

**DEVELOPING A REOFFENDING
DATABASE:
Design, processes, problems
and solutions**

Jiuzhao Hua and Bronwyn Lind

OESR Criminal Justice Modelling Workshop

11- 12 November 2003, Brisbane

Outline of presentation

- How did we develop our database?
 - Design and structure
 - Matching process: problems and solutions
- Some interesting statistics from the database

How did we develop our database?

- Set up a relational database using SAS
- Used existing databases of court appearance records from
 - Local Courts 1995 -- 2002
 - Higher Courts 1995 -- 2002
 - Children's Courts 1991 -- 2002
- Linked court appearance records for the same person by matching on identifying information:
 - name
 - date of birth
 - CNI (Central Names Index)
 - juvenileid (juvenile identifier)

Design and Structure: Tables and Variables

PERSON: master ID, DOB, gender, ATSI, age and postcode at first appearance

PSNDTAL: master ID, alias ID, CNI, juvenileid, surname, first name, DOB

JUVENILE: juvenileid, alias ID, surname, firsts name, DOB

CRT_APR: case ID, alias ID, jurisdiction, courtno, date of finalisation, bail status, conviction indicator, prison indicator, prior conviction counter

ADDRESS: case ID, postcode, LGA, statistical division

Design and Structure: Tables and Variables -- continued

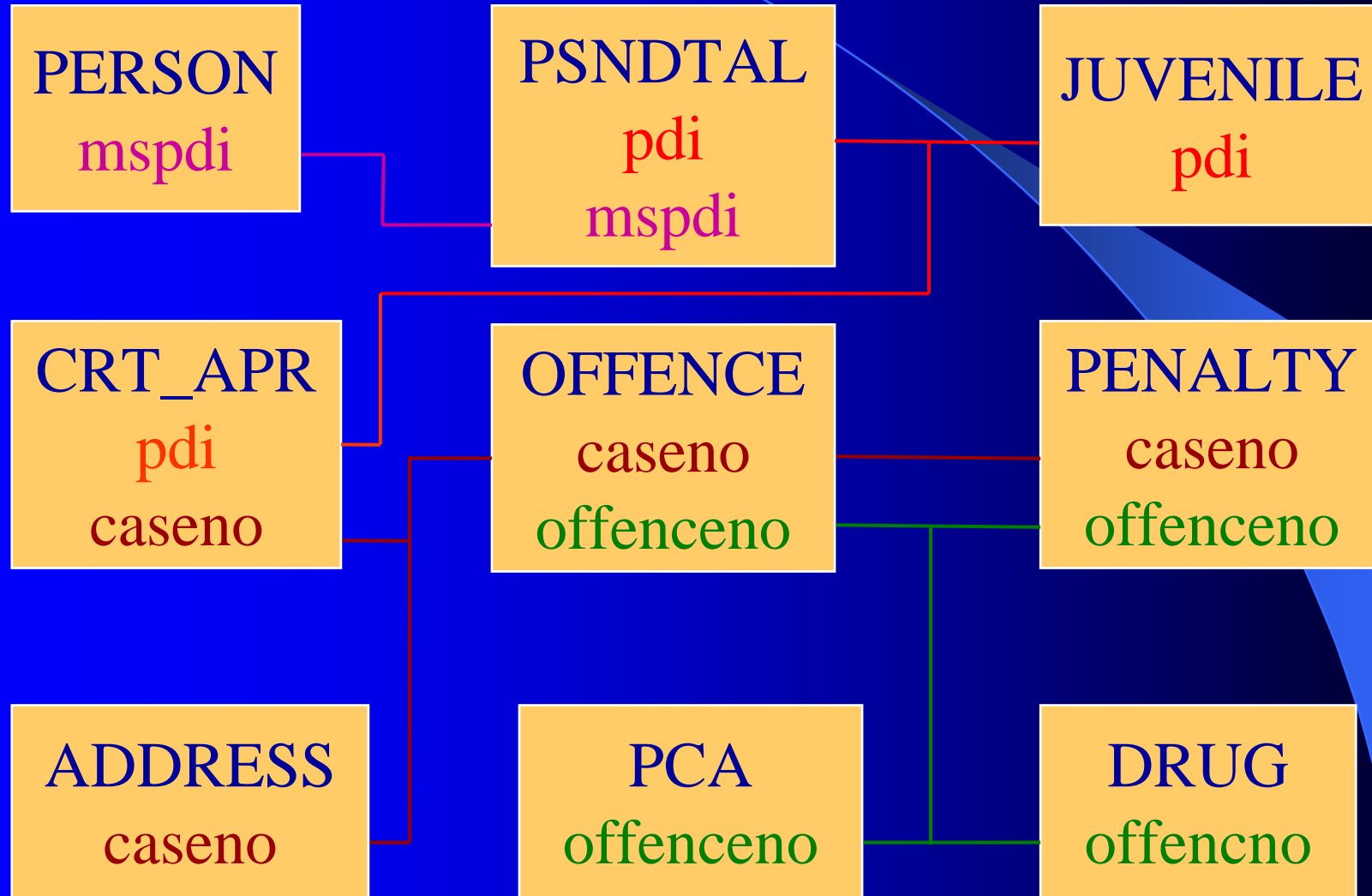
OFFENCE: offence ID, case ID, lawpart code,
ASOC, offence date, plea, outcome,
principal offence indicator

