



**Australian Government**

**Australian Institute of  
Health and Welfare**

# **Developing a linked data collection to report on the relationships between child protection and youth justice supervision**

DATA LINKAGE SERIES NO. 20





Australian Government

Australian Institute of  
Health and Welfare

*Authoritative information and statistics  
to promote better health and wellbeing*

DATA LINKAGE SERIES

Number 20

# **Developing a linked data collection to report on the relationships between child protection and youth justice supervision**

Australian Institute of Health and Welfare  
Canberra

Cat. no. CWS 55

**The Australian Institute of Health and Welfare is a major national agency which provides reliable, regular and relevant information and statistics on Australia's health and welfare. The Institute's mission is authoritative information and statistics to promote better health and wellbeing.**

© Australian Institute of Health and Welfare 2015



This product, excluding the AIHW logo, Commonwealth Coat of Arms and any material owned by a third party or protected by a trademark, has been released under a Creative Commons BY 3.0 (CC-BY 3.0) licence. Excluded material owned by third parties may include, for example, design and layout, images obtained under licence from third parties and signatures. We have made all reasonable efforts to identify and label material owned by third parties.

You may distribute, remix and build upon this work. However, you must attribute the AIHW as the copyright holder of the work in compliance with our attribution policy available at <[www.aihw.gov.au/copyright/](http://www.aihw.gov.au/copyright/)>. The full terms and conditions of this licence are available at <<http://creativecommons.org/licenses/by/3.0/au/>>.

This publication is part of the Australian Institute of Health and Welfare's data linkage series. A complete list of the Institute's publications is available from the Institute's website <[www.aihw.gov.au](http://www.aihw.gov.au)>.

ISSN 2205-5045 (PDF)

ISSN 1833-1238 (Print)

ISBN 978-1-74249-833-1 (PDF)

ISBN 978-1-74249-834-8 (Print)

### **Suggested citation**

Australian Institute of Health and Welfare 2015. Developing a linked data collection to report on the relationships between child protection and youth justice supervision. Data linkage series no. 20. Cat. no. CWS 55. Canberra: AIHW.

### **Australian Institute of Health and Welfare**

Board Chair

Dr Mukesh C Haikerwal AO

Acting Director

Ms Kerry Flanagan PSM

Any enquiries about copyright or comments on this publication should be directed to:

Digital and Media Communications Unit

Australian Institute of Health and Welfare

GPO Box 570

Canberra ACT 2601

Tel: (02) 6244 1000

Email: [info@aihw.gov.au](mailto:info@aihw.gov.au)

Published by the Australian Institute of Health and Welfare

This publication is printed in accordance with ISO 14001 (Environmental Management Systems) and ISO 9001 (Quality Management Systems). The paper is sourced from sustainably managed certified forests.



**Please note that there is the potential for minor revisions of data in this report. Please check the online version at <[www.aihw.gov.au](http://www.aihw.gov.au)> for any amendments.**

# Contents

- Acknowledgments..... iv
- Abbreviations..... v
- Summary ..... vi
- 1 Introduction.....1**
  - 1.1 Children and young people receiving child protection services.....1
  - 1.2 Young people under youth justice supervision.....2
  - 1.3 What are the relationships between abuse or neglect and criminal activity? .....3
- 2 Linking data on child protection and youth justice supervision.....5**
  - 2.1 Key-based linkage.....5
  - 2.2 Multi-step linkage process.....8
- 3 Exploring relationships between child protection and youth justice supervision.....15**
  - 3.1 Young people in both sectors in the same financial year .....15
  - 3.3 Pathways through child protection and youth justice.....16
  - 3.4 Data availability and reporting.....16
  - 3.5 Linkage with other data collections .....19
- Glossary.....21**
- References.....22**
- List of tables .....23**
- List of figures .....24**
- Related publications .....25**

# Acknowledgments

Rachel Aalders wrote this report. Kristy Raithel, Tim Beard and Justine Boland provided guidance and support.

