Ministry
of Justice

Offender Management Community Cohort Study Universal Dataset Technical Report

Jack Cattell and Alan Mackie (GtD)
Martin Wood (NatCen Social Research)

Disclaimer

Acknowledgements

The authors

Jack Cattell, Director, Get the Data Limited
Alan Mackie, Director, Get the Data Limited
Martin Wood, Research Director, NatCen Social Research

Contents

List of tables

List of figures

1.   Introduction

1.1   Overview of the OMCCS

The Offender Management Community Cohort Study (OMCCS) was a longitudinal study that described a cohort of offenders aged 18 and over who commenced Community Orders between October 2009 and December 2010. The Ministry of Justice (MoJ) commissioned NatCen Social Research and Get the Data (GtD) to carry out the OMCCS. This involved the collation of:

- **Central administrative records** which were collected for all offenders starting a Community Order between October 2009 and December 2010 (144,407 offenders). These describe the Community Order received, the index offence committed, the number of previous offences and the assessed risks and needs of individual offenders.

- **Local administrative records** from 10 selected Probation Trusts,[1] covering 50,935 offenders. These describe how offender management operated and how offenders completed or breached their sentences.

- **A longitudinal survey of offenders** –three waves of interviews were conducted with a representative sample of offenders drawn from the 10 Probation Trusts as the local administrative data (2,919 offenders interviewed at Wave 1). The survey covers offender attitudes, needs and experiences whilst on a Community Order.

These data sources were matched and then linked for individual offenders to form a Universal Dataset. This dataset allowed offenders' experiences and outcomes to be traced through different data sources, so the study provides a broad and in-depth picture of how the Community Order was operating. The broad aims of the study were to:

- assess the effectiveness of interventions in reducing offending behaviour; and

- evaluate aspects of offender management.

---

[1]   This report describes selecting 10 Probation Trusts; however, at the time of the OMCCS fieldwork the process of forming Trusts was still under way, and the selection was actually based on Probation Areas. There were 42 Probation Areas in England and Wales (covering the police force areas), which have become 35 Probation Trusts. The Trusts were set up following the Offender Management Act 2007, when some of the Probation Areas joined to form Trusts. The 10 areas were: Derbyshire, Hertfordshire, Kent, Lancashire, London, North Wales, South Yorkshire, Staffordshire and the West Midlands, Teesside and Wiltshire.

In its initial conception, the OMCCS comprised three main stages:

- the baseline stage – which described the commencement of Community Orders and each offender's position at an early stage in their sentence;

- the mid-Community Order stage – which described the implementation of Community Orders, including what was delivered and to whom; and

- the end of Community Order stage – a final stage that explored the outcomes for the cohort, particularly in relation to breach of the Community Orders and reoffending.

This technical report describes the data sources collected for the OMCCS, and the construction and analysis of the Universal Dataset. It builds on two previous technical reports that covered the baseline stage of the study and the longitudinal survey of offenders (Wood *et al.*, 2013;[2] Wood and Hussey, 2013[3]).

1.2   Structure of the report

This report covers the content, construction and use of the Universal Dataset:

- Chapter 2 describes the data sources included in the OMCCS;

- Chapter 3 describes the Universal Dataset's design, how it was built and its final structure.

The report does not contain results from the study. A number of reports presenting findings from the OMCCS are available on the GOV.UK website.[4]

---

[2]   Wood, M., Hussey, D. and Cattell, J. (2013) *Offender Management Community Cohort Study (OMCCS) Baseline Technical Report*. London: Ministry of Justice.
[3]   Wood, M. and Hussey, D. (2013) *Offender Management Community Cohort Study: Wave 2 & 3 Technical Report*. London: Ministry of Justice.
[4]   https://www.gov.uk/government/collections/the-offender-management-community-cohort-study

2. Data sources

The OMCCS data were collated from three sources. This chapter describes each source to provide a comprehensive account of the types of data available.

2.1 Overview of data sources

The three data sources all contain information on different aspects of Community Orders. The linking of individual offenders across the data sources allows the experiences and outcomes of the offenders to be described over time:

- The first source of data is the central administrative data available from National Offender Management Service (NOMS) and the Ministry of Justice (MoJ). These data were collected for all offenders who commenced a Community Order between October 2009 and December 2010 (144,407 offenders).

- The second source is local administrative data, which were collected in 10 selected Probation Trusts (50,935 offenders).

