Ministry of JUSTICE

Justice Data Lab

Methodology Paper

April 2013

Contents

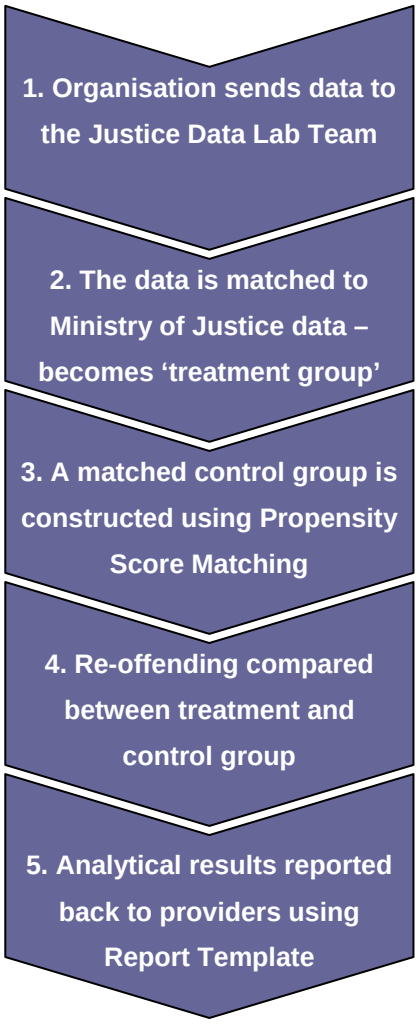
1. Overview

This paper presents the methodology that will be used to conduct the analysis for all requests made to the Justice Data Lab.

The aim of the analysis will be to assess whether data held by the Ministry of Justice suggests that a change in re-offending behaviour has occurred due to a participating organisation's offender rehabilitation work. The Justice Data Lab is therefore concerned with demonstrating impact, and the analysis of each request might be described as an impact evaluation.

Figure 1 presents an overview of the methodological process that will be used during the Justice Data Lab pilot.

Figure 1: An overview of the methodological process used in the Justice Data Lab, following receipt of data from providers.



Once a provider has submitted offender and intervention data to the Justice Data Lab which meets the minimum level of criteria (see the User Journey Document for details about what needs to be provided), the methodological process described in Figure 1 will be applied. The end result of the methodology will be:

- The one year proven re-offending rate[1] for the cohort of offenders that providers worked with, known as the treatment group;
- The one year proven re-offending rate for a matched control group made up of similar offenders to those in the treatment group but who did not receive the provider's intervention; and,
- The results of a statistical comparison between the two groups which allows an assessment of whether the provider's intervention has led to a change in re-offending rates.

The first step in the methodological process is to match the data sent by organisations to the Police National Computer (PNC)[2]. There are two parts to this:

1. Identifying the correct offenders on the PNC based on the data sent to us by the provider;

2. Identifying the appropriate custodial and/or community sentence relevant to the time period when providers delivered their intervention to offenders. This is important in identifying the appropriate point in time from which to measure re-offending.

[1] The **one year proven re-offending rate** is defined as the proportion of offenders in a cohort who commit an offence in a one year follow-up period which received a court conviction, caution, reprimand or warning during the one year follow-up or in a further six month waiting period. The one year follow-up period begins when offenders leave custody or start their sentence in the community.

[2] The Police National Computer (PNC) is the administrative data system used by all police forces in England and Wales. It is managed by the Home Office. The Ministry of Justice receive monthly extracts of data from the PNC.

For a variety of reasons, offenders may be lost from the original dataset during this process (see section 2.3 below). Those offenders who remain form the treatment group.

For information about the data needed to submit a request to the Justice Data Lab please consult the User Journey Document which accompanies this paper.

On its own, the proven re-offending rate for the treatment group does not allow for an assessment of impact because it is unknown what would have happened to the offenders if the provider's intervention had not taken place. A comparison with the national average would not be valid because the composition of offenders in the treatment group will likely be very different to those more generally across England and Wales.

It is not possible to know precisely what the proven re-offending rate for the treatment group would have been without the intervention, so this needs to be estimated. This estimate is known as the 'counterfactual'. The proven re-offending rate for the counterfactual is compared to the treatment group rate, and the difference between the rates is analysed in order to make an assessment of impact.

The counterfactual is estimated by constructing a matched control group of similar offenders to those in the treatment group using a technique known as Propensity Score Matching (PSM). After applying PSM, the only major difference between the characteristics of offenders in the treatment and control groups should be that one group received the intervention whilst the other did not. This isolates the effect of the intervention, allowing for increased confidence in attributing any difference in re-offending between the groups to the provider's intervention. Creating a matched control group allows a researcher to control for other factors that might influence re-offending (provided data is available on those factors). For the Justice Data Lab, these are factors which influence the likelihood of an offender receiving the intervention and/or re-offending - (e.g. criminal history).

