Home Office

HOW TO CONDUCT AN EVALUATION OF DIP IN YOUR AREA

CONTENTS

1. OVERVIEW AND INTRODUCTION

The Home Office has made a significant commitment to delivering a consistent and evidence-based approach to managing problems around drug use and offending. This has led, nationally, to a number of publications that emphasise the effectiveness of interventions but there is a recognition that we need, at both a national and a local level, to continue to evaluate and improve our work and our delivery, and to do so on the basis of the best evidence available. The information presented in this data sheet should offer some useful guidance on how to go about evaluating the key questions of how effective DIP is in your local area and its impact on drug related offending.

This guidance provides information about how to evaluate key aspects of DIP programmes and is intended to support DIP teams to conduct evaluations and research. It is not a standalone document and there is a considerable literature on specific research techniques and methods. It should provide you with some ideas about the scale and scope of the evaluation and some of the issues that you will need to consider but it will be useful to liaise with people who are experienced in conducting evaluations and research.

A more generic approach to evaluation methods can be found in "Passport to Evaluation v.2.0". This outlines a problem-solving model called SARA which identifies eight simple stages to conducting an evaluation:

1. Identify what the research question should be
2. Decide how you will measure the process or project to be evaluated
3. Identify what data you will need to collect
4. Decide how you will analyse the data and initiate the data analysis process
5. Look at the logistics of who will do the project and how its goals will be realised
6. Conduct the evaluation
7. Publish your findings
8. Develop a model for how the findings will be acted upon

Although this should not be taken as a rule, it offers a useful overview of the things that you need to take into account. This model provides an overview of more general ways of managing the evaluation process. The remainder of this guide attempts to focus down specifically on DIP and uses examples of real experiences of evaluating DIP initiatives and activities to explain some of the core activities.

It is important that evaluation be considered as a part of the routine for DIP services to ensure that core aspects of professional practice are adhered to – to show that practice is effective and to demonstrate the impact of DIP at a local level.

2. HOW TO GET STARTED

The starting point for any evaluation is 'what do you already know' and what is the reason for undertaking a particular activity? Evaluating your work is part of the process of professional development and ensuring that the service is committed to maximising resources and providing the best value for money for commissioners and for the clients of the service. An ongoing process of review and service development is best undertaken within an evaluation framework that is transparent and is replicable – in other words, it can be reproduced and re-tested in other areas or at other times.

There is a considerable national knowledge base that has helped DIP to become established and respected, and your work will be to build on that. The likeliest sources for existing information that will be available locally are:

- monitoring data – DIP compact data and NDTMS
- audit – can provide a crucial starting point about what standards apply and whether they are being met around areas of interest
- existing evaluation work – what information has been collected – perhaps in partnership with local universities or colleges, with the police or with treatment providers
- needs assessment – each DAT area will produce a needs assessment and a treatment plan that will set targets and will provide a summary of the local evidence base
- recent research
- establishing information partnerships – it may well be that key partner agencies, such as treatment providers or local authorities may have some relevant information and it will be important to work through partners both in terms of identifying what is already known and how to build on existing knowledge

Which of these you utilise will depend on the question, but consultation with other bodies who will have conducted needs assessments and audits, such as NHS treatment providers and Drug and Alcohol Action Teams, may provide some early insight into what is available, how useful and accessible it is, and what preliminary questions it may answer.

3. CASE STUDY: EVALUATING AN INTENSIVE INTERVENTION FOR HIGH CRIME CAUSING USERS (HCCU) IN COVENTRY

The starting point for this project was a recognition that a significant number of individuals were being arrested three times or more, were testing positive on each occasion for Class A drugs, but were failing to engage with the treatment provider. This analysis was conducted by a DIP police inspector supported by analysis undertaken by a police analyst.

Following the review of arrest and treatment uptake, a new programme was developed mirroring PPO but for this population who were not engaging with treatment services and who did not qualify for PPO status.

Coventry DIP and partners established a more intensive partnership working arrangement between dedicated police officers and treatment workers to target this group and to offer them more intensive engagement into treatment. As only some people were offered this, an opportunity arose to compare their outcomes compared to those receiving the 'treatment as usual' package through DIP. This kind of opportunistic approach is known as a 'quasi-experimental' design in that people were not allocated to conditions but some people received the intensive intervention while others did not.

To support this project, the DIP team partnered up with researchers from the University of Birmingham who conducted the evaluation of the intervention and so provided external and more objective analysis.

Decision on method: The question was simple – "does intensive partnership working improve outcomes for high crime-causing users?" Therefore, the initial stage of the discussion was to identify a primary measure that was available and consistent for assessing impact and to supplement this with additional information where possible. The decision that was then taken was that arrests were the most reliable indicator of offending and that this was the primary measure. This could be obtained from the police analysts for all candidate participants. Additionally, we wanted to assess what level of treatment participants received and their views – and it was agreed that we should review the casefiles and ask participants where possible.

The next big decision was about the sample and the design used is called a 'quasi-experiment' – some people received the intervention and others did not. This happened naturally because of available 'slots' in the caseload of the team but meant that there were roughly equal numbers of those eligible who received the intensive intervention and did not, and the only difference between the two groups was whether there was a place available at the time. This is a robust research design that means it is possible to compare changes in arrests in both the group receiving the intervention and those not receiving it.