- The third source is a longitudinal survey of offenders sampled from the 10 selected Probation Trusts as in the local administrative data (2,919 in total, 2,595 of whom agreed to allow their survey responses to be linked to administrative data).

Each source is described in more detail below. The central and local administrative data were collected over three periods: June to August 2011; November 2011 to January 2012; and November 2012 to January 2013. The overall population sizes increased slightly between each point of collection and the data holders corrected data errors (for instance, tier levels were updated for individuals as they progressed through their Community Order, or where tier level was previously missing).[5] As a result, frequency counts of variables such as tier, ethnicity and age changed over time as offenders were added to the data and as their details were amended. Also, the way in which data were used in the study changed over time as understanding improved from the successive, detailed analyses of different aspects of the data. Consequently analyses of the central and local administrative data completed at the beginning of the study may have produced slightly different results from analyses completed later.

---

[5] The National Offender Management Model has a tiering framework for matching resources and offender management styles to different types of offenders; a tier is allocated to an offender based on a number of factors, including their risk of reoffending. There are four tiers: Tier 1 is the lowest tier, and these offenders have minimal levels of interventions in their sentence. As the tier increases the risks posed by the offender increase and the sentence imposed increases in complexity. The OMCCS survey excludes Tier 1 offenders as they have minimal levels of interventions in their sentence.

2.2   Central administrative data

Central administrative data were collated on all offenders who commenced a Community Order between October 2009 and December 2010 in all 35 Probation Trusts in England and Wales. The data describe the Community Order received, the index offence committed, the number of previous offences and assessed risks and needs of individual offenders. The central administrative data were drawn from:

- Form 20, a centrally collated dataset that describes each Community Order commencement and termination including information on: tier; dates of commencement and termination; requirements; sentence length; demographics of the offender; and reason for termination;

- Offender Assessment System (OASys) data, which describe an individual's risks and needs, and their sentence plan;[6]

- details of the offenders' attendance on any accredited programmes from the Interim Accredited Programmes System (IAPS); and

- Police National Computer (PNC) records, which contain information on proven reoffences for which a conviction or caution is given;[7] this includes when the reoffence was committed, and the offence type.[8] PNC records also contain information on offence history.

Table 2.1 shows the central administrative data sources, the type of data collected and their purposes.

---

[6]   The first OASys research compendium can be found here:
http://webarchive.nationalarchives.gov.uk/20100505212400/http://www.justice.gov.uk/publications/compendium-research-analysis-offender-mgt-system.htm. This compendium provides details of the content, development and research of OASys.

[7]   PNC only covers recordable offences and so a number of non-recordable summary offences, such as low-level driving offences, are not included. Please see
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/311307/guide-to-criminal-justice-statistics-may-2014.pdf for further information about what data PNC records cover.

[8]   Previous reports described the PNC data as a fourth source of data. For the purposes of describing the Universal Dataset, the PNC data are presented as part of the central administrative data.

Table 2.1: Main data obtained from central administrative data sources

Source	Main data	Main purpose	Derived data
Form 20 – commencements	Community Order type Commencement date Probation Trust Case Reference Number Police National Computer (PNC) numbers Tier OGRS3 score Sentence length Index offence Requirements Age Gender Risk of serious harm	To provide variables to describe the offenders, specifically: their basic demographic characteristics; their risk assessment; the offence for which they were sentenced; their Community Orders (length and the purpose of their sentence)	Likelihood of reoffending groups (OGRS3)[9] Grouping of offender and sentence characteristics such as age groups, sentence length and offence type
Form 20 – terminations	Termination date Reason for termination	To describe when the Community Order(s) ended and the reason it ended	Breach rate
OASys assessments (commencement and termination of Community Order)	Needs and risk assessments Sentence plan OGRS3 Predictors of reoffending – general and violent Risk of serious harm	To describe the offender's offence history, what criminogenic needs he or she might have and what was contained in the sentence plan	Identified criminogenic needs at the beginning, during and end of the sentence Needs addressed in the sentence plan Whether identified needs are addressed in the sentence plan
IAPs	Programme types Programme start Programme completion Number of sessions Dates of sessions Completion dates	To describe what accredited programmes the offender attended, when they started, and whether they completed the programme	
Police National Computer (PNC)	Reoffence date Reoffence type Main offence (if more than one reoffence committed on the same date) Number of previous offences	To determine the rate of reoffending	Proven reoffending – whether committed a reoffence that received a court conviction or caution within 12 months from commencement or from interview date[10]

---

[9] The Offender Group Reconviction Scale (OGRS3) uses static factors (such as age at sentence, gender, offence committed and criminal history) to predict the likelihood of proven reoffending within a given time (usually one or two years after starting their Community Order). OGRS3 groups scores into low (less than 50% chance of proven reoffending within two years), medium (between 50% and 74%), high (between 75% and 89%) and very high (90% or more chance of proven reoffending) risk bands.