The final stage of the methodological process is to compare the one year proven re-offending rates of the treatment and control groups using statistical significance testing. This returns a statistical significance value (sometimes referred to as a ‘p’ value), which denotes the level of confidence with which we can conclude that an intervention may have had a real impact on re-offending. The standard acceptable level of statistical significance necessary to demonstrate impact is 0.05 – which means you can be 95% confident that the difference was not down to chance. If the ‘p’ value is greater than 0.05 then there is insufficient evidence to confidently conclude that an intervention has impacted upon re-offending. The lower the ‘p’ value, the more confidence we can have in asserting that the data held by the Ministry of Justice indicates that an intervention may have led to a change in re-offending - provided the control group is closely matched to the treatment group and remembering that there may be other factors involved not covered by Ministry of Justice held data.

The results of the analysis will be reported back to the provider using a standard reporting template and will include commentary explaining the results and what they mean. An example reporting template can be found accompanying this document.

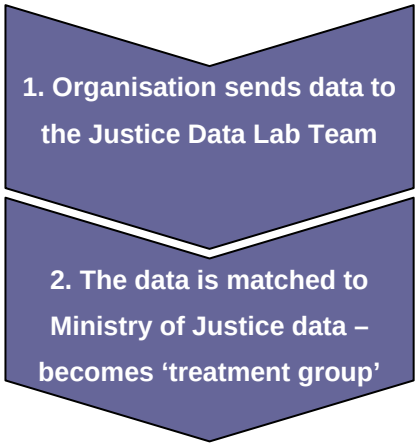
In understanding and interpreting the results (including what can and cannot be said about a piece of analysis) providers should consult the User Journey Document. When reporting results back to providers, commentary will be provided on how the data they originally sent progressed through the methodological process described in this report and what the results mean.

This paper describes the methodological process outlined above in full. A glossary is included at the end to aid understanding.

As the Justice Data Lab is in a pilot stage, the methodology may be subject to alterations, particularly where changes might enable the aims of the Justice Data Lab to be better met. For the duration of the pilot, we welcome feedback on the methodology proposed in this document.

2. Matching data to Ministry of Justice data

Once the Justice Data Lab has received data which conforms to the guidelines and criteria set out in the User Journey Document and the Data Upload Template, the data will be matched to Ministry of Justice data.



This process is done in two main stages:

1. First the offenders are matched to the Police National Computer (PNC) using the information provided such as forename, surname and date of birth. This is necessary to enable the proven re-offending rate to be calculated for the appropriate group of people.

2. The appropriate sentence is then found for each offender to ensure that the one year re-offending rate is measured from the appropriate starting point. This might be the sentence that the offender was serving at the time the provider delivered its intervention to the offender. It may also be the closest custodial sentence before a provider's work if they worked with offenders upon release.

Identifying the correct offender and appropriate sentence enables a range of offender characteristics to be extracted, relevant to each offender at a point in time just prior to receiving the intervention. These characteristics are used to construct the matched control group so as to ensure that the control group is similar to the treatment group. After this matching, the re-offending rate for the offenders the provider worked with is calculated.

If the correct offender or appropriate sentence cannot be identified on the PNC then the re-offending rate for the treatment group will not include this offender. This is explained in more detail in section 2.3. Organisations should be aware that offenders will be dropped from the sample and will not be

included in the analysis if the one year proven re-offending rate cannot be calculated.

2.1 Locating the offender

In order to successfully locate the correct offender on the PNC, it is important for organisations to have followed the guidelines on supplying data to the Justice Data Lab set out in the User Journey Document.

There are over 8.5 million offenders on the PNC, many of whom have the same name and many who have multiple alias names. If only forename and surname are supplied by the provider it is very difficult to correctly identify any individual offender. However, when date of birth and gender are also provided, testing on sample datasets suggests an offender identification rate of around 90%.

A matching algorithm has been developed to locate offenders on the PNC. This categorises offenders as ‘successful matches’, ‘suspect matches’ or ‘non-matches’. Successful matches proceed to the next stage of the methodological process.

There are many reasons why an offender can be a ‘suspect’ match. For example, the forename and surname supplied by the provider could be alias names used by the offender. Alternatively, the forename or surname recorded on the PNC may be spelt differently to the spelling of the names supplied – e.g. ‘Martyn’ as opposed to ‘Martin’.

When manually accepting or rejecting offenders flagged as suspect matches, it can be helpful to have additional offender information. The algorithm will produce a range of potential matches and without additional information it can be hard to work out which offender is correct. The User Journey Document sets out examples of additional information that can be useful in this regard, such as postcode.

Suspect matches will not be accepted unless it is clear that the correct offender has been identified.

Non-matches will occur if providers do not conform to the guidance set out in the User Journey Document and the Data Upload Template. Non-matches will be removed from the sample and therefore excluded from the re-offending rate calculated for the treatment group.

The final report will inform providers of the number of offenders who were successfully identified on the PNC.

