Home Office
BUILDING A SAFE, JUST AND TOLERANT SOCIETY

Findings 170

The Police National Computer and the Offenders Index: can they be combined for research purposes?

Brian Francis, Paul Crosland and Juliet Harman

The current standard research tool used in studying patterns of offending and reconviction in England and Wales is the Offenders Index. It has been suggested that a 'complete criminal record' should be created using information from the police, the courts, the prison service and the probation service (Allnutt, 2001). This study investigates the feasibility of merging relevant police records held on the Police National Computer into extracts from the Offenders Index in order to maximise the information available to researchers. However, external researchers are not given access to individual data as they are made anonymous.

Key points

- Research and evaluation studies would be enhanced if the currently used Offenders Index (OI) was augmented with additional information from the operational Police National Computer (PNC).
- Records from the OI and the PNC data sources were matched at the individual and court date levels for offenders in five research studies.
- The matching process was carried out for over 18,000 individuals using purpose-written software. Records for 92% of these individuals were found in both the PNC and OI data and 91% of individuals were found to be matches (or partial matches).
- Automatic matching of court dates by the purpose-written software was supplemented by manual intervention to match records with similar dates, using summary offence information, and to identify composite OI records. There were more than 178,000 court dates in the merged files with a 71% match (including 2% matched manually).
- The results of the matching process were used to propose an automatic matching algorithm to link court date level records with minimal intervention from the user.

The standard research tool used for conviction studies is the Offenders Index (OI). This is a database intended to contain all court disposals relating to standard list offences since 1963 in England and Wales. The Police National Computer (PNC) is an operational policing database for the UK (excluding Northern Ireland), which contains additional information, particularly on cautions, warnings and dates of offence. It is not suitable for research into criminal histories, as 'weeding' of records takes place periodically. Police records from the period prior to its launch in 1995 are being computerised and integrated into the PNC database when more recent convictions are recorded.

Work by Friendship et al. (2001) compared these two sources for a small sample of 134 offenders and concluded that a merged database is most valuable for researchers.

The views expressed in these findings are those of the authors, not necessarily those of the Home Office (nor do they reflect Government policy)

The Research, Development and Statistics Directorate exists to improve policy making, decision taking and practice in support of the Home Office purpose and aims, to provide the public and Parliament with information necessary for informed debate and to publish information for future use.

Findings are produced by the Research, Development and Statistics Directorate. For further copies contact:

Communication Development Unit
Room 275,
Home Office,
50 Queen Anne's Gate,
London SW1H 9AT.

Tel: 020 7273 2084
Fax: 020 7222 0211
publications.rds@homeoffice.gsi.gov.uk

Editor: Carole Byron
Printed by: TABS

ISSN 1473-8406

The merits of the two record systems as research tools are compared in Tables 1 and 2.

The research study

These Findings focus on two aims of the study:

- for selected subgroups of cases, to match OI data with PNC data and to assess rates of matching
- to develop a strategy whereby, for any set of offenders, the two sets of information could be merged into a single data source.

Five research datasets were examined in detail. For some of the datasets, names were collected from an external agency (such as the probation or prison service) and the names tracked on both the PNC and OI. For other datasets, offender names were collected from one source and traced on the other.

The OI and PNC data are recorded at a number of levels. At the individual level, they contain personal information including name, date of birth, gender and Criminal Record Office number. Within each individual record there is a list of court appearances in date order. Court date level information includes police and court codes. Finally, each court appearance will have a list of offences. Offence level information includes the offence code, the plea and the court disposal. The PNC data contain offence dates, crime location, co-offender details, cautions, warnings and impending prosecutions, none of which is available on the OI.

Table 1 Advantages and disadvantages of the PNC as a research tool

Advantages
Includes Scotland, England, Wales and British Transport Police
Complete history for older offenders if back records are converted
Information usually available more promptly than OI but some delays in data entry
Criminal histories built up by fingerprint verification
Information on cautions, warnings and impending prosecutions available
Dates of offence available
All offences recorded
Postcode information on crime location available
Disadvantages
Excludes Northern Ireland
Cannot search the database directly
Offence codes and disposal codes differ from the standard codes used by the Home Office Research, Development and Statistics Directorate
Disposal information of poor quality
Criminal histories may be split
Criminal histories 'weeded' – less important offences removed
Criminal histories deleted on death

This research focused on matching records at the court appearance level, as the PNC and OI use different coding systems to identify offences and disposals. Additionally, these codes have changed over time and the currently available conversion routines are not reliable enough to ensure accurate matching at the offence level.

Purpose-written software was developed for matching individual and court date level records, allowing manual intervention where necessary. The results of the matching process were used to develop an automatic matching algorithm.

Matching process

For each study, the PNC and OI files were 'levelled' to ensure that certain types of information present on the PNC but not on the OI were removed before comparing records. This included non-standard list offences, cautions, warnings, reprimands and impending prosecutions, convictions before 1964 and convictions outside England and Wales.