Contributions, comments and advice from the Juvenile Justice Research and Information Group, which consists of representatives from the state and territory departments responsible for youth justice supervision, and the Children and Families Data Network, which consists of representatives from the state and territory departments responsible for child and family services, are gratefully acknowledged.

# Abbreviations

AIHW	Australian Institute of Health and Welfare
CP NMDS	Child Protection National Minimum Data Set
JJ NMDS	Juvenile Justice National Minimum Data Set

# Summary

Research shows that children and young people who have experienced abuse or neglect are more likely to commit offences than those who were not abused or neglected, and a high proportion of those in youth justice supervision have a history of abuse or neglect. Understanding the characteristics of children and young people who are both in the child protection system and under youth justice supervision would assist support staff, case workers and policy makers to achieve optimal outcomes for children and young people, and for their families.

Currently, there are no national data on children and young people who are both in the child protection system and under youth justice supervision. An earlier Australian Institute of Health and Welfare (AIHW) project (AIHW 2012a) found that linking child protection and youth justice supervision data was both feasible and beneficial. With the introduction of a national unit-record child protection data collection, this is now possible. This report describes how these data collections can be linked and how the relationships between child protection and youth justice supervision can be explored.

## **Linking data on child protection and youth justice supervision**

Because the child protection and youth justice supervision data collections do not contain full names or common person identifiers, the collections need to be integrated using key-based linkage. Key-based linkage allows existing data collections to be linked while protecting individual privacy. A multi-step process will increase the possibility of linking records belonging to the same person, while at the same time decreasing the possibility of linking records belonging to different people. This method allows records with missing data (such as unknown date of birth) and alternative information (such as alias names) to be linked.

## **Exploring the relationships between child protection and youth justice supervision**

A linked data collection will enable the relationships between child protection and youth justice supervision to be explored in two ways.

The first involves analysing the characteristics and pattern of events for children and young people who are involved in both child protection and youth justice supervision in a given financial year. This requires data from both collections for the same financial year for young people aged 10–17 (as youth justice supervision is not possible before the age of 10). This analysis is possible with the data currently available.

The second involves analysing the pathways that children and young people take through the two systems. With the data currently available, it is possible to analyse partial pathways. As data accumulate over time, more comprehensive longitudinal analyses will be possible.

# 1 Introduction

Every year, almost 100,000 children aged 0–17 have a notification of abuse or neglect investigated by a department responsible for child protection (AIHW 2015a) and almost 13,000 aged 10–17 are supervised by a department responsible for youth justice in the community or in youth detention centres (AIHW 2015b). Research shows that children and young people who have been abused or neglected are at risk of subsequently committing criminal activity and entering the youth justice system (Stewart et al. 2005).

Currently, there are no national data on the number of children and young people who are involved in both child protection and youth justice supervision. Developing a data collection that links existing child protection and youth justice data would allow the levels of dual involvement and the pathways through these two sectors to be analysed.

Measuring the extent of involvement in both child protection and youth justice supervision and analysing the pathways that these children and young people take would provide a number of benefits to policy makers and service providers, and to those who work directly with these children, young people and their families.

Previous work by the Australian Institute of Health and Welfare (AIHW) linking data on youth justice supervision with limited child protection data showed that this linkage is both feasible and beneficial (AIHW 2012a). Since the conclusion of this earlier work, a national data collection on children and young people receiving child protection services has been implemented in a format suitable for linkage. Developing and implementing a method to link these collections on an annual basis would enable ongoing analysis and reporting of the overlap between these two sectors.

Initial reporting would focus on the extent of involvement within a particular financial year, while more detailed analyses of the pathways through these sectors would be possible over time as more data become available. It would also be possible to integrate data on other sectors with the linked child protection and youth justice supervision data collection, either on an ad hoc or ongoing basis.