2.2 Locating the sentence

Having identified the correct offenders on the PNC, it is then necessary to identify the correct point in time from which to measure re-offending, for each offender.

The one year proven re-offending rate reflects the proportion of offenders who commit at least one offence within a one year follow up period. The date on which the one year follow up begins will vary from offender to offender, and is known as the index date. For the Justice Data Lab, the index date is likely to be one of the following:

- The date when an offender is released from custody. This is appropriate for interventions delivered in prison or following release from prison (as in through the gates projects).
- The date when an offender begins a court order (such as a community sentence or suspended sentence order), appropriate for sentences served in the community.
- The date when an offender received a non-custodial sentence like a fine.

However, many offenders have received multiple custodial and/or community sentences over time. To produce re-offending data appropriate to the provider's work with offenders, the sentence corresponding to the provider's work needs to be identified.

To enable identification of the correct sentence, it is mandatory for organisations to provide some form of date information relevant to their time working with each offender. Guidance on date information can be found in the User Journey Document.

The final report will inform providers of the number of sentences that were successfully found on the PNC.

In certain cases, offenders may have delivered their intervention to the same offender across multiple sentences, for example during two different custodial sentences. For each offender where this is the case, only re-offending data relating to the first sentence will be included in the analysis as re-offending data for any subsequent sentences may be influenced by receiving the intervention the first time. It would not be appropriate to include the same offender multiple times in the treatment group.

2.3 Reasons for losing offenders

It is likely that some offenders will be lost from the original sample during the process of matching to the PNC. There can be no guarantee that re-offending rates can be calculated for every offender originally supplied by a provider. Reasons for offenders being dropped from the treatment group include:

- The supply of inaccurate or missing name and date of birth information, or the supply of alias names and dates of birth for offenders which are not recorded on the PNC. In the interest of robust analysis, offenders will be dropped from the sample if there is doubt over whether they have been correctly identified on the PNC.

- Offenders still being in prison or having only been released recently so that the one year proven re-offending rate is not yet available. At present, the re-offending rate can not be calculated for offenders who were released from prison in 2011 or later, or who are still in custody.

- The one-year proven re-offending data not existing for all custodial and community sentence spells. This is likely to cause the biggest loss in offenders not making the treatment group. The reasons for the absence of some re-offending data are explained in the Definitions and Measurement document published alongside the Proven Re-offending Statistics Quarterly Bulletin[3]. This is a limitation with any request made to the Justice Data Lab and with re-offending statistics in general.

The number of offenders not making the treatment group will be communicated to organisations in the final report along with the main reasons for reductions. Prior to conducting the analysis the Justice Data Lab team will work with providers to understand and improve the data they submit in order to minimise losses.

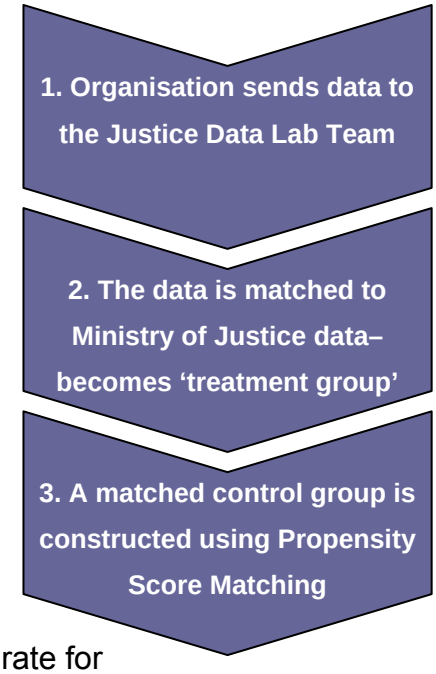
2.4 Minimum number of offenders

After matching to the PNC, treatment groups will only proceed through the methodological process if they consist of at least 30 offenders. Providers are asked to provide a minimum of 60 offenders initially, to account for potential losses as a result of matching to the PNC.

From a statistical perspective, any sample of fewer than 30 offenders will be very unlikely to produce a statistically significant result. This means there would be a low likelihood of concluding with an acceptable level of confidence that any difference in re-offending between the treatment and control groups was real rather than the result of chance.

[3] For the latest version, see Table 2 in: Ministry of Justice (2012). Proven re-offending statistics: definitions and measurement, www.justice.gov.uk/downloads/statistics/reoffending/proven-reoffending-definitions-measurement-apr10-mar11.pdf

3. Creating a Matched Control Group



Following successful matching of at least 30 offenders to datasets held by the Ministry of Justice, the next step in the methodological process is to construct a matched control group. This will consist of similar offenders to those in the treatment group.

3.1 The rationale for using a control group

It is not possible to assess the impact of an intervention by looking at the one year proven re-offending rate for the treatment group in isolation. Neither is it valid to compare the rate to the overall rate for England and Wales. This is because the composition of offenders in the treatment group will likely be very different to those more generally across England and Wales.