Three types of data were assessed overall:

1. arrest data in the six months prior to and after the start of the process
2. treatment case files assessing how much treatment was received and what effect it had
3. questionnaire follow-up to assess offenders' perceptions of the treatment they were offered and received

The results were unequivocal – offenders who received the more intensive support reported a drop in their offending rates by almost 50% while there was no change in the other group.

Not all of the evaluation was successful – the response rate to the questionnaire was poor, especially among the 'non-intensive' group but the advantage of having three types of data were that the eggs were not all in one basket. The results were sufficiently robust that the evaluation was subsequently published in an academic journal, Policing and Society.

This case study will be used as an example of various processes as we proceed.

4. BENEFITS OF REVIEWING PROCESS

The process of reviewing potential collaborators and previous local activity will allow you to:

1. Avoid duplication
2. Build on existing local knowledge
3. Identify local expertise and skills for the evaluation work
4. Identify processes and activities that allow for comparison
5. Identify success and build upon it

This initial review process will also allow you to work out who the key potential resources and collaborators are as you move from the review of what has gone before to planning. Once this initial stage of surveying existing knowledge and current research and evaluation activity locally has been completed, there is then a task to review what is known about your question nationally and internationally.

One starting point for this is the Home Office website and possible research literature review – for criminal justice issues, a potential starting place is the Campbell Collaboration website. There are two main purposes to this:

1. What has recently been done and what lessons have been learned in practice about application and implementation
2. What lessons have been learned about the strengths and weaknesses of different evaluation methodologies and techniques.

This process will provide a level of academic compatibility and will ensure that you both do not fall into the mistakes that others have learned from and also that you can build on existing knowledge. One of the key benefits of the academic review process is to build on what other people have done and work out what that may mean for the question you have identified.

5. BUILDING HYPOTHESES

It is important that you do not over-interpret data from a single source or time-point and to consider alternative explanations and predictors of behaviour and data. Thus, access to appropriate comparison datasets both locally and nationally will help you contextualise and understand the findings you obtain. The challenge here is to sift the information that you have acquired from the review process – including academic and local information sources – and use them to develop a research question.

While you may end up with multiple questions it is important that your questions can be clearly specified. Thus, in the case example we have used the basic question was "Do people who receive more intensive interventions after DIP contact have better outcomes?". Putting questions in an operational context allows you to consider whether the question can be answered and whether the resources and materials you have at your disposal will allow you to do this. In the case study, our primary concern was to ensure that we had enough different ways of testing whether the two groups would differ on a range of measures.

There are a number of core questions that you might want to ask – there are overall questions of effectiveness ("has DIP made a difference?") and the effectiveness of specific components (e.g. the proportion of positive testers seen by a DIP and then the proportion of these who access treatment services), and more complex questions of value for money. The latter is an incredibly complex and debated question that will often involve estimating how much of a behaviour has been prevented, what impact that might have and how it can be costed relative to the costs of running the DIP service.

With increasing focus on outcomes rather than process measures, there is increasing emphasis on changes in long-term recovery. Thus, indicators of change – such as reduced arrests or positive tests, and overall volume of drug-related crime may be seen as more important and timely measures than indicators of the number of people engaging in treatment. Similarly, establishing links between changes in the behaviours of drug-using offenders and positive effects on fear of crime and reported crime will be important markers of effective provision.

It is also important to be clear about what it is you are trying to change – thus, whether it is the behaviour of a group of high risk offenders as in the HCCU example in this document or whether it is offending that occurs at an area level. If the former, then the data will be about changes in behaviour of the individuals (e.g. number of arrests or positive tests), if at an area level it may be more appropriate to look at the total number of arrests or offences. These issues may also influence the time window that you are assessing – but typical windows for comparing changes in arrest would be between 6 months and 1 year depending on the data available and the frequency with which the target behaviour occurs.

WHAT ARE THE AREAS OF DIP THAT COULD BE EVALUATED?

Although there will be local variations in what you may consider to be key areas of the programme, it may be useful to consider the stages of the process:

1. Are all arrestees assessed for suitability? In other words, are there individuals who might benefit from engagement with DIP staff who are not seen? If so, what can be done about this?
2. Are all of those identified as suitable actually seen by a DIP worker? One question might be about the availability of DIP workers in the evenings, early mornings and weekends – are DIP workers seeing the right people? If not, what are the approaches that could be used and evaluated to address this?

3. Does seeing a DIP worker help? Irrespective of whether arrestees engage with treatment, does seeing a worker motivate drug-using offenders to change?
4. What proportion of offenders seen by workers actually make contact with treatment services? What predicts variability in rates of engagement and what evidence is available to improve this?
5. Does engagement with DIP and/or treatment appear to affect re-offending and re-arrest rates? In other words, do people keep coming back into custody regardless of treatment status and engagement with DIP workers?
6. What predicts variability in test results?
7. How successful is integration with the police team and what impact does this have on early engagement and effectiveness of DIP interventions?
8. Are the DIP staff satisfied, motivated and engaged?
9. Do they have sufficient supports and resources to do a good job?
10. Is the DIP team delivering value for money?

These are among the core questions that you may wish to address – but some of them are complex questions and you may wish to break them down into simpler and staged processes and questions. The most basic starting point for working out the priorities will be to look at monitoring and performance management data and then to conduct basic audits of what is the current situation. Once you have done that, it will be appropriate to move onto the next stage.