PENALTY: offence, case and penalty IDs
type and amount of penalty,
principal penalty indicator

DRUG: offence ID, drug type, drug amount

PCA: offence ID, blood alcohol content, range

Design and Structure: Links between tables



Matching Process: Problems and Solutions

- **Identifying variables:**

CNI, surname, first name, middle name, DOB, juvenileid

- **Basic matching criteria:**

surname, first name, DOB

surname, first name, middle name, 2/DOB

CNI, DOB

CNI, surname, 2/DOB

CNI, first name, 2/DOB

CNI, surname, firstname

juvenileid

Variations in dates of birth -- examples

- **Swapping day and month**

CNI	surname	first name	DOB
3546092	Bayley	David	05/08/1972
3546092	Bayley	David	08/05/1972

- **Swapping the two digits in day**

1234122	Rudson	Mathew	12/09/1972
1234122	Rudson	Matthew	21/09/1972

- **Swapping the two digits in month**

56432302	Rodgers	Wayne	15/10/1976
56432302	Rodgers	Wayne	15/01/1976

Variations in dates of birth -- continued

- Swapping the last two digits in year

1208432	Downey	Dylan	15/06/1967
1208432	Downey	Dylan	15/06/1976

- Swapping the last digits in day and month

80123489	Dawson	Trevor	13/06/1964
80123489	Dawson	Trevor John	16/03/1964

- Swapping the first digits in day and month

3546092	Marsh	Jesse Lee	05/12/1967
3546092	Marsh	Jesse	15/02/1976

Variations in dates of birth - solutions

Define derived variables from DOB, then match on
surname, first name, derived DOB

- ‘swap_ddmm’, place ‘day’ and ‘month’ in numeric order,
eg., ‘12/09/1965’ and ‘09/12/1965’ → ‘09121965’
- ‘swap_day’, place the digits in day in numeric order,
eg., ‘12/09/1965’ and ‘21/09/1965’ → ‘12091965’
- ‘swap_month’, place the digits in month in numeric order,
eg., ‘14/10/1972’ and ‘14/01/1972’ → ‘14011972’

Variations in dates of birth – solutions -- continued

- ‘**swap_year**’, place the last two digits in year in numeric order,
eg., ‘14/10/19**65**’ and ‘14/10/19**56**’ → ‘141019**56**’
- ‘**swap_dmdm**’, place the last digits in day and month in numeric order,
eg., ‘**15**/07/1965’ and ‘**17**/05/1965’ → ‘**1507**1965’