In order to assess whether an intervention has had an effect on re-offending it is necessary to estimate what the re-offending rate for the offenders would have been without the intervention. If this estimate is compared to the actual rate an assessment can be made about the impact of the intervention on re-offending. This estimate is known as a 'counterfactual'.

Of course, it is not possible to know for sure what would have happened to offenders in the treatment group in the absence of the provider's intervention. From a statistical perspective, the most robust research method available for negating this problem is a randomised control trial design. This would involve random assignment of offenders to either receive the provider's intervention or not followed by a comparison of the two groups' re-offending rates. However this design is impractical for a number of reasons, such as providers needing to randomly select offenders to work with, and the cost and time

needed to design the trial, put it into operation and obtain results which will often be years away.

When constructing a robust counterfactual, researchers need to isolate the effect of the intervention by ensuring that in all other important observed offender characteristics the treatment group and the counterfactual are as similar as possible. The counterfactual therefore comes in the form of a matched control group – where offenders in the control group are matched to offenders in the treatment group based on the similarity of offender and offence characteristics.

Isolating the effect of the intervention allows a researcher to be more confident in concluding that an observed difference in re-offending is the result of the intervention itself rather than other factors. Statistical testing can reveal the extent of the impact and the level of confidence appropriate to the result.

For requests made to the Justice Data Lab, a Propensity Score Matching approach will be used to construct the matched control group.

3.2 Propensity Score Matching

For the Justice Data Lab, the purpose of Propensity Score Matching (PSM) is to find offenders for the control group who are similar to those treated by the intervention in terms of their offender and offence characteristics. In PSM, offenders in the treatment group are matched to non-treated offenders with similar propensity scores. These matched, non-treated offenders form the control group.

Propensity scores are derived for each offender using a logistic regression model. This includes a range of offender and offence characteristics associated either with being selected onto the intervention program or with re-offending. Such characteristics include gender, ethnicity, age at offence and criminal history. The propensity score is a value between 0 and 1 which

reflects the likelihood that an offender received the provider's intervention given the recorded characteristics. Values closer to 1 represent an increased likelihood.

To illustrate the matching process, consider offender A who the provider worked with in prison and offender B, who did not receive the provider's intervention but did receive a prison sentence at a similar point in time. The logistic regression model produces propensity scores of 0.31 for both offenders – suggesting that each offender was as likely as the other to receive the provider's intervention based on the variables included in the model. Because the offenders have comparable propensity scores offender B is a suitable match for offender A and their actual re-offending can be compared; say in this case offender A did not re-offend whereas offender B did. Given that the propensity score is a function of offender and offence characteristics related either to selection onto the intervention or re-offending, it is likely that these underlying characteristics of offenders A and B are similar when the propensity scores are close.

Matching offenders in this way, based on the proximity of propensity scores to one another, is repeated for all offenders in the treatment group provided each finds one or more similar non-treated offenders. After the matching process has been completed, the quality of the matching can then be assessed.

There will often be cases where more than one non-treated offender has a similar propensity score to an offender in the treatment group. For Justice Data Lab analyses, all offenders who fall within an acceptable boundary of proximity will be included in the control group. This increases the precision of the estimate for the re-offending rate of the control group.

When developing the chosen PSM methodology using sample datasets, at least one acceptable control match could be found for all treated offenders and this resulted in closely matched treatment and control groups. However, it cannot be guaranteed that there will always be at least one non-treated

offender with a similar propensity score to treated offenders. In situations where no non-treated offenders can be found with similar propensity scores to certain treated offenders, unmatched treated offenders will have to be dropped from the treatment group. This is because no valid comparison is available for those offenders.

3.3. Developing the PSM model

The PSM model is used to calculate propensity scores for treated and non-treated offenders. The scores are subsequently used to create a matched control group.

The model includes all the offenders in the treatment group and a large pool of non-treated offenders who may potentially form a similar comparison group. Non-treated offenders would not qualify for the matched control group if, for example:

- They have different sentences to those in the treatment group; someone on a community sentence could not be matched to someone with a custodial sentence.

- The years over which re-offending is measured are different to that for any offenders in the treatment group; offenders released from custody in 2003 could not be matched to offenders in the treatment group if these were all released in either 2008 or 2009.

- They are a different gender to those in the treatment group; if the treatment group only included males, they would not be matched to females and vice versa.

- They are in the treatment group.

A series of variables will be tested for inclusion in the model. These variables represent a range of pre-treatment offender and offence characteristics, a full list of which can be found in Appendix A.

It is important that only variables measured prior to the treatment are included as variables measured after the intervention has commenced may have been affected by the treatment itself. This is why P45 employment, for example, is only measured in the year and month prior to the conviction date rather than afterward where it may have been affected by the intervention – particularly if the intervention itself is designed to help offenders find employment.