6. IDENTIFYING POSSIBLE DATA SOURCES

Data access can be a complex process and it is essential that it is both ethical and practical to access the information that you will need for the results to answer the question you have set. One particular challenge in this respect can be access to health service data – and there are two generic approaches:

1. Information sharing protocols across partnership agencies
2. Agreement in advance from offenders that any personal data can be used for evaluation purposes

If you are looking at how many clients take up addiction treatment, this may not require prospective research and might be managed on data held by the providers. What level of information you can access will depend on the relationship you have with them (e.g. will they give you client identifiers, data on engagement and retention in treatment, whether they are scripted etc) and on what performance management agreements are in place.

It is much simpler to be able to address this kind of question by shared information than by following clients up yourself – on average it is estimated that one single client follow-up in the community takes roughly one day of researcher time. This is a huge resource, so accessing information from partner agencies can save lots of time and effort!

It may be useful to match up what data you have access to and the extent to which it is likely to meet your needs. Thus, DIP engagement data, test results, and offence data may be accessible but may not satisfy all of your requirements. An overview of possible types of data and their strengths and weaknesses are outlined below:

Controlling for extraneous variables: In the 'hard' sciences, the gold standard of research is usually considered to be the randomised control trial (RCT) in which subjects are allocated, on a random basis, to different levels of the intervention. This is extremely limiting and resource-intensive but allows researchers to control unwanted factors that might influence the result. In applied evaluation, the opportunity for RCT design is unlikely to be possible or even desirable. However, without that level of control it is extremely important that possible confounds are accounted for – thus, is a reduction in positive drug tests a consequence of changes in policing policy or testing procedure rather than because of the success of the treatment model? This is one of the reasons why repeated testing in different settings or at different times may increase confidence in the results.

Checking what is already out there should help you to frame the question you want to ask and identify options for the evaluation method you will employ. How you frame the question is crucial to ensure that:

1. You are asking a question that can be answered
2. With the information at your disposal or that you can access
3. Within the resources available to you
4. Within the time available to carry out the work

There is likely to be a process of review where the question is revised in terms of what can be answered based on the data and the resources available. Evaluation can be time-consuming and requires some knowledge and expertise and it will be useful to consult with colleagues who have undertaken evaluations previously about what the time and resources required will be.

Evaluation and research will often benefit from an impact assessment – what this

means is assessing what the costs are of conducting the evaluation – as well as direct costs to the evaluators, what the impact will be on workers and clients and will it reduce the effectiveness of the service, compared to what the likely benefits are going to be. It is also worth bearing in mind that the broader the question to be addressed the more difficult it will be to pin down a precise answer.

Confidentiality and data protection: There are key issues around Data Protection that need to be addressed to protect the rights of individuals. This primarily concerns individual level data with the potential for identification of individuals. For clients accessing services, this can be dealt with by signing consent forms at the time of initial contact, but is much more problematic for retrospective analysis of information or data. To ensure that the rights of all participants – staff as well as service users/offenders – are protected, it is strongly recommended that the Caldicott Guardian is contacted at the planning stage of the research or evaluation process to identify and address both statutory and local requirements about information management. If the partnership working involves NHS service providers, local research and development (R&D) teams should be contacted and they will also be able to assist with whether NHS ethical clearance is likely to be needed.

In considering issues of anonymity, it is important that the individuals will not be identifiable on the basis of the information reported or presented – i.e. data on the one outlying individual who has committed the highest level of crime. One of the benefits of partnering up with an NHS Trust or a university is that it will enable access to an ethics committee who will be able to advise on such issues and who will provide guidance on what local guidance is about information and consent forms; on processes for anonymising information and on the appropriate methods for reporting at individual and group level.

7. SPECIFICATION AND FORMULATION OF THE EVALUATION METHOD

So how does DIP work in your area? This is a very broad question and you would need to consider what exactly it is that you want to know:

1. Are drug using offenders being successfully identified and assessed?
2. Are they engaging with the treatment service?
3. Are they retained effectively by the treatment service?
4. Does this impact on their offending?
5. Does this result in a reduction in the overall levels of reported offending for trigger offences?
6. Is it cost effective?
7. Does it have greater impact than other possible interventions might have?
8. Are there mediating variables (eg treatment staff and effective housing and employment support) that contribute to its effectiveness?

These are large and complex questions and you may have to think with each of them how they can be broken down into simpler questions that can be answered more immediately and some of these will constitute programmes of research. However, the collation of any available evidence will allow you to start an evidence-based model and will allow subsequent attempts to be more directed and more precise. Establishing a culture of evidence-based practice and getting the team into the habit of using evaluation and data to address questions of delivery and improvement is a gradual process.

8. SELECTING A METHOD

When assessing the effectiveness of DIP and its associated interventions, it is generally not sufficient to measure whether clients show improvement (e.g. stop offending, or becoming drug free), as this could result from a host of other factors. Knowing that clients are less likely to re-offend or test positive following contact with DIP or after spending time on the caseload is a good start, but the next level of research question is to assess how confident you are that these improvements are due to the intervention or part of the DIP programme you are evaluating. The sorts of question you might want to answer include:

1. Do people who get this intervention do better than those who don't get it? Here it is important that you control for any self-selection bias – in other words, that those who engage and are retained are not just those who are the most motivated to engage in treatment or who have the most supports and resources.
2. Have overall results (e.g. overall volume of crime, reported fear of crime, numbers engaged in treatment and so on) improved since DIP or a particular aspect of the programme such as Testing on Arrest was introduced? You also need to consider such factors as the timescales you are going to look at, the size and composition of the area and any other local factors that you think might influence the results.
3. Does your service perform better [measured how] than equivalent services which don't have this approach?