This report describes how this new national child protection collection could be linked with the existing youth justice supervision data collection to report on the relationships between child protection and youth justice supervision. This first chapter provides an overview of the children and young people receiving child protection services and youth justice supervision. It also presents an overview of the research on the relationships between child protection and youth justice supervision, and more broadly, between abuse or neglect and criminal activity. The second chapter outlines how the two data collections could be linked. The third chapter describes how the relationships between child protection and youth justice supervision could be explored and reported.

## 1.1 Children and young people receiving child protection services

In Australia, state and territory departments responsible for child protection provide assistance to vulnerable children and young people aged under 18 who have been, or are at risk of being, abused, neglected or otherwise harmed, or whose parents are unable to provide adequate care or protection (AIHW 2015a).

Children often first come into contact with a child protection department when a report of concern about the child is made, either by themselves, a family or community member, or a professional. Child protection departments screen these reports to determine if further action is required. In some situations, advice or referrals to secondary support services are provided, while in others, the report is investigated to ascertain whether the child has been, is being or is likely to be abused, neglected or otherwise harmed; if the investigation finds that this is the case, the notification is recorded as substantiated.

In some situations, the child protection department may apply to a court to place the child on a care and protection order (a legal order or arrangement that gives child protection departments some responsibility for a child's welfare) (AIHW 2015a). The level of departmental involvement varies depending on the type of order – see *Child protection in Australia 2013–14* (AIHW 2015a) for more details.

When parents are unable to provide adequate care or children require a more protective environment, the department may place children in out-of-home care, which is overnight care where the department has made or offered a financial payment to the carer (AIHW 2015a). Children can be placed in out-of-home care because they require a more protective environment; their parents are unable to provide adequate care; or because alternative accommodation is needed during family conflict. Out-of-home care is considered a last resort, consistent with the principle of keeping children with their families wherever possible. For more information on the child protection policies in the states and territories and their similarities and differences, see AIHW 2015a.

A child or young person is considered to have received child protection services when they have been the subject of an investigated notification, been placed on a care and protection order, or been placed in out-of-home care. In 2013–14:

- around 143,000 children and young people received 1 or more child protection services, which is almost 3% of the Australian population aged 0–17
- almost 100,000 children (2%) were the subject of an investigation of a notification that was received in 2013–14, and of these, around 41,000 were the subject of an investigation that was substantiated in 2013–14
- around 55,000 children (1%) were on care and protection orders
- just over 50,000 children (is just under 1%) were in out-of-home care (AIHW 2015a).

One-third (33%) of those receiving child protection services in 2013–14 were on both a care and protection order and in out-of-home care.

Emotional abuse and neglect are the most common types of substantiated abuse and neglect. For 40% of children and young people who were the subject of substantiated investigations in 2013–14, emotional abuse was the primary type of abuse or neglect, followed by neglect at 28%.

## **1.2 Young people under youth justice supervision**

In Australia, the states and territories are responsible for dealing with young people who have committed, or who are alleged to have committed, criminal offences. In all states and territories except Queensland, the youth justice system applies to children or young people who were aged 10–17 at the time of the offence. In Queensland, the youth justice system applies to those aged 10 to 16. Children aged under 10 cannot be charged with a criminal offence in any state or territory.

There are several stages in the youth justice system. The first involves apprehension and investigation by police. If charges are laid, a court will hear the criminal matter. Then, if the young person is proven guilty, the court will hand down a sentence. At each of these stages, the young person is either unsupervised in the community or supervised by a youth justice department. This youth justice supervision can occur either in the community or in a youth justice detention centre. When the supervision relates to the young person's sentence, it is termed sentenced supervision, while unsentenced supervision refers to any supervision that occurs before the young person is sentenced.

It is possible for a young person to be under multiple types of youth justice supervision in the same year. It is also possible for a young person to be under multiple types of supervision at the same time where these supervision orders relate to different court matters. For more information on youth justice supervision in Australia, see AIHW 2015b or <http://www.aihw.gov.au/youth-justice/>.

In 2013–14, almost 13,000 young people were under youth justice supervision during the year, which is around 0.5% of the Australian population aged 10–17 (AIHW 2015b). At some time during the year:

- around 11,400 were supervised in the community
- almost 5,500 were detained in youth justice centres
- almost 4,000 were both supervised in the community and detained at different times during the year.