In order to be included in the model, variables will generally need to be related to either the likelihood of receiving treatment or of re-offending[4]. This will be decided by systematic testing of the variables' relationship with both the likelihood of receiving treatment and with re-offending[5].

Because each intervention is different, the variables used in the models will be different for each request submitted to the Justice Data Lab. One issue with PSM that applies to all requests, however, is that there may be variables that are not recorded for offenders (and so cannot be included in the model) but which may influence re-offending. The Ministry of Justice cannot feasibly keep a record of all variables that might influence re-offending, so this is a limitation for all analyses involving a matched control group. Please see the section 4.3 on caveats and limitations with the methodology for further discussion.

[4] On the whole, variables included in the model will have probability (p) values of less than 0.2 either relating to receiving treatment or re-offending. This means that the outcome observed in the data would be less than 20% likely to occur if there is genuinely no relationship between the variable and receiving treatment or re-offending. There may occasionally be exceptions where the variable does not have a p value of less than 0.2 but where it makes a statistically significant contribution to the goodness-of-fit of the model. The level of significance of 0.2 is consistent with the academic literature (e.g. Apel & Sweeten, 2010; Hahs-Vaughn & Onwuegbuzie, 2006; Rosenbaum, 2002) and previous Ministry of Justice research (Ministry of Justice, 2012).

[5] Using a Backward Elimination Stepwise procedure. This involves entering all the variables to begin with then removing those which do not warrant inclusion and finally checking that none of the discarded variables should be added back into the model.

Details on the variables included in the propensity score model and technical model output will be available on request following the production of the report.

3.4 Matching treated and non-treated offenders

Propensity scores for treated offenders are compared with those of non-treated offenders. All non-treated offenders whose scores are in a pre-defined close proximity to scores of treated offenders will be matched to the treated offenders, even if they have previously been matched to other treated offenders, and will form the matched control group. This approach is known as Radius Matching[6] (with replacement) (Dehejia & Wahba, 2002).

There will be no limit placed on the number of potential matches to individual treated offenders. By matching as many similar offenders as possible (rather than selecting the most similar offender or placing a limit on the number of non-treated offenders that can be matched to each treated offenders), the precision of the re-offending rate estimate is increased. This is thought to be particularly advantageous to Justice Data Lab analyses as providers will not generally have worked with a large number of offenders.

The level of proximity in propensity scores deemed appropriate to create a match will vary between Justice Data Lab requests. The ultimate aim is to achieve a matched control group that is as similar as possible to the treatment group on important variables to enable a valid estimation of the difference in re-offending whilst at the same time avoiding losses in the number of treated offenders where no similar match could be found. Treatment group sizes for Justice Data Lab requests are likely to be small, not least due to losses in the process of matching to the PNC, so it is desirable that as many treated offenders should be matched as can be.

[6] This 'matching with replacement' approach was introduced by Dehejia and Wahba (2002).

To be matched, propensity scores do not need to be exactly the same, only similar. The maximum distance between scores is known as a caliper[7] and it is this that will vary between requests. If the starting point for the caliper does not enable the groups to be well matched, it will be reduced (reduction means propensity scores have to be closer to be matched) iteratively until the groups are well matched (see section 3.5 below on assessing the quality of the match) but no further in order to avoid unnecessary losses of offenders from the treatment group.

3.5 Assessing the Quality of the Match

Once a matched control group has been constructed it is necessary to assess whether the control group is similar to the treatment group on important offender and offence characteristics. If the groups are well matched then re-offending for the control group and the treatment group can be compared.

The purpose of PSM is to produce a situation where the only major difference between the treatment and control group should be that the treatment group had the intervention and the control group did not. This isolates the effect of the treatment, increasing confidence in asserting that a difference in re-offending rates was the result of the intervention rather than other factors which are controlled by the matching. In order to assess the quality of the match, then, the two groups should be comparable on the 'other' factors – these are those identified when developing the model as being related to either the likelihood of receiving treatment or of re-offending.

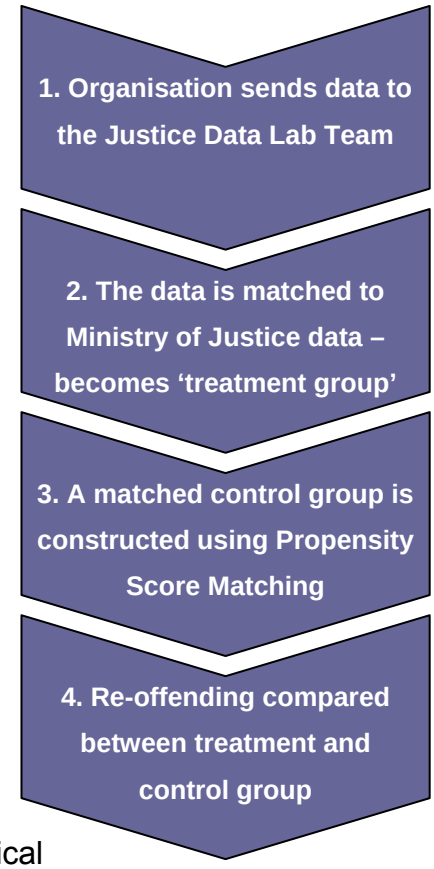
For all requests made to the Justice Data Lab, the quality of the match will be assessed by the standardised differences[8] between the means of the treatment and control group variables associated with receiving treatment and/or re-offending.