The choice of research method may encourage you to identify an appropriate comparison group – another DIP area that is similar in other respects, or other groups of offenders who don't receive the intervention. If you want to assess whether things were better than before, you are relying on the quality of the information that was available in the period prior to establishing your intervention. If you have enough time, it is a good idea to collect 'baseline' data before your intervention is introduced – it will give you a comparison and it might also inform you early about potential problems of data collection.

Comparison groups are difficult to find and the problem may often be that the areas compared differ on more than the factors of interest in the research making comparisons problematic. For this reason, time series data (comparing the same location before and after the introduction of a new programme or intervention) may be beneficial. In the evaluation of intensive interventions for HCCUs in Coventry this was the approach adopted with comparison both between groups and over time used to make comparisons more meaningful.

So it is critical that the question is precise and that this is matched against a data collection or analysis method that will be robust enough to address the question. So, if you wanted to know whether motivational enhancement provided in custody influenced the uptake of treatment, it would be necessary to compare a group who received this intervention with a group who did not – preferably in the same setting, and if possible, randomly assigned to active intervention or control.

Ideally multiple sources of information will be available – such as monitoring data, self-reported data from clients, worker reports, arrest or conviction data – that allow the researchers to have enough confidence that the results are 'triangulated' – in other words, that data from different sources are compared to assess whether they display consistent outcomes.

The role of partnerships: In making all of these research decisions, it is critical that you consider what resources you may have available and what the partnership

should be for the evaluation project. Thus, in the first instance, police data analysts might be able to provide you with some initial information about crime patterns and trends, and other organisations may have access to statistical supports and resources. In the absence of this, it may be worth considering partnerships with local colleges and universities who may welcome the collaboration and the opportunity for applied research. It is also crucial that you find out whether other local organisations have undertaken similar work that you could build on or that would allow you to shape your research questions and design. It is also important that you consider having an evaluation steering group – this will not only allow you to generate commitment to the project from partners, it will also increase their commitment to the outcome of the process. Finally, the involvement of partner agencies should also help to deal with issues of support and access and enable effective project integration in partner organisations.

9. PILOT PROJECTS

This is in effect the equivalent of doing feasibility work and it is crucial in assessing whether it is worth investing large amounts of time and resource in a project. By conducting pilot work you may well discover strengths and weaknesses of your design, organisational and practical obstacles to delivery and possibly give you early indication of unanticipated effects.

Case example: Delivering a brief intervention to cocaine users: In Birmingham DIP, there was a concern that primary cocaine powder users were engaging with treatment at a very low rate. A brief intervention was developed in consultation with the University of Birmingham and training was provided to the team. The pilot phase was to assess the feasibility of delivery, the number of sessions delivered, their acceptability to the identified users and their subsequent uptake of ongoing drug treatment. This was done initially for one month to see whether workers were able and willing to use it, if the materials were satisfactory and if it successfully engaged the client group. The aim was to decide whether there was sufficient justification to role this out on a larger scale.

10. WHAT CAN GO WRONG?

There are a range of possible problems that can go wrong in the course of an evaluation that you should be aware of and mitigate against if at all possible before the start of the project.

1. **Failure of delivery** – if the programme is not delivered as planned, the findings will not be an adequate test of the intervention – and to do so it is essential that you have some measure of 'implementation fidelity' and also that participating staff and organisations will commit to sticking with it for the duration of the study. Thus, in the above example, workers required considerable support and training before they were confident to deliver a cocaine intervention and the early weeks had to be excluded from the evaluation as offenders were not being offered the intervention in custody.
2. **Failure of measurement** – particularly if you are using monitoring data, it is essential that you test that enough of it is collected, recorded, accurate and available – there is no point in having large numbers of case notes if the key information is missing on most of them. One of the other benefits of piloting can be to test that the sources of data actually have what you need. For example, if you are interested in test results and you are using case notes, then taking a random selection of 10 case files and testing whether the information is complete in each case is an important preliminary step.
3. **Failure of data analysis** – if you do not perform what is called a power calculation to test the likely sample you will need, you will not know if a failure to find an appropriate effect is a result of lack of numbers in the study (or lack of sufficiently robust data) or because there is 'really' no effect. Power calculations are a statistical test that you can download for free from the internet that will offer you a range of methods for calculating if you have enough statistical information to test the question you have in mind.
4. **Failure of implementation** – you need to be sure that what you find will actually influence ongoing service delivery or policy. Thus there is little point in delivering the strongest evaluation or research if nobody is going to read it or pay any attention to the outcomes. One specific problem that needs to be considered if projects are to be undertaken over long periods of time is how the context might change in the meantime – this would include changes in police resources and priorities, changes in personnel, in treatment delivery or in partnership membership and participation.