Using the purpose-written software the individuals in each research dataset were matched with the OI and PNC records on personal details such as name, date of birth and gender.

If the personal details agreed (or partially agreed), then the level of matching at the court date level was tested. Two conviction records were automatically accepted as belonging to the same person if the level of matching was high.

Court dates were sometimes matched manually if they were close and other details, such as court and police identifiers together with summaries of the number of offences at each court date for each of ten offence groups, agreed. Matches were, on occasion, 'partial', if the OI record appeared to contain composite information on more than one individual.

Table 2 Advantages and disadvantages of the OI as a research tool

Advantages
Complete conviction history for all offenders since 1963
Information on conviction dates, offences, disposals and courts
Can search OI by name, by offence code etc.
Subset of OI data (individuals are made anonymous) available at ESRC data archive
Disadvantages
Excludes Scotland and Northern Ireland
Not a complete history for older offenders
Delay in collection and processing of information
Criminal histories may contain composite information on more than one individual
Only standard list offences recorded
No cautions, warnings
No dates of offence – problem of pseudo-reconvictions

Table 3 Example of output from the matching software

Police National Computer data							Offenders Index data					
Record No.	Court code	Police code	Offences		No.of offences	Date	Court code	Police code	Offences		No.of offences	Date
1	1249	17	1t		1	12 Nov 1982	1249	17	1t		1	12 Nov 1982
10	9998	17	1b		1	1 Feb 1988	460	17	1b		1	1 Feb 1988
17							1249	17	1v		1	9 Mar 1998
18	1249	17	1v	1o	2	19 Mar 1998						
19	1249	17	1v	1o	2	17 Apr 1998	1249	17	2v		2	17 Apr 1998

Notes: 1. Offence codes are v = violence, b = burglary, t = theft, o = other. 2. Only 5 court dates shown from a record of 21 entries; missing court code indicated by 9998.

Results

The matching process was carried out for over 18,000 individuals over the five research datasets. Records for 92% of these individuals were found in both the PNC and OI data, and the records of 91% of individuals were found to be matches (or partial matches).

Table 3 shows the partial record of an individual produced by the matching software, with three examples of matching court dates and two unmatched court dates that may be manually matched. The court dates from the PNC and OI data are listed with information on court and police codes, numbers of offences in ten offence categories (e.g. violence, burglary, theft and criminal damage) and total number of recorded offences. The details may not always match. There were 21 conviction dates for this offender and 17 (80%) were common to both files. In addition to the matched dates, unlinked records likely to refer to the same conviction, such as lines 17 (OI) and 18 (PNC), which differ by only one digit in the day of conviction may be manually matched.

Figure 1 Sources of court dates in the merged files

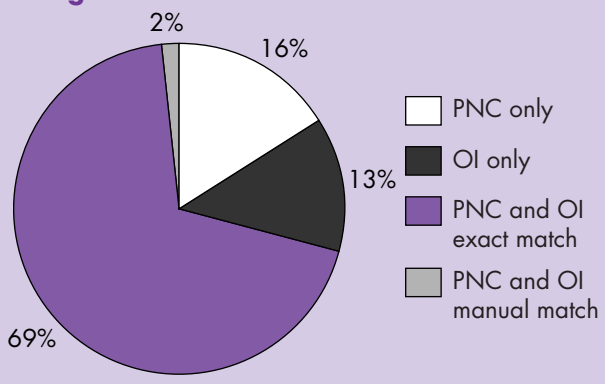
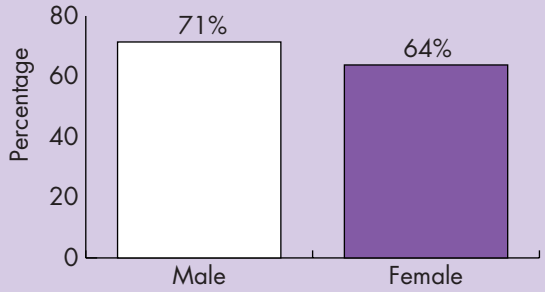


Figure 2 Match rates for court dates by gender



There were more than 178,000 court dates in the merged files covering the five studies, of which 71% were found to be matches (Figure 1).

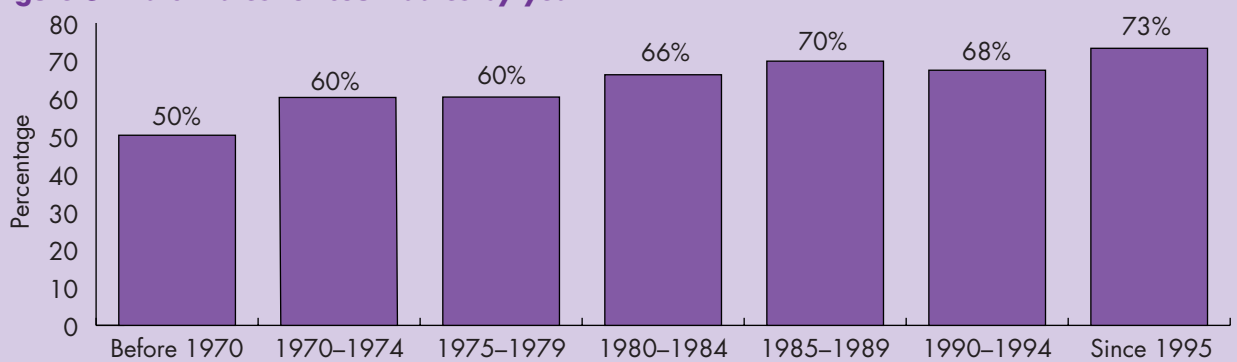
The difference in the rates of matching for men (71%) and women (64%), illustrated in Figure 2, was statistically significant.