[7] Calipers are measured in terms of the number of standard deviations of the logit of the propensity score. For example a caliper of 0.2 (the starting point in Justice Data Lab Analyses (Austin, 2011)) would mean a match needs to be within 0.2 standard deviations of the logit of the treated offender's propensity score.

[8] Standardised mean differences take account of the variation in the treatment and control group variables as well as the differences between the means. For Data Lab analyses, they will be calculated according to the process set out in Austin (2008).

When reporting the results, the standardised mean differences for each variable will be provided in an annex. Standardised differences of 5% or less would indicate that the groups were closely matched for that variable. For example the treatment group may have a mean age at first offence of 14.1 compared to 14.2 for the control group with a standardised difference of 1% between the two – this would indicate that on this variable the groups were closely matched. Standardised differences of between 6% and 10% would suggest the groups were reasonably matched for that variable. Initial work suggests that most standardised differences will be below 5% with most of the rest below 10%. The groups will only be considered well matched if the vast majority of the standardised differences are below 10%. Differences above 10% would indicate a poor match and may alter the interpretation of the final result, particularly if the variable(s) concerned are highly related to the likelihood of receiving treatment or of re-offending.

4. Comparing the Groups



Once a well matched control group has been constructed to estimate the counterfactual, the re-offending rates for the treatment and control groups can then be compared.

When reporting the results of the analysis, the one year proven re-offending rate for the treatment and control groups will be provided, as will the percentage point difference between the two. This percentage point difference is a best estimate of the difference in re-offending rates between the two groups. However it is not on its own sufficient evidence to suggest an intervention has had an impact on re-offending. Statistical significance testing needs to be applied to the data to be confident about whether the estimate represents a real difference in re-offending between the

groups or is just the product of chance. This will be illustrated in the final report through placement of a 95% confidence interval[9] around the re-offending rate for both groups.

4.1 Statistical Testing

Statistical significance testing indicates the level of confidence providers can have in drawing conclusions about whether the difference in re-offending rates between the treatment and control groups is likely to be real.

If a reduction in re-offending is suggested by the percentage point difference between the two groups, this does not necessarily mean that there was a genuine difference between the two groups. This is because despite controlling for pre-treatment characteristics that might have affected the result, the matched control group is only an estimate of what would have happened to the offenders in the treatment group had they not received the treatment. Because the re-offending rate is only an estimate the percentage point difference between the groups is also an estimate; there is always the possibility of error. The likelihood of error will be partially reduced by including as many suitable matches in the control group as possible (so as to make the estimate more precise) but it cannot be removed completely.

So even though a comparison of re-offending rates may suggest a difference in re-offending rates between the groups, this change could be reflecting 'chance' or error, rather than a real difference between the two groups. This is why statistical significance testing is important, because it provides information about how confident you can be that your results were not down to chance.

[9] Each 95% confidence interval will show a range within which the true mean rate of re-offending in the treated or non-treated population is estimated to be in 95 such samples out of 100, Where 95% confidence intervals for the treated and non-treated groups don't overlap but the 90% confidence intervals do, this may be commented upon in the final report text.

4.2 Statistical significance and confidence intervals

The statistical significance testing[10] applied to the re-offending outcome data will be illustrated in the final report by placement of a 95% confidence interval around the re-offending rate for each group. This reflects the standard level of statistical significance in demonstrating impact in impact evaluations. A statistically significant difference is where there is a less than 5% likelihood that the observed result was down to chance.

To illustrate this, consider a treatment group with a re-offending rate of 35%, and a control group with a rate of 41%. These rates are the best estimates for each group, because the groups are samples of larger populations, and so are subject to error. A 95% confidence interval for each rate will give the upper and lower bounds in which there can be 95% confidence that the real rate for each group lies. For the treatment group this might be 32% to 38%, and for the matched control group 40% to 42%[11]. Taking these figures, the final report to providers would state that the reduction in re-offending found by the analysis could be anywhere between 2% and 10%. If the confidence interval boundaries of the treatment and matched control groups overlap then the difference will not be statistically significant at the 0.05 significance level.

For the Justice Data Lab, a statistically significant difference between the groups will still be subject to certain caveats.

[10]For the Justice Data Lab pilot, this will generally be a Weighted T-Test where the numbers of observations in the treatment group are sufficiently large. If a variable is related to either selection onto the intervention or re-offending, and the standardised mean difference is high, it may be appropriate to instead use logistic regression on the final matched groups to control for the remaining imbalance on this variable.