11. WHAT ARE THE TYPES OF DATA THAT CAN BE COLLECTED?

There are a range of types of information that can be collected that can be used individually or concurrently. In the analysis below, they are presented by likely populations:

1. **Offenders** – the most common sources of data on offenders would be arrests or convictions as 'objective data' although they are skewed by policing and other criminal justice system processes. Thus, one of the main weaknesses of arrest and conviction data is that once individuals are known to the police they may be disproportionately likely to be arrested again and arrests do not establish guilt. Convictions are a more solid measure of this but reflect variations in decision-making by a whole range of bodies (such as the Crown Prosecution Service and by juries) and it is important to make a clear decision on how TICs should be managed in the process. On the other hand, these are objective events that should have quantitative data accessible for them and convictions are a matter of public record and so should not lead to issues around consent or data protection (if the client is anonymised in reporting). The alternative approach is to measure self-report which has the potential to reveal a much wider array of undetected crime but is rendered problematic by issues of under-reporting and self-presentational biases. Thus, not only are offenders less likely to report undetected crime, they are also less likely to report undetected crimes that are highly stigmatised such as sex offences or domestic violence. In addition to deliberate under-reporting, memory effects are also problematic particularly for lifetime events and there is research evidence to suggest that onset ages will be reported differently as people age. The more stigmatised or threatening the nature of the question, the more grounds there are to question the validity of self-reports – thus, asking about undetected crime and sensitive areas of life (e.g. parenting or sex work) may provoke under-reporting. However, for measures of treatment motivation, satisfaction and engagement this may be essential, and self-report may be improved in reliability if the offender knows that the researcher has other sources of information and the information sought is more proximal in time.
2. **Offender collaterals** – if there are concerns about the reliability of the offender, one option is to gather data from partners, family members or other professionals linked to the target, but this is both expensive and intrusive and does not guarantee against the same types of bias in reporting.
3. **DIP or other criminal justice staff** – for interventions that are introduced, it may be critical to collect information on acceptability, ease of deployment, satisfaction and perceived impact. They may also be the source of 'activity' information – contact frequency, client behaviour etc, but there may be issues around self-presentational bias and it is essential that ethical clearance and organisational agreement has been obtained for this.
4. **Other key informant groups** – this may include drug workers, court officials, custody and other police officers and other partner agency staff. Questionnaire and face to face interviews may be the most likely sources of information here but questions may arise about how these will be analysed and what will be done with the outputs. Qualitative data from interviews are difficult to analyse and this is a time consuming process if it is to be done systematically.

5. **General public** – the kind of assessment that is common here is fear of crime or perceived impact of interventions. For this kind of analysis, sampling (both the numbers of people accessed and how representative they are of the community) will be key questions. General population surveys tend to use stratified samples to ensure that appropriate groups by gender, age and ethnicity are represented. For local issues, it is important that the general public sample actually lives or works in the area of the intervention and so decision making is important. You also have to be confident that the intervention will be sufficiently successful that the effect will be felt by members of the public. It will also be important to assess what they actually know about the DIP service or any interventions linked to them. One measure of general public impact is reported crime. Thus, in assessing the effects of DIP in Birmingham, one of the measures used was the reported number of burglaries. This removes the problem of sampling groups of the public and asking their perceptions but will only work for offences that have a high rate of reporting. In other words, this would be less effective for drug dealing or shoplifting.
6. **Victims** - this is an increasingly common area of criminal justice research yet there are significant ethical and practical problems about accessing this group and very strong justification would be required for including this sample in a DIP evaluation. The kind of case where it might be appropriate is restorative justice projects. Linked to the point at the end of bullet 5, this may be an advantageous way of accessing basic enumerative data, but awareness of variations in reporting need to be taken into account.

The key process is to consider what resources and time are available, what personnel and skills would be needed and to match them against the sources of data and methods outlined above. The design to be used will be the reconciliation of pragmatic and conceptual issues. One last academic principle is worth bearing in mind – this is the Principle of Parsimony (sometimes called Occam's Razor). This is the idea that the best method will be the simplest one that can explain the phenomenon under investigation. And do not try to do too much at once – a systematic and phased approach to answering clear and simple questions is much more likely to succeed than a convoluted approach that will lead to large amounts of information that make little sense!

12. IMPLEMENTING THE METHOD

Having decided on the appropriate design and target population(s), it is critical to think about what data collection approaches are to be implemented. The main options here will be:

1. **Case-file data:** Has the advantage of accessibility but may not contain precisely the information required and may not always be available or complete
2. **Monitoring data:** Particularly in the form of arrest data, this is consistent and statistically reliable, but there may be problems of access, of limited detail available and of validity as indicated above
3. **Questionnaires:** This is the most common form of prospective data collecting and is easier with audit where data collection can be managed against an agreed standard. For evaluation and research the problem is often what measure or battery of measures to use – while standardised instruments have advantages (generally published information on the psychometric properties of scales and ideally population norms as well), they can be cumbersome and time-consuming, and you may be tempted to design your own measures. However, this is not a simple task and advice should be sought from a research methodologist before doing this. There are a number of online resources about designing questionnaires but the basic principles include:
   - Only ask what you need to know
   - Do not ask questions that ask more than one thing at a time
   - Do not ask either or questions where there could be a third option
   - Avoid leading or loaded questions
   - Use a standard scale if rating e.g. 1-5
   - electronic forms are easy for data collating but paper forms can be more user friendly
   - Informed consent must be given
   - Pilot test the questionnaire to make sure the questions make sense and are not ambiguous
   - Provide a contact name and email/ telephone on the form and for respondents to keep
   - Make sure you know what you are going to do with the answers and have a plan for how they are to be analysed
4. **Interviews:** Can provide the most in-depth data but are very resource-intensive in terms of managing face-to-face contact and in terms of the collation and analysis of the results. Clear protocols are required for analysis of this type of qualitative data and permissions must be obtained around use of quotations and identifiable information. If you are asking sensitive questions, one of the key issues to be addressed from an ethical perspective is what the conditions for disclosure will be. Thus, if the individual reports things that suggest that they are at risk or that others are at risk, promises of confidentiality or anonymity may have to be breached and participants need to know this in advance. Workers may also feel compromised with police colleagues if they are made aware of offending and are not able or willing to report this. For all forms of direct contact, but particularly interview, there may be concerns if the evaluator is a member of the team about how this may change their relationship with offenders who are also clients, and they need to be clear about their role.
5. **Objective tests:** This will include urine or oral fluid tests, as well as paper and pencil tests such as cognitive

function tests, assessments of mental health and assessments of physical functioning. Access and permission may need to be sought for some tests (as well as payment to use them) and there may be training requirements involved. Analysts will have access to DIP test results locally (DTR), but partnership with services may be required for individual level data and similarly for matching up engagement with DIP teams.

The key issue is that the methods you employ address the question that you regard as important and are proportionate to the resources and supports available. While an assessment of outcomes may appear to be the most scientifically valid approach, there are circumstances where a survey of offenders, some structured interviews with key informants or retrospective analysis of monitoring data may provide you with all of the information you need. Thus, if you are looking at the extent to which integrated working with DIP treatment teams enhances treatment engagement, then assessing the concordance between DIP assessment and treatment using monitoring data may be all that is required. Alternatively, if you are wanting to assess impact on key stakeholders a short survey may provide the appropriate evidence. Indeed, if you are going to be systematic in your programme of ongoing evaluation, it may make most sense to do something relatively brief and simple as the mechanism for generating support for a more complex and sophisticated type of research or evaluation. There is no point in undertaking a long and expensive outcome evaluation if a survey of custody sergeants would tell you that none of the arrestees receive the target intervention in any case.

So who should carry out the evaluation? Ideally an individual or group independent of the delivery of the DIP intervention and who have the skills both to collect the appropriate data and to subject it to rigorous and replicable analysis. It is important that it is not those who are delivering the intervention or have a stake in its outcome who are involved in delivery, as they may (unconsciously) influence their behaviour and so the results because of their beliefs and expectations. The method will also be shaped by what is being measured and what the appropriate indicator is – whether that is a process measure of implementation or practice (such as effective delivery of brief or motivational interventions) or an outcome indicator (like reduction in rates of re-arrest or positive testing).

13. ANALYSING AND MAXIMISING THE RESULTS

The data analysis method should have been agreed in advance to avoid the situation that large amounts of data are collected that are not subsequently used. A basic framework for analysis should be considered as part of the evaluation design and this should be adhered to as much as possible. This may require the input of a researcher to know what statistical tests are to be used and what implications this might have for the collation and entry of data. It may well be that the input of a local university or college researcher would be helpful.

If you are collecting quantitative data, you should ensure that your project team includes the data entry, data management and data analysis skills required to test the questions you wish to address. This is likely to mean some specialist research input and may also mean access to the appropriate data analysis packages. Although most computer systems will have MS Access and Excel, these may not be sufficient to analyse the information particularly if you are looking to compare differences between groups. For social science quantitative data, SPSS is the most commonly used package but this can be expensive and is another reason why collaboration with existing researchers may be beneficial.

In reporting the results you should include unexpected findings as well as those that confirm your original hypotheses. This is important as part of an ongoing learning cycle and it is important that you are as objective as possible in making sense of the information that you collect.

What are you going to do with the results? A clear dissemination approach should be in place to identify who will be given the report, what process there is for interpreting the findings (i.e. is there going to be a project steering group of multiple stakeholders who will be involved in working out what the findings mean and what is to be done with them) and where we go from here.

Finally, it is critical that the evaluation is not carried out in isolation – whatever the findings are it is essential that there is a clear process for implementing recommendations and testing whether they have made a difference. In an ideal world, the evaluation process will be an ongoing one that is tied in to service planning, workforce development and delivery.

14. DISSEMINATING THE FINDINGS AND MOVING FORWARD

It should be clear by now that evaluation projects should not exist in isolation within the SARA model, the Assessment phase is about testing implementation and effect. As with audit, the process of evaluation should be ongoing and ideally cyclical with consistent and systematic performance management and quality assurance supplemented by bespoke projects that test specific interventions or processes. Thus, for the case example of HCCU's in Coventry, the next questions that arise are about sustainability of the initial effect and the identification of 'matches' – which particular groups benefit from the intervention and what the key ingredients of effectiveness would be.