Variations in CNI -- examples

- Swapping two adjacent digits

3546092	Wilson	Daniel	18/08/1972
3546902	Wilson	Daniel	08/08/1972

- Dropping a digit or adding a digit

123422203	Hookey	Vincent	17/09/1972
12342203	Hookey	Vincent	17/08/1972

- Replacing one digit by another

56432302	Richards	Witney	15/10/1968
57432302	Richards	Witney	15/10/1967

Variations in CNI - solutions

- Dropping one digit at a time, then match on

CNI, surname, first name, 2/DOB

Eg., dropping '0' will allow the following to match

3546092	Walford	Fred	18/08/1972
3546902	Walford	Fred	08/08/1972

- Replacing one digit by another, then match on

CNI, surname, first name, 2/DOB

Eg., replacing '6' by '7' will allow the following to match

56432302	Richards	Witney	15/10/1965
57432302	Richards	Witney	15/10/1966

Variations in names -- examples

- Dropping a letter or adding a letter

20945214	Rodgers	Dean	22/08/1972
	Rogers	Dean Andrew	22/08/1972

- Swapping two adjacent letters

1234122	Fraser	Jeffery	17/09/1972
1234122	Fraser	Jeffrey	17/09/1972

- Similar sound but different spelling

56432302	Bright	Betty	15/10/1959
56432302	Bright	Bettie	15/10/1959

- Replacing one character by another

20945214	Whalan	Lyden	22/08/1972
	Whalen	Liden Andrew	22/08/1972

Variations in names -- continued

- Misinterpreting handwritten letters

20945214	OHANIAN	CAGLAR	22/08/1972
20945214	OHANLAN	CACLAR	22/08/1972

- Letters disordered

1234122	Kalaitzis	Mathew	17/10/1972
1234122	Kalaitsiz	Mathew	17/01/1972

- Shuffling words around

46187341	Van Deven	Thimy	28/03/1958
46187341	Deven Van	Thi My	28/03/1958

Variations in names - solutions

- Dropping one letter at a time, eg.,
dropping 'R': 'Berlinda', 'Belinda' → 'Belinda'
- Replacing one letter by another (for common variations), eg.,
replacing 'E' by 'A': 'Mohamed', 'Mohamad' → 'Mohamad'
- Derived name: put words in alphabetical order
'Van Deven', 'Deven Van' → 'Deven Van'
- Defining soundex: group letters and assign a numerical code for each group, then shrink double codes.
Eg, if 'I', 'E' and 'Y' are in the same group, then the followings
have the same soundex:
'Betty', 'Bettie', 'Betti', 'Bettey', 'Betic', 'Bette', 'Bietty'

Soundex examples

Version	Names that have the same soundex as 'michael'	Code
soundex 1	michaael, michael, michaell, michail, mychael	M82B08C
soundex 2	miccheal, michaael, michaeal, michael, michaell, michal, micheal, michel, michell, mickhael, mikhael, mychael, mychal	M73608
soundex 3	michaael, michael, michaell	M827039
soundex 4	machiel, maichael, maichel, malachie, micahael, micaheal, micahel, miccheal, michaael, michaeal, michael, michaela, michaele, michaelea, michaeli, michaell, michaella, michaelle, michale, michcael, micheal, micheala, micheale, michealla, michela, michella, mihcael	MACEHIL
soundex 5	machaal, machiel, maichael, maichel, mchael, mechele, micahael, micaheal, micahel, michaael, michaeal, michael, michaela, michaele, michaelea, michaeli, michail, michal, michale, micheal, micheala, micheale, michel, michela, michele, michial, michile, mychael, mychal	MCHL