### **1.3 What are the relationships between abuse or neglect and criminal activity?**

There is extensive research showing that children and young people who suffered abuse or neglect are more likely to engage in criminal activity than those who did not. One study found that being maltreated roughly doubles the probability of committing a crime (Currie & Tekin 2006). A 2012 AIHW linkage study found that, of those born in the financial year 1990–91 who had been the subject of one or more child protection notifications in Victoria, 8% also had some form of youth justice supervision in Victoria. The rate was highest for Indigenous males: almost two-thirds of Indigenous men who had had a child protection notification as a child or young person had also been under youth justice supervision at some time (AIHW 2012a).

Similarly, those who are in the youth justice system, particularly in detention, are highly likely to have had a history of abuse or neglect. A survey of young people in youth detention in New South Wales found that 81% of young women and 57% of young men had been abused or neglected, while 49% and 19%, respectively, had suffered 'severe' abuse or neglect (Indig et al. 2011).

Research has also highlighted factors that are predictive of later involvement in criminal activity. The age at maltreatment has been shown to be related to the likelihood of later involvement in youth justice, with one study showing that those whose maltreatment started in adolescence or persisted into adolescence were more likely to offend than those who were only maltreated when they were children (Stewart et al. 2005). The type of maltreatment has also been linked to later criminal activity, although the findings are inconsistent. Some research has found that those who are physically abused or neglected are more likely to be involved in the youth justice system than those who suffered other types of abuse or neglect

(Stewart et al. 2005), while other studies have shown that those who were neglected were more likely to be detained than those who were physically or sexually abused (Jonson-Reid & Barth 2000), and conversely that those who were sexually abused were most likely to go on to offend (Currie & Tekin 2006). Studies have also shown that children and young people placed in out-of-home care are around twice as likely to subsequently offend as those who received child protection services but were not placed in out-of-home care; however, being placed in out-of-home care may be a proxy for the frequency and severity of the abuse or neglect (Stewart et al. 2005; Widom 1989).

Children and young people who receive child protection services may also experience different levels of involvement in the youth justice system, irrespective of the type of offence. A study conducted in the United States found that children entering youth justice from child welfare were just as likely to have their first case dismissed as those who had not come from welfare. However, children entering from child welfare were less likely to receive probation and more likely to be placed in a detention centre or group home, even after controlling for the effects of variables such as sex, age, race and offence type (Ryan et al. 2007).

There are several possible reasons why children and young people who have been abused or neglected are more likely to commit offences. A key factor in young people's involvement in delinquent and criminal activity is lack of adequate supervision (Weatherburn & Lind 2006). Children and young people who have been abused or neglected may have parents or guardians who are unable to provide adequate supervision due to social and economic stress. In addition, young people who have been involved in the child protection system are more likely to be homeless (Johnson et al. 2010), have low levels of educational attainment and employment, and more likely to have drug and alcohol problems (Cashmore 2011), which may make them more likely to commit survival crimes such as theft.

While the rate of criminal activity is higher for those who have been abused or neglected, it is important to remember that most of the children and young people who receive child protection services do not go on to offend. For example, of children born in Queensland in 1983 and 1984, as many as 10% were the subject of suspected or substantiated abuse or neglect, and of these, around one-quarter (26%) went on to offend (Stewart et al. 2005).

## 2 Linking data on child protection and youth justice supervision

Information on children and young people receiving child protection services is available from a longitudinal person-based data set known as the Child Protection National Minimum Data Set (CP NMDS). This data set has records for each year from 2011–12 and contains information on the demographics of children and young people who receive child protection services, the details of the notifications received by child protection departments, and the child protection orders and out-of-home care placements relating to these children and young people.

Information on young people under youth justice supervision is available from a longitudinal person-based data set known as the Juvenile Justice National Minimum Data Set (JJ NMDS). This data set has records for each year from 2000–01 and contains information on the demographics of young people who are supervised by youth justice departments and the details of their unsentenced and sentenced supervision both in the community and in youth justice detention centres.