Not only will a steering group be important in establishing an evaluation, they will also have a crucial role to play in deciding what the results mean and what comes next. It is important that people do not get dissatisfied that one evaluation study will not answer all of the questions. Managing expectations is important; you must make sure that your customers/steering group are clear about what the evaluation will and will not be able to deliver. It is crucial that there are plans for implementing any recommendations, for committing to an ongoing process of audit and evaluation and for assessing what the next steps would be in a systematic process of ongoing service development and improvement for DIP.

APPENDIX 1: APPROACHES TO RESEARCH AND EVALUATION

There are a range of approaches that can be adopted to gather information systematically – each of the four main approaches outlined below – audit, evaluation, needs assessment and research – have slightly different rationales and approaches to data collection and management. What is outlined below is an initial guide of what each of these things means and involves. They are not mutually exclusive and can be used in stages depending on the type of question that is being asked and the resources available to address them.

1. AUDITING CLINICAL ACTIVITY IN THE CITY

The most basic systematic form of data collection is audit, and the most common version of that in the addictions field is the clinical audit. Clinical audit was defined by the UK Department of Health as "systematically looking at the procedures used for diagnosis, care and treatment, examining how associated resources are used and investigating the effect care has on the outcome and quality of life for the patient" (Department of Health, 1993). In essence it is a method for reflecting on and reviewing practice in a systematic way, and utilising this to improve the quality of service. The key benefits of clinical audit are to:

- Identify and promote good practice
- Provide opportunities for training and development
- Ensure the most efficient use of resources
- Improve communication and working relationships, and the consistency of practice

Audit is not the same as research, as it generally involves the identification of an accepted standard and the measurement of performance against this standard. In contrast, research may have more fluid objectives and is often aimed at the generation of new knowledge. In this case, an example of audit would be the completion of the DIR form against the guidance issued by the Home Office with omissions identified having possible implications for training and supervision of the assessment process.

A key component of the information cycle that has significant implications for applied clinical research is around the audit cycle. In many treatment services (including the NHS) there is a governance requirement to engage in a programme of clinical audit in which processes and practices are mapped against established standards – such as the National Treatment Agency's Models Of Care (NTA, 2002; updated, 2006). One of the advantages of the method used is that it can be readily evidenced – the care plans and reviews, if that is what is being assessed - will either be in the case files of each client or not and they will either be complete or not. However, this can form the basis for a more sophisticated analysis of quality and also of impact – so the first stage of the applied research process may be to audit one particular aspect of treatment – such as the implementation of care planning. A second stage might be then to evaluate the quality of the care plans, and to use qualitative interviews with clients and workers to assess how important and useful they are. This would provide the foundation for an intervention to improve the process that would itself be the subject of evaluation. The advantage of a care planning cycle is that the routine audit of the target activity would in itself constitute an established and useful measure of effectiveness of change and its durability. In this sense, audit is both the first stage of initiating research agendas but also the basic building block of monitoring and measurement that, if used in a meaningful and cyclical way, can make an important contribution to the planning and delivery of strategic research aims.

This is a key principle to bear in mind – while one-off processes are useful, a systematic process of intervention, evaluation, review and re-consideration is likely to provide a much more consistent method for delivering improved services and standards.

2. NEEDS ASSESSMENT

Although needs assessments are a well-recognised part of developing treatment services in the addictions field, the adequacy of the science in this area is not good, and relatively little has been published to guide researchers in conducting structured needs assessments. Ford and Luckey (1983) suggested four steps that underpin systematic needs assessment, to identify the size of the in-need population:

1. Determine the geographic size of the area under investigation: this is basically to ensure that there is a specified geographic area and that each of the candidate sources fall within this boundary – this is generally referred to as the issue of 'co-terminosity'.
2. Estimate the number of problem users within each population group: using existing data sources from epidemiological analysis – which in our example included Home Office assessments of the size of the drug problem (Hay et al, 2008) and local estimates available for alcohol from each Public Health Observatory in the UK.
3. Estimate the number of users from step 2 that should be treated within a particular year (defined as the demand population): this is a complex question that will depend on the data sources available and on the assumptions made – such as whether the number of people requesting treatment should be weighted according to such factors as mortality, blood-borne viruses and so on.
4. Estimate the number of individuals from step 3 that will require a service from each component of the treatment system: this part of the analysis will attempt to match what is available in the current system (e.g. number of beds for detox or rehab, number of community treatment slots) against what is known about the characteristics (such as problem severity) of the candidate treatment population.

This approach was extended by Rush (1990) in assessing the need for alcohol treatment in Canada. Using an epidemiological model, Rush (1990) estimated that 15% of problem drinkers in Canada can be considered as the 'target group' for treatment in any given year. While the sources for this estimate were national alcohol-related mortality data, survey data on population drinking levels and population-level consumption data, the model developed was akin to a survival model for treatment seeking, based on individuals making each transition through the process of treatment seeking, assessment, engagement and completion. In this approach, for every 10,000 adult problem drinkers, 1,500 (15%) will seek treatment of whom around two-thirds (n=1,000, 10%) will make it as far as assessment and just over half to the point of treatment referral. However, the proportions who will seek treatment will vary as a function of local treatment systems and the methodological challenge will be to work out a mapping process for assessing clients reasons for seeking treatment and the types of treatment that would be acceptable to them.