[10] In addition to the 12-month follow-up period for reoffending, a six-month waiting period is included to allow any prosecutions to reach a conclusion.

2.3   Local administrative data

Local administrative data were collected from 10 Probation Trusts selected at random with probability proportionate to size, covering 50,935 offenders. Two types of data were collated: contacts data and breach data. The contacts data describe what contact the offender had with the offender manager and other probation and delivery staff. The breach data identify offenders who did not comply with their Community Order and the outcome of these breaches. Table 2.2 lists the local administrative data sources, types and purposes.

Table 2.2: Main data obtained from local administrative data sources

Source	Main data types	Main purpose	Derived variables
Breach	Date of breach Outcome (charge of breach withdrawn, proven, etc.)	To describe the level of compliance, beyond that available in central data sources	Breaches that resulted in a modified Community Order within 12 months Overall breach rate
Contacts	Contact date Contact type Contract description Contact free text Missed contacts	To describe when contacts between the offender manager and the offender occurred, and the number of missed contacts	Total number of supervision contacts Total number of missed supervision contacts When Drug Rehabilitation Requirement (DRR) testing began and ended When unpaid work began and ended When an education, training and employment intervention began and ended

2.4   Longitudinal survey of offenders

Three waves of the survey were conducted, between December 2009 and July 2011. Tiers 2–4 offenders were sampled from the 10 Probation Trusts that were selected for the local administrative data (see section 2.3)I. n the survey sample cohort, the Wave 1 survey linked to the central and local administrative data for 2,595 offenders and the Wave 2 or 3 surveys linked for 1,640 offenders.

- The first longitudinal survey of offenders (the Wave 1, or baseline, survey) was carried out approximately three months after the start of the offender's Community Order, and 2,919 offenders were interviewed.

- The second wave of the survey (Wave 2) was conducted seven months, on average, into the sentence and 1,827 offenders were interviewed.

- The third wave (Wave 3) of the survey was conducted following the sentence's expected end point and 440 offenders were interviewed (fieldwork was not completed due to concerns about value for money given response rates).

These interviews provided information on offenders' perceptions and experiences of Community Orders, their backgrounds (e.g. employment, family life), attitudes and needs, and how these changed over time. The previous technical reports provide details about how these surveys were administered (Wood *et al.*, 2013; Wood and Hussey, 2013). Table 2.3 lists the sources, data types and main purposes of the longitudinal survey data.

Table 2.3: Main data obtained from longitudinal survey of offenders

Source	Main data types	Main purpose	Derived data
Survey Waves 1, 2 and 3	Basic demographics Sentence characteristics Evaluation of the Community Order Pre-sentence report Probation officer relationship Requirements Risk assessment Health and drug use Family Housing situation Attitudes to offending Offending behaviour	To describe offenders' experiences of, and attitude to, their Community Orders. The data also fill gaps that exist in other data sources such as when requirements started and how requirements such as the unpaid work were delivered.	When requirements started Risks, attitudes and needs Change over time (each wave of the survey) in these measures Whether started programmes

3. Creating the Universal Dataset

The previous chapter described what data the OMCCS collected and the purpose of each data type. This chapter continues by explaining how these data were linked to form a Universal Dataset. First the overall design of the Universal Dataset is explained and then the steps taken to construct the Universal Dataset are set out. These steps were: selecting and processing the data to include in the dataset; linking the individual data sources; checking the content of the dataset; and constructing the final version of the Universal Dataset.

3.1 Design of the Universal Dataset

The Universal Dataset contains data on all offenders who commenced a Community Order between October 2009 and December 2010. The design of the Universal Dataset was conceived as three cohorts of offenders that correspond with the three main sources of data described in the previous section. The three cohorts are:

- full cohort – central administrative data (Form 20, OASys, IAPS and PNC) linked for up to 144,407 offenders;[11]

- local cohort – local administrative data linked to the central administrative data for 50,935 offenders; and

- survey sample cohort – Wave 1 survey linked to the central and local administrative data for 2,595 offenders and the Wave 2 or 3 surveys linked for 1,640 offenders.