Matching with popularity checking

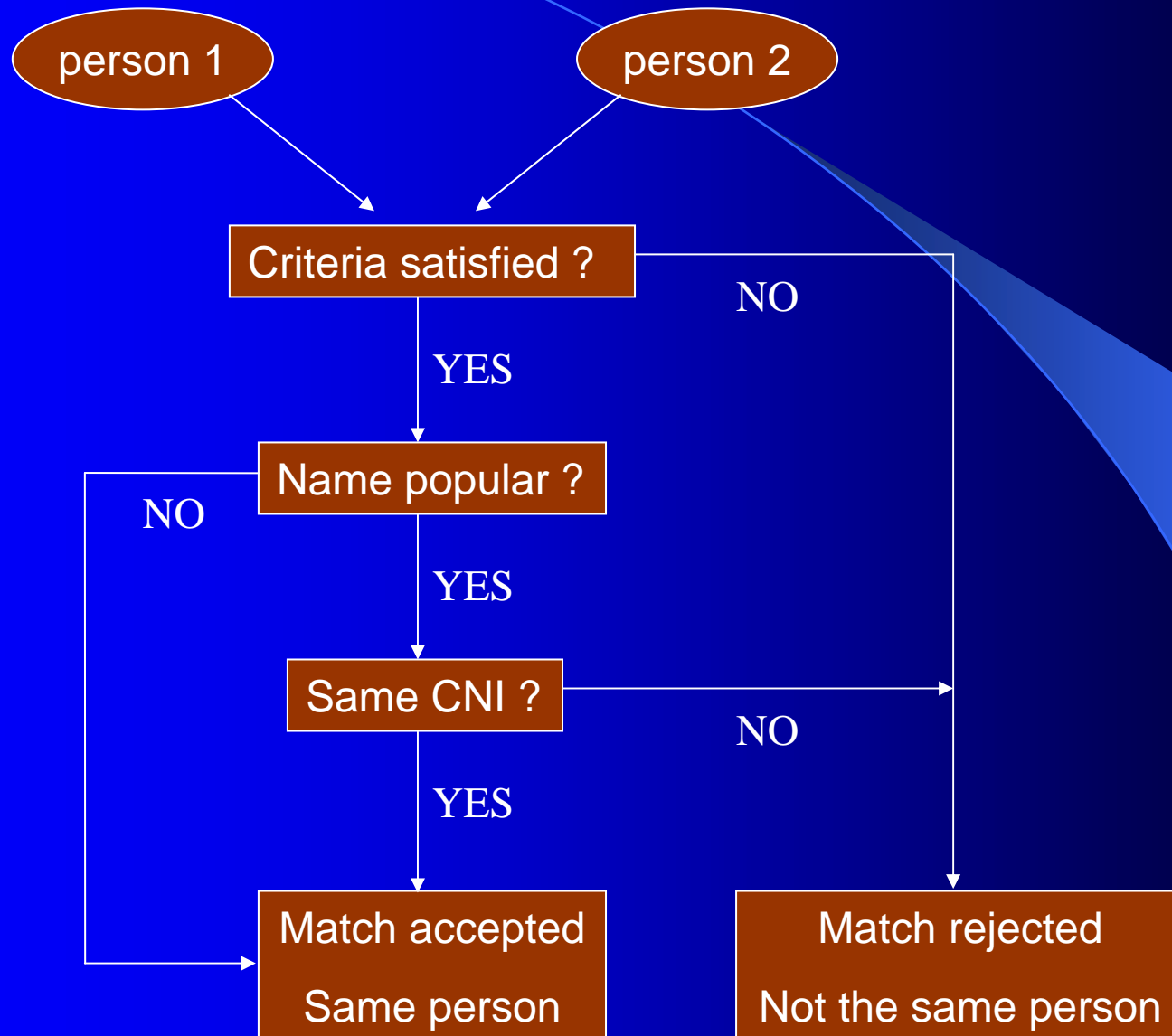
- Mismatch is more likely to occur where the name is a common one, such as 'SMITH' and 'MICHAEL'.
- Popularity of names is derived from the frequency of the name appearance in the database. (There are about 500,000 unique persons in the database.)
- Each criterion involving DOB or its variant is modified. A popularity 'check mark' is assigned.
- Matches with popular names are rejected unless both persons have same CNI or JUVENILEID.

Facts on probability

Number of distinct persons with same name combination (N)	Prob. all persons having distinct DOB	Prob. at least two persons having same DOB	Popularity of the name combination (500,000 / N)
5	0.998631	0.001369	100,000
12	0.990995	0.009005	41,667
27	0.953000	0.047000	18,519
50	0.845194	0.154806	10,000
100	0.506017	0.493983	5,000