In the alcohol treatment system assessed by Rush, just over half of the referrals were to out-patient services, around 30% to day programmes, 10% to short-term inpatient treatment and 5% long-term residential treatment. This approach to needs assessment is highly rigorous and systematic but relies exclusively on the availability

and quality of adequate epidemiological data and says nothing about the quality of the treatment service. In developing this kind of approach to needs assessment, the challenge is to combine quantitative measures of existing treatment activity (such as demand and uptake, supplemented by such things as measures of waiting times and early drop-out) with more qualitative assessments of perceptions of treatment delivery and need from a range of key stakeholders, such as clients, treatment providers and commissioners, and family members and community representatives.

For DIP, some of the key questions around needs are what the needs of arrestees are in terms of their substance use and its links to ongoing offending behaviour.

3. QUALITATIVE RESEARCH APPROACHES

There are also issues that arise about the types of data that are collected – most broadly categorised as 'qualitative' and 'quantitative'. The easiest data to manage are quantitative data in which closed category methods give rise to data that are enumerable and amenable to being combined according to a range of statistical manipulations. However, this information may miss some of the richness or complexity that may be more accessible in a qualitative data collecting format.

Strauss and Corbin (1998) argue that qualitative research is "any type of research that produces findings not arrived at by statistical procedures or other means of quantification" (Strauss and Corbin, 1998, p.10). While this issue is dealt with more generically in the chapter on qualitative methods, it is important to emphasise the role that this approach can play in applied research settings. Qualitative approaches can be used to assess organisations and how they function, social movements and cultural phenomena, and can be used at different stages of the applied research process, as is often used at the start of a programmatic investigation for example where knowledge about an area is limited (eg Stern, 1980).

Thus, it would not be unusual in undertaking a large research project, such as the topic of delivering structured interventions in applied settings, to use qualitative methods at the start and at the end of the research programme. This can be done to generate hypotheses for more structured investigation and also as part of the process of assessing implementation effectiveness. Equally, qualitative methods can be used running alongside quantitative methods as a method of assessing the experienced process of the intervention or activity taking place. Thus, for instance, the introduction of a new treatment facility may be tested using formalised treatment evaluation processes, but the impact on the local community assessed using more qualitative and perhaps ethnographic methods.

What would be the point of using qualitative methods at the start of the research process? At the outset, when attempting to define the research question, it may be appropriate to use open-ended interviews with a range of stakeholders to allow an initial assessment of what the issues are and what the level of consensus and consistency is across the stakeholder group. Similarly, at the end of the research programme, qualitative methods may be used to assess the impact of the work or, using a reflexive approach, to enable the researchers to consider the impact of the research process on the subject and context (Steier, 1991). It is one of the key differences between qualitative and quantitative methods that the assumption of objectivity and invisibility is not assumed in qualitative methods – in other words, the researcher's own role in the production and interpretation of data is more readily acknowledged. In addiction research, one of the key roles that qualitative research may perform is to give a 'voice' to users – and one of the key politicised objectives of

much qualitative research is as an act of empowerment for disenfranchised groups (e.g. Edwards and Potter, 1992).

Thus, for the question we have set about delivery of effective structured interventions, a qualitative approach may have involved a combination of interviews with key stakeholders, some observations of both treatment sessions and wider activities within treatment provider organisations. Feeding this back to participants and using the early interpretations as part of a reflexive process would contribute to what is often referred to as 'action' or dynamic research. Action research is a reflective process of progressive problem solving and takes research from a passive observation of naturally occurring events to an active participant in the process of managed change. In our case example, the process was dynamic but would not be classed as action research as it did not have that participative quality.

Qualitative analysis may be appropriate, for instance, in assessing what female stimulant users may need and to understand why they may be unwilling to access certain types of treatment. In this context, some open-ended and unstructured interviews may generate hypotheses and themes that can be explored further in later phases of the research programme.

4. EVALUATION RESEARCH

The basic purpose of evaluation research is to test whether something has worked or not. There are really two separate issues to be addressed – the first is the question of 'implementation fidelity' or whether the method or intervention tested has been implemented as intended and this falls within the area of 'process evaluation' which will attempt to assess how and how well the target intervention has been delivered. The second type of evaluation research is outcome evaluation which is an assessment of whether the intervention has delivered the desired results. There are a wide range of methods for undertaking this kind of evaluation but the most straight-forward and the one to be explored here is the 'pre- / post' method. What this involves is assessing the level of functioning or activity prior to the implementation of an intervention, then delivering the intervention and then assessing change in the period following implementation. Please note that this is not a controlled experiment and so does not exclude the possibility that unpredicted or extraneous variables may have influenced the observed change. In other words, this kind of design is susceptible to unpredicted intervening effects. To explain this process, we have cited an example from our own work, using the delivery of a training package around implementing a brief manualised intervention for drug users as the activity to be assessed.

This process is based on a sequenced implementation of training in a novel intervention in which baseline measures are taken of client and staff satisfaction with core aspects of treatment prior to the delivery of a training package. The training is then evaluated in the standard way of assessing receptiveness and impact, but an additional evaluation is then carried out with staff around three months later to measure their views on what aspects of the training have actually been implemented. This follow-up is a direct measure of obstacles to training implementation. Finally, the cycle is completed by repeating the initial satisfaction surveys of workers and clients and measuring changes that can be linked to the training implementation.

The combination of methods and techniques will depend on the resources available and the skills and knowledge of the team working on the project. It will also be shaped by the local cultural context and the requirements and timetable of the specific project.