the two groups in the number of breach proceedings initiated using either of the above two strategies.

¹¹ The variables which were significantly related to termination in terms of lack of compliance with probation conditions were Offence category; Relationship status; Understands other people's views; Highest risk of serious harm in community across all categories; Age first police contact; Thinking and behaviour as a criminogenic need score; OVP score. When these variables were entered into a logistic regression, along with the propensity score in ten strata and group membership (trained versus comparison $N=614$, $R^2=0.164$ (Cox & Snell), 0.304 (Nagelkerke). Model $\chi^2(27)=109.934$, $p<0.001$.

In relation to the propensity score strategy on all three measures (successful/unsuccessful outcome, whether breach initiated, number of breaches initiated), almost identical results were obtained whether one used the propensity score categorised into five or ten strata in the regression analyses, or if the propensity score was entered into the regressions as a continuous variable.

In essence, therefore, the results from the propensity score method of analysis are identical in terms of what is and is not significant to those achieved by regression analysis controlling for significant differences between the groups on variables significantly related to compliance. Propensity score analysis additionally compensates for any non-significant differences in background variables between the trained and comparison groups and non-significant relationships with compliance, as described above.

Discussion

In this article we have provided an account of our experience of using compliance with probation supervision as an interim outcome measure in evaluating a probation initiative.

Compliance was partly used as an interim outcome measure to provide a somewhat more immediate measure of impact, as compared to reconviction. Determining whether the initiative was having an impact on compliance with the actual supervision was also of interest in its own right. However, one limitation of using compliance data to assess impact is that, unlike reconviction, offender managers are involved in breach and recall decisions. It is possible that rather than making service users more likely to comply with supervision, an initiative such as SEED training may just relate to more appropriately taken enforcement proceedings. The analyses we performed in which we considered separately unsuccessful terminations because of further offending and unsuccessful terminations because of failure to comply with probation conditions does provide some evidence in relation to this. SEED appeared to have an impact on further offending during the order, as well as compliance with probation conditions and seemed to have a greater impact on the former than the latter. Taking enforcement action for reoffending is outside the offender manager's control so it seems unlikely that the effect is entirely due to more appropriately taken enforcement proceedings. That said, although in the short term assessing the impact of an initiative on formal compliance is important, it should be considered supplementary to rather than as a substitute for a longer term reconviction study. Amongst other things, only a

reconviction study can indicate whether an initiative has had an impact on longer term desistance from crime.

The evaluation of SEED could not be set up to be absolutely ideal in terms of research design. There could be no random assignment or matching. We therefore used two different data analysis methods to adjust for prior underlying differences between groups: regression adjustment of treatment covariates that are also related to the outcome measure in the sample data and regression adjustment using propensity scores that have been derived from a wide range of baseline variables. On the basis of our experience, given a sufficient sample size, we would suggest using propensity scores combined with further regression adjustment of known confounders is the most helpful technique because it controls for non-significant differences between the groups on background variables, though in our case both gave almost identical results.

Substantively, we found some small but statistically significant positive effects on compliance. SEED training was related to whether supervision terminated successfully or unsuccessfully, when the background factors differentiating between the groups were controlled. This was the case for both of the analysis methods. It seemed to prevent both reoffending during the supervision period *and* breaches of probation conditions. In other words, it seemed to be having some effect on desistance, at least

immediately after the offence, as well as on formal compliance. There were, however, no significant differences in relation to other measures of compliance with community sentences, such as the decision to initiate breach proceedings. This may be because initiation of proceedings does not imply breach is then found to be proved at court – acceptable reasons for not meeting the probation condition may be discovered or the court may dismiss the breach proceedings, or simply put the person back on supervision.