Assuming all persons were born within a period of 20 years

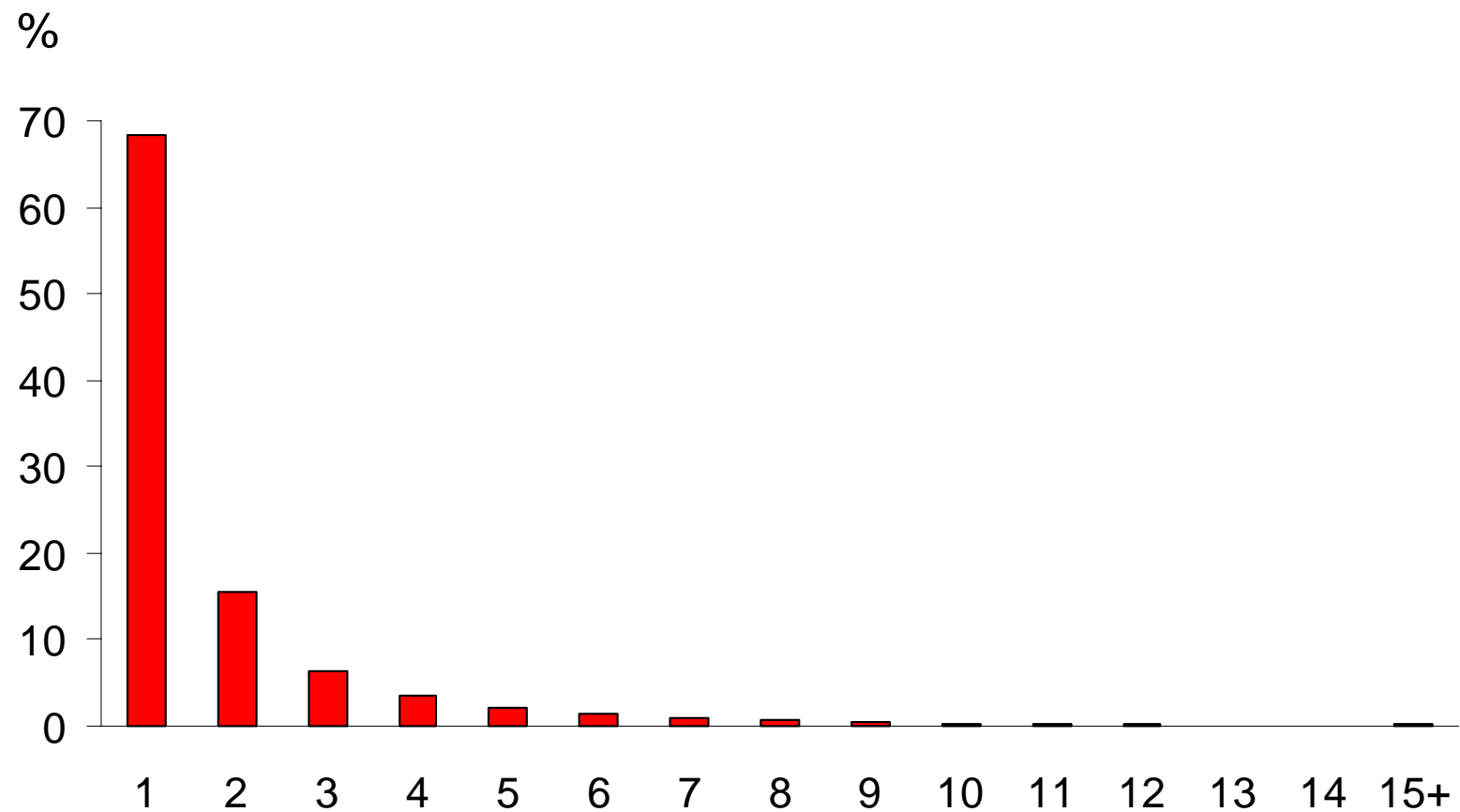
Popularity checking -- logical flow diagram



An example of a person after all matching

CNI	Juvenileid	Surname	Firstname	DOB
	203948	Kalaitzis	Mathew Henry	17/10/1978
	203948	Calaitzis	Matthew	17/10/1978
		Kalaitzis	Mathew Henry	17/11/1978
		Kalaitzis	Matthew Henry	17/10/1978
		Mathew	Kalaitzis Henry	17/10/1978
1234122		Kalaitzis	Mathew Henry	17/10/1978
1234122		Calaitzis	Mathew	07/10/1978
123412		Kalaitzis	Mathew	17/01/1978

Frequency distribution of number of criminal court appearances, NSW, 1997-2001



Proportion of NSW population appearing in court in 2001 (about 103,000 people with a court appearance in 2001)

- 1.9% of all NSW residents aged 10+
- 5.1% of all NSW residents aged 20-24
- 8.4% of all male NSW residents aged 20-24
- 12.8% of all Indigenous NSW residents aged 10+
- 27.7% of all Indigenous NSW residents aged 20-24
- 41.3% of all Indigenous male NSW residents aged 20-24

Thank you !