[11] The precision of the estimate is likely to be much higher for matched control groups than for treatment groups in Justice Data Lab analyses due to the much larger size of matched control groups which occurs as a result of the matching methodology.

4.3 Limitations of the methodology

There are a number of limitations to the methodology that providers should be aware of when submitting data to the Justice Data Lab. Where relevant these will be discussed in the final report along with any caveats specific to the analysis itself.

One important limitation from a provider perspective may be the impact of small treatment group sizes on the statistical significance of the result. The smaller a sample size, the more room for error there is in estimating the re-offending rate for the group as a whole, and the less chance of detecting a genuine difference in re-offending between the two groups.

If the analysis returns a non-significant result, it does not mean that the intervention has failed to reduce re-offending. It means there is insufficient evidence to draw a conclusion about whether re-offending in the treatment group is different to that in the control group. The major reason for imposing a minimum sample size limit for the Justice Data Lab is that this conclusion will almost certainly be drawn for treatment samples featuring fewer than 30 offenders.

When results are close to demonstrating a statistically significant reduction in re-offending it may lead to a recommendation that the provider resubmits to the Justice Data Lab if and when a larger sample becomes available (perhaps after another year's worth of data has been collected).

Another major limitation of the method is that propensity score matching can only account for offender characteristics captured by the Ministry of Justice. Despite the data held by the Ministry of Justice covering a wide range of offender and offence characteristics (see Appendix A), there may be factors that strongly influence either the likelihood of having treatment or of re-offending that cannot be used in the analysis because they have not been captured. These could include: an offender's motivation to receive

treatment, geographical variation and interventions delivered by other providers. This means a statistically significant difference in re-offending may be due, at least in part, to unknown factors rather than the intervention itself.

This is a particular concern when a variable not recorded on the PNC is very important in selecting offenders. For example, interventions may select offenders based on their relationship status but this is not captured by the PNC. There would be no way, in this example, of knowing whether the control group contains offenders with different relationship statuses to those in the treatment group and thus this variable's impact cannot be accounted for by a matched control group. Issues such as this will be discussed in the report where appropriate and providers should bear them in mind when interpreting their results.

Providers should also acknowledge that analyses in the Justice Data Lab only concern re-offending outcome data. Although outside the scope of Justice Data Lab analyses, there may be other important outcomes to consider for certain rehabilitation interventions. Examples of this might include increased employability or positive attitudes.

In addition, the one year proven re-offending rate might be considered a blunt measure of re-offending. Whilst it is certainly a very important indicator when assessing the impact of an intervention, it is not the only indicator available. Further re-offending data, such as the frequency of re-offending or the two year proven re-offending rate may be provided on a case by case basis where appropriate (e.g. if enough time has passed for a two year rate to be calculated, or if releasing the data conforms to our obligations under the Data Protection Act).

References

Austin, P. C. (2008). Assessing balance in measured baseline covariates when using many-to-one matching on the propensity score. *Pharmacoepidemiology and Drug Safety, 17*(12), 1218-1225.

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics, 10,* 150–161.

Apel, R. J., & Sweeten, G. (2010). Propensity score matching in criminology and criminal justice. In A. R. Piquero and D. Weisburd (eds.), *Handbook of Quantitative Criminology.* New York: Springer.

Dehejia, R. H., & Wahba, S. (2002). Propensity score matching methods for non-experimental causal studies. *The Review of Economics and Statistics, 84*(1), 151-161.

Hahs-Vaughn, D. L., & Onwuegbuzie, A. J. (2006). Estimating and using propensity score analysis with complex samples. *Journal of Experimental Education*, *75*, 31–65.

Ministry of Justice (2012). Peterborough Social Impact Bond: an independent assessment. Development of the PSM methodology. *Ministry of Justice Research Series 8/12.* London: Ministry of Justice.

Rosenbaum, P. (2002) *Observational Studies* (2nd ed.). New York: Springer.

Wermink, H., Blokland, A., Nieuwbeerta, P., Nagin, D., & Tollenaar, N. (2010). Comparing the effects of community service and short-term imprisonments on recidivism: A matched samples approach. *Journal of Experimental Criminology, 6*, 325–349.