The impact of SEED training on formal compliance was relatively small. There may be a variety of reasons why the impact was not greater. The long term aim of SEED was to promote desistance. Compliance with supervision arguably requires somewhat more than desistance from crime: it requires attendance at appointments. A broad range of individual and social factors may impact on supervision attendance and hence compliance with community orders. Ugwudike (2010) outlines a number of obstacles to formal compliance with orders. In addition to factors which would be considered potential obstacles to both desistance and supervision compliance, such as substance misuse and unemployment, obstacles to formal compliance also include practical issues such as childcare problems and transportation costs which have nothing to do with criminality. It is conceivable that even where service users have a good relationship with their supervisor, feel they are gaining something from supervision

and are committed to desisting from crime, they may still struggle to comply with the order itself due to practical difficulties. The quality of probation supervision is just one of a number of factors that may impact on compliance. In addition, although SEED training includes problem solving, probation practitioners cannot possibly be expected to ameliorate all the practical problems that service users may face. Furthermore, SEED training emphasises assisting service users to solve their own problems rather than the Offender Manager solving the problems for them. This emphasis on service users taking responsibility is important in the long term, particularly from the point of view of desistance, but may provide less immediate solutions to problems.

SEED training is designed in accordance with RNR principles. RNR specifically targets or prioritises criminogenic needs to reduce recidivism (Andrews et al. 1990; Andrews and Bonta 2010). RNR specifically does not prioritise non-criminogenic needs that are only weakly related to recidivism (Andrews and Bonta 2010). As obstacles to compliance with probation supervision may potentially be broader than obstacles to desistance, as we outline above, it is possible that non-criminogenic needs, although only weakly related to recidivism, may relate to a somewhat greater extent to compliance with supervision itself. Non-criminogenic needs and non-criminogenic practical obstacles to compliance with the order, such as childcare and transportation issues, do not specifically form part of SEED. This may be part of the reason why SEED training

appears to have had a greater impact on offending during the order than it did on failure to comply with probation conditions. In addition, our use of OASys variables to control for differences between the trained and comparison groups on these variables does not control for potential differences between the two groups in relation to non-criminogenic needs or for changes in criminogenic needs during the course of supervision.

Ugwudike (2010) found that some probation practitioners use various strategies to manage practical obstacles to compliance (such as offering flexible appointments, reminding service users to attend appointments and making home visits to assist with childcare and transport problems); other practitioners did not use these strategies, because they considered compliance with the order to be the service user's responsibility. SEED training did not specifically address the use of such strategies to facilitate compliance, although they did come up during discussions in the course of training (which was observed by the evaluators). On the one hand, the focus on relationship building in SEED training may have made practitioners more likely to use such strategies but, on the other hand, the focus in SEED training on the service user taking responsibility may have made practitioners less willing to use these strategies, thereby potentially reducing any impact on compliance with the order itself. An additional point in relation to this issue is that discussions which took place during

training indicated that some of these strategies, such as the use of home visits, may be difficult in practice due to time and resource limitations. At a more general level, the potential impact of SEED training on compliance may have been lessened by the obvious time pressures that practitioners were under, something which came out in our observations of training and also in questionnaires which were administered to practitioners at each training event.

Additional reasons why it may be unrealistic to expect SEED training to have a dramatic impact on compliance are related to the actual difference in supervision practice that training might be expected to create between the trained and comparison groups. The range of skills used by practitioners in one-to-one supervision has been studied by Raynor et al. (see for example Raynor et al. 2014) using videotapes of supervision sessions conducted by probation practitioners in Jersey. In that research there was a particular focus on the use of Core Correctional Practices, the skills which are also the focus of SEED training. Raynor et al. found considerable variation between practitioners in their use of these skills. We would therefore expect there to be similar pre-training differences between practitioners in our study in the use of skills which are the focus of SEED. There may also have been pre-training differences between the SEED and comparison practitioners in the extent to which they were already using these skills. In addition to pre-existing differences between practitioners in their use

of SEED skills there may also be differences between practitioners in the impact of SEED training and the extent to which they feel able to actually implement the skills in practice. Those who were making less use of these skills before training may have found it difficult to immediately adapt their supervision practice, especially in the face of limited time for planning and supervisions.

In view of the number of factors that may impinge on practitioners' use of the skills addressed by SEED training and in view of the number of factors that may affect service users' compliance with supervision, it would probably be unrealistic to expect SEED to have a dramatic impact on compliance. The small but significant impact which it did have is however a step in the right direction.