The local and survey sample cohorts are subsets of the full cohort, with more information being available for these smaller subsets. This concept is summarised in Figure 3.1.

---

[11] See section 3.3 for the total number of records from each source and the populations they cover.

Figure 3.1: Universal dataset cohorts[12]

Full
- Central administrative data
- 144, 407 offenders that commenced Community Orders between October 2009 and December 2010

Local
- Local administrative data
- Collected from 10 Probation Trusts
- 50, 935 offenders sampled

Survey
- 3 survey waves
- Random selection of offenders who agreed to data linking
- 2, 595 offenders at Wave 1
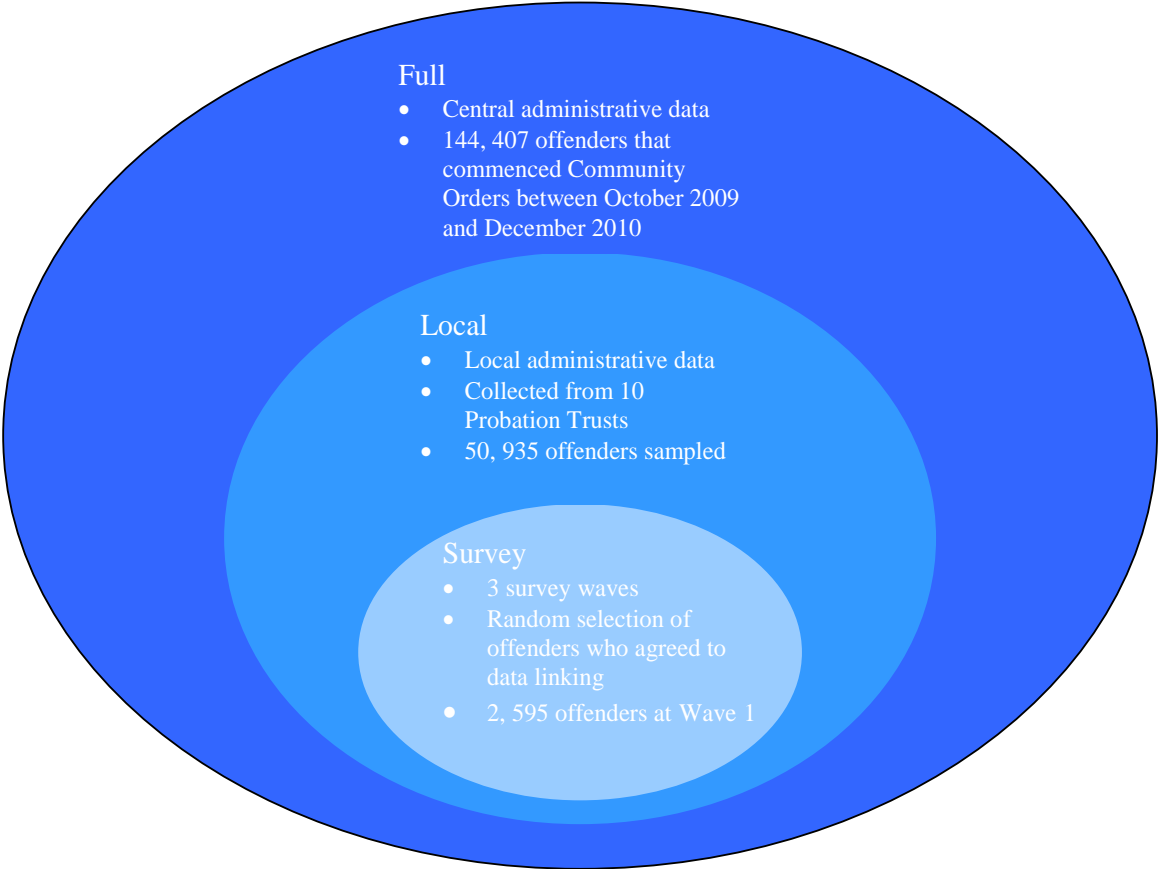
3.2   How the data were selected and processed

Cohort sample criteria

The sampling criteria for the central and local administrative data were: all commencements between October 2009 and December 2010, and an offender's first commencement in that period (if they commenced more than one Community Order). Sometimes an offender had two concurrent Community Orders that started on the same day. In these circumstances the Community Order with the highest tier[13] rating was sampled.