Glossary of Terms

- **Caliper** – The maximum distance between the propensity scores of treated and non-treated offenders deemed acceptable for offenders to be a viable match.
- **Confidence Interval** – A range of values within an upper and lower bound. A 95% confidence interval would mean you could be 95% confident that the real value for a population of interest lies within the upper and lower bound. Confidence intervals will be a key output for Justice Data Lab analyses as the re-offending rates for the treatment and control groups are essentially samples of larger populations.
- **Control Group** – A group of offenders who did not receive the intervention being analysed. The control group is made up of offenders with similar characteristics to those in the treatment group.
- **Counterfactual** – An estimate of what would have happened with regards to the re-offending of offenders in the treatment group if they had not received the treatment.
- **Index Date** – The date at which the follow up period for measuring re-offending begins.
- **Logistic Regression** – A technique used to predict a binary, categorical outcome; for the Justice Data Lab this will mainly be used to ascertain the likelihood of an offender receiving treatment or not. Predictions are based on the variables used in the regression.
- **One year proven re-offending rate** – The proportion of offenders in a cohort who commit an offence in a one year follow-up period which received a court conviction, caution, reprimand or warning during the one year follow-up or in a further six month waiting period. In the Justice Data Lab analyses, one year proven re-offending rates will be provided for the treatment and control groups as the main descriptive output.
- **Police National Computer** (PNC) - An administrative data system used by all police forces in England and Wales, managed by the Home Office. The PNC records offender, crime and disposal details.

- **Propensity Score Matching (PSM)** – The methodology used for constructing a matched control group in Justice Data Lab analyses. Uses logistic regression to predict the likelihood of each offender receiving treatment; these predicted probabilities called propensity scores. Treated and non-treated offenders are matched based on the closeness of their propensity scores.
- **Standardised mean difference** – The standardised difference in means between the treatment and control groups, for an individual variable. The standardised mean difference is expressed as a percentage; the smaller the percentage the more similar the groups are on that variable.
- **Treatment Group** – The group of offenders that the provider delivered their intervention to. In other words, the offenders who received 'the treatment'.

Appendices

Appendix A: List of variables to be tested for inclusion in logistic regression models used to calculate propensity scores[12].

Offender Demographics

- Gender
- Ethnicity
- Nationality (whether the offender was born in the UK or not)
- Cohort Year (for analyses where offenders in the treatment group span years)
- Government Office Region and/or Prison of Discharge may be considered in ensuring that offenders in the matched control group come from the same area as those in the treatment group.
- Employment – whether the offender was in P45 employment in the year/month prior to the conviction date for the index offence.
- Benefits – Whether the offender received out of work benefits in the year/month prior to the conviction date for the index offence[13]

Index Offence (this is the offence that led to the sentence appropriate to the provider's work with offenders)

- Age at date of index offence*
- Length of custodial sentence (for custodial sentences only)
- OGRS offence Code (condensed 20 categories for the index offence, e.g. robbery, violence and so on, as in the Offender Group Reconviction Scale 3)
- Severity of Index Offence (ranked 1 to 3 with 1 being the most severe).

[12] A * besides the variable indicates that a squared term will be tested for inclusion in the model. Squared terms are able to account for any non-linear relationships between variables and the likelihood of receiving treatment or of re-offending (Wermink et al., 2010).
[13] Out of work benefits include Job Seeker's Allowance, Income Support and Incapacity Benefit.

Offending History[14] (all prior to index offence)

- Number of previous offences*
- Copas Rate[15]
- Number of previous custodial sentences*
- Number of previous court convictions*
- Number of previous court orders*
- Age at first contact with the criminal justice system*
- Number of previous offences for each of the 20 OGRS offence categories
- Number of previous offences split by severity of offence*

Other Interventions[16]

- If the offender has attended the following programmes a) during the sentence related to the provider's work with offenders and b) at some point prior to receiving the conviction for the index offence:
    - General Offending Behaviour Programme
    - Sexual Offender Treatment Programme
    - Drug Treatment Programme
    - Domestic Violence Programme

OASYS assessment data, which covers assessments made of offenders at various points in their sentence, will not be included at the outset of the Justice Data Lab pilot. Roughly half of all offenders have OASYs assessments completed. Given that the treatment group sample sizes are likely to be small in Justice Data Lab analyses, it was considered too risky to include OASYs data as the lack of completion would lead to losses of offenders from the treatment group due to missing data. Inclusion of OASYS data will be reviewed throughout the pilot.

[14] All offending history variables exclude Penalty Notices for Disorders.
[15] The Copas Rate controls for the rate at which an offender has built up convictions. The formula is a natural log of the number of court appearance or cautions + 1/lengthof criminal career in years + 10).
[16] Intervention data is only available for interventions delivered in custody or during a community sentence between 2008 and 2010. It cannot be included in the model if providers worked with offenders in 2007, for example.

Contact Points

Press enquiries should be directed to the Ministry of Justice press office:

Tel: 020 3334 3536

Other enquiries about the analysis should be directed to:

Justice Data Lab Team
Ministry of Justice
Justice Data Lab
Justice Statistical Analytical Services
7th Floor
102 Petty France
London
SW1H 9AJ
Tel: 0203 334 4770
E-mail: Justice.DataLab@justice.gsi.gov.uk

General enquiries about the statistical work of the Ministry of Justice can be e-mailed to: statistics.enquiries@justice.gsi.gov.uk

General information about the official statistics system of the United Kingdom is available from www.statistics.gov.uk