Funding

The research was supported by the National Offender Management Service (NOMS) under its Offender Engagement Programme. Any opinions, findings and conclusions or recommendations in the article are those of the authors, not necessarily shared by NOMS or the Ministry of Justice.

References

Andrews, D.A., Bonta, J. and Hoge, R.D. (1990) Classification for Effective Rehabilitation: Rediscovering Psychology, *Criminal Justice and Behaviour* 17(1): 19-52.

Andrews, D.A. and Bonta, J. (2010) *Psychology of Criminal Conduct*. Fifth Edition. Newark, NJ:LexisNexis.

Austin, P. (2011) An introduction to propensity score methods for reducing the effects of confounding in observational studies, *Multivariate Behavioral Research* 46(3): 399-424.

Bottoms, A.E. (2001) Compliance with Community Penalties. In: A. Bottoms, L. Gelsthorpe and S. Rex (eds) *Community Penalties: Change and Challenges*. Cullompton: Willan.

Bourgon, G., Rugge, T., Guiterrez, L., Simpson, K., Bonta, J., Scott, T., Yessine, A., Li, J. and Helmus, L. (2008) *Strategic Training Initiative in Community Supervision (STICS)* Presentation to the Canadian Psychology Association's 69th Annual Convention, Halifax, Nova Scotia.

Bourgon, G., Bonta, J., Rugge, T., Scott, T. and Yessine, A. (2010) The Role of Program design, Implementation, and Evaluation in Evidence-Based “Real World” Community Supervision’. *Federal Probation* 74: 2-15.

Brunton-Smith, I. and Hopkins, K. (2014) *The factors associated with proven re-offending following release from prison: findings from Waves 1 to 3 of SPCR – Results from the Surveying Prisoner Crime Reduction (SPCR) longitudinal cohort of prisoners*. London: Ministry of Justice Analytical Series.

D’Agostino, R.B. (1998) Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine* 17: 2265-2281.

Dowden, C. and Andrews, D.A. (2004) The Importance of Staff Practice in Delivering Effective Correctional Treatment: A Meta-Analytic Review of Core Correctional Practice. *International Journal of Offender Therapy and Comparative Criminology* 48(2):203-214.

Faries, D., Leon, A.Haro, J. and Obenchain, R. (2010) *Analysis of Observational Health Care Data Using SAS*. Cary.NC: SAS institute Inc.

Gyateng, T., McSweeney, T. and Hough, M. (2010) *Key predictors of compliance with community supervision in London*. London Criminal Justice Partnership Research report. Available at

<http://www.icpr.org.uk/media/10306/Key%20predictors%20in%20compliance%20mcsweeney%20gyateng%20hough.pdf>

Harper, G. and Chitty, C. (2005) *The impact of Corrections on Re-Offending: a review of 'What Works'*, Home Office Research Study 291, London: Home Office.

Hedderman, C. and Hough, M. (2004) *Getting Tough or Being Effective: What Matters?* In: G. Mair (ed.) *What Matters in Probation*. Cullompton: Willan.

Hucklesby, A. (2013) *Compliance with Electronically Monitored Curfew Orders*. In: A. Crawford and A. Hucklesby (eds) *Legitimacy and Compliance in Criminal Justice*. NY, London: Routledge. McNeill, F. and Robinson, G. (2013) *Liquid Legitimacy and Community sanctions*. In: A. Crawford and A. Hucklesby (eds) *Legitimacy and Compliance in Criminal Justice*. NY, London: Routledge.

Ministry of Justice (2010) *Offender management caseload statistics 2009*. Ministry of Justice statistics Bulletin. London: Ministry of Justice.

Moore, R. (ed.) (2015) *A Compendium of Research and Analysis on the Offender Assessment System (OASys) 2009-2013*. Ministry of Justice Analytical Series. London: Ministry of Justice.

Pasta, D.J. (2000) Using Propensity Scores to Adjust for Group Differences: Examples Comparing Alternative Surgical Methods. Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference, Indianapolis, IN, 261-25. Available at <http://www2.sas.com/proceedings/sugi25/25/st/25p261.pdf>.