One method of investigating the relationships between child protection and youth justice supervision is to link these data collections at the client level. Neither collection contains full names; instead, children and young people are identified by a state-specific identifier (a string of letters and numbers). This identifier is not common to the child protection and youth justice departments within a state or territory, which means that if a person receives both child protection services and youth justice supervision, whether in a single state or territory or different states or territories, their records will have different person identifiers. The multi-step key-based linkage method (described below) aims to link records with a high degree of precision (by minimising incorrect links and maximising correct links) while maintaining the privacy of individuals in the collections.

### 2.1 Key-based linkage

Key-based linkage allows data collections without common person identifiers or full name information to be linked. The aim of key-based linkage is to minimise the likelihood of both false positives (where records that belong to different people are incorrectly identified as belonging to the same person) and false negatives (where records that belong to the same person are incorrectly identified as belonging to different people).

Linkage keys are formed using data items available in both collections. The child protection and youth justice supervision data collections share a number of data items that can be used to form linkage keys, including selected letters of name, date of birth, sex, Indigenous status and information about the suburb or town of their usual residence (complete address information is not available). Using linkage keys ensures the privacy of individuals is maintained, and it also reduces the burden on data providers because existing data collections can be used.

The simplest key-based linkage method is the simple deterministic method. This method, which requires relatively few resources, uses a single linkage key (for example, a string of letters from family name and given name and the date of birth) and considers records to belong to the same person where the linkage keys match exactly (see Box 2.1 for examples of

linkage results). This method works well in situations where the linkage key is adequately distinctive (that is, relatively few people share the same linkage key) and the components of the linkage key are recorded consistently. However, this method will not work as well in situations where the key components are not recorded consistently, either due to clerical errors (such as date of birth being incorrectly recorded) or because the components change (such as family name changing due to adoption or marriage, or where alias names are used). In addition, records with missing data (such as unknown date of birth) cannot be used with this method.

In contrast to the simple deterministic method, the multi-step key-based linkage method aims to minimise both false negatives (missed links) and false positives (incorrect links) by using a series of keys that vary in distinctiveness. This method reduces the possibility that records belonging to different people are incorrectly recorded as belonging to the same person. At the same time, it increases the possibility that records belonging to the same person will be identified, even where components such as last name have changed. This method can be used where values are missing (such as unknown dates of birth), and where available, it can also use alternative information such as alias names. This method has been used in a number of AIHW linkage projects, including the earlier project linking child protection notifications with youth justice supervision data (AIHW 2012a) and is the method that will be used to link child protection and youth justice supervision data.

### **Box 2.1: Examples of linkage results using simple deterministic linkage**

#### **True positive (correct link – the records are correctly identified as belonging to the same person)**

Emily Brown appears in two data collections. In both data collections, her date of birth is correctly recorded as 2 February 1997. A linkage key consisting of the 2nd, 3rd and 5th letters of family name, the 2nd and 3rd letter of the given name and the date of birth will result in her records being correctly identified as belonging to the same person.

Data set 1: RONMI02021997

Data set 2: RONMI02021997

Result: link

#### **True negative (no link – the records are correctly identified as belonging to different people)**

Emily Brown appears in one data collection with a date of birth of 2 February 1997. A different person who shares the same date of birth, Emily Smith, appears in another data collection. A linkage key consisting of the 2nd and 3rd letter of the given name, the 2nd, 3rd and 5th letters of family name and the date of birth will result in their records being correctly identified as belonging to different people.

Data set 1: RONMI02021997

Data set 2: MIHMI02021997

Result: no link

#### **False negative (missed link – the records are incorrectly identified as belonging to different people)**

Emily Brown appears in two data collections. In one collection, her date of birth is correctly recorded as 2 February 1997. In the other collection, her date of birth has been incorrectly recorded as 20 February 1997. A linkage key consisting of the 2nd and 3rd letter of the given name, the 2nd, 3rd and 5th letters of family name and the date of birth will result in their records being incorrectly identified as belonging to different people.