The longitudinal survey of offenders used different sampling criteria from those used for the central and local administrative data. Offenders for the survey were selected from the 10 Probation Trusts that had been sampled for the local administrative data. Offenders were selected at random by NatCen each month based on monthly flows of Form 20

---

[12]   The central administrative data include PNC data, but this is not shown in this figure for clarity.
[13]   There are four tiers: Tier 1 is the lowest tier and Tier 4 is the highest tier. Higher-tier offenders receive more intervention in their sentence.

commencement data from the MoJ. Black minority ethnic and female offenders were oversampled in order that a large enough sample would be achieved to enable analysis.

For one hundred and fifty offenders who started more than one Community Order in the sampling period, the Community Order commencement that was sampled in the survey was not the first Community Order they had started. In 39 cases, the Community Order commencements sampled by the survey did not match those in the local administrative data. For these 39 offenders, the closest date match in the local administrative data was used.

OASys data selection

The OASys assessments were matched to the Form 20 data by NOMS, using surname, date of birth and gender. OASys assessments are completed at multiple time points across an offender's sentence including at the start and termination of sentence points. All start and termination assessments completed within 90 days before and after an offender's commencement date and within 90 days of the termination date were matched to the Form 20 data.

These were then restricted to one start and one termination assessment per offender by selecting:

- the start OASys assessment closest to the commencement date; if a start assessment was unavailable, the assessment made for the pre-sentence report (PSR) was used; and
- the termination assessment closest to the termination date.

OASys data were linked for:

- 91,097 start assessments and 51,894 termination assessments in the central administrative data;
- 32,615 start assessments and 18,146 termination assessments in the local administrative data; and
- 2,401 start assessments and 1,556 termination assessments in the survey sample.

IAPS data selection

The NOMS Interim Accredited Programmes System (IAPS) was linked to the Form 20 data, again using surname, date of birth and gender. The IAPS data describe each accredited programme undertaken as part of a Community Order.

IAPS data were linked for:

- 17,859 offenders in the central administrative data;
- 5,958 offenders in the local administrative data; and
- 352 offenders in the survey sample.

Data on contacts

Each of the Probation Trusts provided individual contact records that describe the frequency and purpose of contact between the offender and the Probation Trust staff. These data were in different formats and had different contact codes to identify the type of contact, e.g. a supervision contact. To create a consistent dataset across all of the Probation Trusts, each Probation Trust's contact codes were mapped to the corresponding requirement, such as supervision or unpaid work. Some of the contact records were shared without a contact date, or the type of contact (for instance whether for supervision or unpaid work) was not described; these records were deleted from the dataset.[14]

Using the case reference number (CRN) to identify unique offenders, contacts data were linked for:

- 48,818 offenders in the local administrative data; and
- 2,336 offenders in the survey sample.

Data on breaches

Each of the Probation Trusts provided breach data. Each Probation Trust used different codes to describe the same outcomes such as whether the Community Order was modified (for instance, a requirement was added) or revoked. To create a consistent dataset, the codes of seven Trusts were mapped to a generic set of outcome descriptions, so the data could be analysed together. The remaining three Trusts provided descriptions of breaches in open text fields where the Offender Manager had noted the type of breach event that occurred. In these circumstances, the study team manually coded the text to identify when a breach had occurred and the type of breach it was. For these three trusts, breach data were only coded for the survey sample cohort. Owing to the variability of data quality in the dataset, the breach variables in the dataset should be seen as indicative measures. More detail is available in Appendix C of the OMCCS report 'Implementation of Community Orders',[15] which includes information that can be used to improve the measure in the future.

---

[14] 1,560 records were deleted – seven without a contact date and 1,553 without a contact type.
[15] Cattell, J., Mackie, A., Capes, T. and Lord, C. (2013) *Implementation of Community Orders: Results from the Offender Manager Community Cohort Study.* London: Ministry of Justice.

Again, using the case reference number, breach data were linked for:

- 5,994 offenders in the local administrative data; and
- 588 offenders in the survey sample.

Reoffending data (PNC)

A record for each reoffence by offenders in the full cohort was provided by the MoJ. In total, 126,673 offenders were matched to the PNC (88% match rate), using PNC number (when that was not available the name and date of birth were used), index offence and index commencement date. More than one reoffence can be committed on one day; in such cases the offence receiving the most severe sentence is defined as the 'main reoffence'. All reoffences are presented in the dataset and the main reoffence is marked accordingly.