Raynor, P. , Ugwudike, P. and Vanstone, M. (2014) The impact of skills in probation work: A reconviction study. *Criminology & Criminal Justice* 14(2): 235-249.

Rex S (2012) 'The Offender Engagement Programme: Rationale and objectives'. *Eurovista: Probation and Community Justice* 2(1): 6–9.

Rex, S. & Hosking, N. (2014) 'Supporting practitioners to engage offenders', in I. Durnescu & F. McNeill (eds.) *Understanding Penal Practice*, London: Routledge.

Robinson, G. and McNeill, F. (2008) Exploring the Dynamics of Compliance with Community Penalties. *Theoretical Criminology* 12(4): 431-449.

Robinson, G. and Ugwudike P. (2012) Investing in "Toughness": Probation, Enforcement and Legitimacy. *Howard Journal of Criminal Justice* 51(3): 300-316.

Robinson, G. (2013) What Counts? Community Sanctions and the Construction of Compliance. In P. Ugwu-dike and P. Raynor (eds) *What Works In Offender Compliance International Perspectives and Evidence-Based Practice*. Basingstoke: Palgrave Macmillan.

Rosenbaum, P.R. and Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1): 41-55.

Shapland, J., Bottoms, A.E., Farrall, S., McNeill, F., Priede C. and Robinson, G. (2012a) *The quality of probation supervision – a literature review*. Sheffield: Centre for Criminological Research, University of Sheffield, Occasional Paper 3, at http://www.sheffield.ac.uk/polopoly_fs/1.159010!/file/QualityofProbationSupervision.pdf

Sorsby, A., Shapland, J., Farrall, S., McNeill, F., Priede, C. and Robinson, G. (2013) *Probation staff views of the Skills for Effective Engagement Development (SEED) project*. Sheffield, Centre for Criminological Research Occasional Paper no. 4. Sheffield: University of Sheffield. Available at http://www.sheffield.ac.uk/polopoly_fs/1.293093!/file/probation-staff-views-seed.pdf

Ugwu-dike, P. (2010) Compliance with community penalties: the importance of interactional dynamics. In Fergus McNeill, Peter Raynor and Chris Trotter (eds)

Offender Supervision: New Directions in Theory, Research and Practice. Cullompton:
Willan

Ugwudike, P. and Raynor, P. (2013) Introduction. In P. Ugwudike and P. Raynor (eds)
*What Works In Offender Compliance International Perspectives and Evidence-Based
Practice*. Basingstoke: Palgrave Macmillan.

Ugwudike, P. and Raynor, P. (2013a) Conclusion: What Works in Offender Compliance.
In P. Ugwudike and P. Raynor (eds) *What Works In Offender Compliance International
Perspectives and Evidence-Based Practice*. Basingstoke: Palgrave Macmillan.

10191 words.

Biographies and contact details

*Angela Sorsby, University of Sheffield, School of Law, Bartolome House, Winter Street S3 7ND,
UK; email a.sorsby@sheffield.ac.uk; telephone 0114 2226809.*

Angela Sorsby is a Lecturer in Criminology at the School of Law, University of Sheffield,
UK. Her research interests include: evaluation of criminal justice initiatives, probation
supervision, desistance from crime and restorative justice.

*Joanna Shapland, University of Sheffield, School of Law, Bartolome House, Winter Street S3
7ND, UK; email j.m.shapland@sheffield.ac.uk; telephone 0114 2226712.*

Joanna Shapland is the Edward Bramley Professor of Criminal Justice in the School of Law at the University of Sheffield, UK. She researches desistance, probation, restorative justice and the informal economy.

Gwen Robinson, University of Sheffield, School of Law, Bartolome House, Winter Street S3 7ND, UK; email G.J.Robinson@sheffield.ac.uk; telephone 0114 2226863.

Gwen Robinson is Reader in Criminal Justice in the School of Law, University of Sheffield, UK. She has published widely in the areas of community sanctions, offender rehabilitation and restorative justice.