For most of the studies the match rate gradually improved over time. The overall match rate increased from 50% for the period before 1970 (with the two data sources contributing approximately equal numbers of unmatched records), to 73% for 1995 onwards, as shown in Figure 3.

Court date match rates by police authority after 1974 for over 15,000 individuals in the three larger studies were examined. They showed low rates of matching (averaging around 65%) for the Metropolitan Police area and for the City of London compared to rates of 80–90% for many Northern, Midlands and South Western forces.

The average number of offences contributed by the two data sources for each matched court date was compared. This indicated that before 1990 the PNC contributed a larger number of offences than the OI, and that from 1990

Figure 3 Match rates for court dates by year



onwards this pattern has been reversed. It must be noted that this finding is after 'levelling' the PNC and OI data so that both contained only standard list offences. The full PNC data also includes non-standard list offences and will therefore contain more offences.

The rates of matching for a number of common family names was often poorer than the rate of matching over all family names. One reason for this may be that the OI record for common family names is more likely, than for less common names, to consist of composite individuals formed in error into a single record.

Automatic matching

Using logistic regression on a sample of results from the matching process on 3,000 individuals with over 36,000 court dates (Copas and Hilton, 1990), a matching score was developed to link court level records from the PNC and OI. The matching score contained 13 measures (Table 4) of potential discrepancy between the OI court level record and the PNC record. Variables measuring agreement made a positive contribution to the score, while variables measuring difference made a negative contribution. Greater discrepancy leads to a lower score – scores above zero (corresponding to a probability of 0.5) for PNC-OI record pairs are accepted as matches.

Table 4 Variables used in the matching score

Agreement on:	Day of conviction	Yes/no
	Month of conviction	
	Court code	
	Police code	
	No. of violent offences	
	No. of burglary offences	
	No. of robbery offences	
	No. of theft offences	
	No. of fraud offences	
	No. of criminal damage offences	
Absolute difference in no. of other offences		Count
Absolute difference in year of conviction		Count
Same court building (youth or adult)		Yes/no

The matching score performed well on a validation sample from the same dataset, with only 0.35% of records mismatched when compared to the 'true' results obtained from the earlier matching process. The matching score also performed well on a validation sample from a different dataset, with only 0.40% of records mismatched when compared to the 'true' results.

Based on these results an automatic matching algorithm has been proposed which will have minimal user intervention.

Main recommendations

Automatic matching

Additional research should be carried out to further develop, test and evaluate the proposed automatic matching of records at the court date level and the automatic detection of composite OI criminal histories which need splitting. Purpose written software capable of matching large numbers of criminal records together efficiently should be developed.

Merging

Work should be undertaken to incorporate all types of disposal and all types of offence from PNC records into the Offenders Index. This includes:

- adding summary PNC information to existing OI court date records
- adding new court dates to OI records, including cautions/warnings/reprimands, impending prosecutions and also court disposals not known to the OI
- creating new records from PNC information for individuals not traced on the OI.

Coding

Reliable code conversion routines should be developed to translate the different and changing offence and disposal codes used by the PNC and the OI.

Research should then follow to review the robustness of the offence/disposal level match, using similar techniques to those developed here.

References

Allnutt, D (2001). *Review of statistics on efficacy of sentencing*. Office of National Statistics, London. http://www.statistics.gov.uk/themes/yourviews/downloads/postconsdrft.pdf

Copas, J.B. and Hilton, F.J. (1990). Record linkage: Statistical models for matching computer records. *Journal of the Royal Statistical Society Series A*, 153, 3, 287–320.

Friendship, C., Thornton, D., Erikson, M. and Beech, A. (2001). Reconviction: A critique and comparison of two main data sources in England and Wales. *Legal and Criminological Psychology*, 6, 1, 121–129.

For a more detailed report, see *The Police National Computer and the Offenders Index: can they be combined for reseach purposes?* by Brian Francis and Paul Crosland (2002) on the Home Office website: www.homeoffice.gov.uk/rds/rfpubs1.html

Brian Francis is Director and Juliet Harman is a Research Associate of the Centre for Applied Statistics, Lancaster University. Paul Crosland was, at the time of the research, a Research Associate at the Institute of Criminology, Cambridge University.