PNC data were linked for:

- 126,673 offenders in the central administrative data;
- 44,000 offenders in the local administrative data; and
- 2,336 offenders in the survey sample.

3.3   How the data were linked for the Universal Dataset

The data collected were stored and linked in an SQL Server database, which had the necessary storage and data processing capabilities. A unique identifier was created for each offender to link the various datasets collected. This identifier was created by combining the case reference number used to identify individual offenders (CRN), the Probation Trust code and the commencement date.[16]

In preparation for merging the survey data with the commencements data, a unique offender identifier was also created by using the CRN and area codes. Only those offenders who gave their consent had their survey data matched into the Universal Dataset. Of the 2,919 offenders interviewed for the Wave 1 survey, 2,595 (89%) gave permission for their survey responses to be linked to the administrative data and 1,640 (56% of the total interviewed in Wave 1) also gave permission to link their Wave 2 or 3 responses.[17]

---

[16]   For example, the CRN *A000000*, the Probation Trust *99* and the commencement date *2009-10-01* would create the unique identifier *A000000-99-2009-10-01*.

[17]   Of the 1,640 offenders who answered more than one wave of the survey, and gave permission to have their responses linked to the administrative data, a match rate to PNC data of around 90% was achieved (1,496 offenders).

3.4   Data checks

Data checks were undertaken on the Universal Dataset following the linking of individual datasets to ensure it was complete and accurate. The data checks that were carried out for each data source are described below:

The Universal Dataset:

- All commencement CRN numbers were complete and unique.
- Commencement dates were between 1 October 2009 and 31 December 2010.
- Non Community Order commencements were removed.
- All CRN numbers had seven digits.
- All US-formatted dates in the commencements, OASys, IAPS and terminations data were converted to UK formats.

OASys data:

- Duplicate assessments were removed from the dataset.
- Assessments with an unknown purpose code were removed from the dataset.
- All first assessment dates were within 90 days of the commencement of the Community Order.
- All termination assessment dates were within 90 days of the termination of the Community Order.

IAPS data:

- The data were linked for individuals with an accredited programme in their Community Order. When this was not the case, it was flagged in the dataset.

PNC data:

- All offence dates were after the Community Order commencement date.

Survey data:

- The date of the commencement of the Community Order given in the survey agreed with the Form 20 commencement date. When this was not the case it was flagged in the dataset.

Local administrative data:

- All contacts had a date and contact type.

- All breaches were coded correctly.

3.5   Final structure of the Universal Dataset

Once the data were linked in the database, a version of the Universal Dataset was built that was complete but also designed to be uncomplicated so future users could engage with the data relatively easily. A key consideration was to allow users to trace an offender through the different data sources so they understood how to take full advantage of the data available. The dataset was therefore presented in seven different tables, which are explained in Table 3.1 below. The first three tables represented the three cohorts described at the beginning of this chapter (full cohort, local cohort and survey sample cohort), and tables 4–7 provide local contacts, reoffence records, individual OASys assessments, and data on contacts respectively. Table 3.1 describes the tables that make up the SQL server database.

Table 3.1: Structure of the SQL Server database

Name of table	Description	Size of table	Sources
1. Central	All data derived from central administrative data sources for all offenders	144,407 records 922 variables	Form 20 OASys IAPS PNC
2. Local	All data from the *1. Central* table plus the local administrative data collected from 10 probation trusts	50,935 records 1,050 variables	Form 20 OASys IAPS PNC Contacts data Breach data
3. Survey[18]	All data available in the *1. Central* and *2. Local* tables data plus data from the three survey waves for those offenders who gave permission to link their survey responses to the administrative data	2,595 offenders 8,421 variables	Form 20 OASys IAPS PNC Contacts data Breach data Survey Wave 1 Survey Wave 2 Survey Wave 3
4. Local contacts	Each individual contact including date and type of contact and information from the free text field	2,285,077 records 9 variables	Contacts data

---

[18]   3. Survey is an SPSS file and not contained in the SQL server database. It is described here for completeness.

Name of table	Description	Size of table	Sources
5. Reoffences	Each reoffence including date and type	337,489 records 5 variables	PNC
6. OASys assessments	Each OASys assessment during the period of the Community Order (plus and minus 90 days)	440,134 records 483 variables	OASys
7. Data on contacts	The open text field where details of each contact were recorded. These data are in a separate table from the rest of the contact records (table 4) because of the size of the data.	2,264,545 records 2 variables	Contacts data