Data set 1: RONMI02021997

Data set 2: RONMI20021997

Result: no link

#### **False positive (incorrect link – the records are incorrectly identified as belonging to the same person)**

Emily Brown appears in one data collection with a date of birth of 2 February 1997. A different person who shares the same date of birth, Amity Aronnes appears in another data collection. A linkage key consisting of the 2nd and 3rd letter of the given name, the 2nd, 3rd and 5th letters of family name and the date of birth will result in their records being incorrectly identified as belonging to the same person.

Data set 1: RONMI02021997

Data set 2: RONMI20021997

Result: link

## 2.2 Multi-step linkage process

The first step of the linkage process involves generating the ordered list of linkage keys. Once an ordered list of linkage keys has been determined, these keys are used to link the data collection using the multi-step key-based linkage method. Following this, the data sets are created for analysis and the quality of the linkage is assessed. This process will be followed on an annual basis to create the linked data set. These steps, which were first outlined in AIHW 2012b, are detailed below.

### Generating the ordered list of linkage keys

The multi-step key-based linkage method involves linking the data collections one key at a time, starting with the most precise key, which is the key most likely to result only in true positives (correct links) and no false positives (incorrect links). That is, it is likely to return links that are almost certainly correct. However, the cost to this precision is that a number of links will be missed. The next most precise key is then used, and so on, until the keys no longer add any value to the linkage because they are estimated to return too many incorrect links.

Generating the ordered list of linkage keys involves:

1. identifying possible linkage keys
2. measuring the utility of each key
3. ordering the keys
4. determining the point at which keys do not add value to the linkage.

#### Step 1: Identifying possible linkage keys

The variables that are shared by the two collections determine the possible linkage keys, which are formed by systematically including and excluding the various key components in all possible permutations. The child protection and youth justice supervision data collections currently share 7 variables that can be used to derive 13 linkage key components, which results in 1,280 possible keys (see Table 2.1). Over time, additional variables could be added to the collections to improve the distinctiveness of the linkage keys.

**Table 2.1: Linkage key components**

Component	Symbol	Details	Data collection	
			JJ NMDS	CP NMDS
Letters of family name	F3		Client file (letters of name)	Client file (letters of name)
Letters of given name	G2		Client file (letters of name)	Client file (letters of name)
Day of birth	D2		Client file (date of birth)	Client file (date of birth)
Month of birth	M2		Client file (date of birth)	Client file (date of birth)
Year of birth	Y4		Client file (date of birth)	Client file (date of birth)
Sex	S1		Client file	Client file
Indigenous status	I1	Indigenous, non-Indigenous or unknown	Client file	Client file
Suburb/town/locality name	TS		Order and detention files	NIS <sup>(a)</sup> and LA <sup>(b)</sup> files
Postcode	P4	Digits 1–4	Order and detention files	NIS <sup>(a)</sup> and LA <sup>(b)</sup> files
Postcode (alternative 1)	P3	Digits 1–3	Order and detention files	NIS <sup>(a)</sup> and LA <sup>(b)</sup> files
Postcode (alternative 2)	P2	Digits 1–2	Order and detention files	NIS <sup>(a)</sup> and LA <sup>(b)</sup> files
Postcode (alternative 3)	P1	Digit 1	Order and detention files	NIS <sup>(a)</sup> and LA <sup>(b)</sup> files

(a) Notification, investigations and substantiations.

(b) Living arrangements.

*Note:* Suburb/town/locality name and postcode are for the child's or young person's usual place of residence at the time of the event.

Each child or young person within the state and territory child protection and youth justice supervision data collections may have multiple values for the components relating to their usual place of residence, because this is recorded as at the start of the event. It is also possible to have multiple values for the person-level components (letters of family and given names, date of birth, sex and Indigenous status), due to the method of data supply for the collections. Data for both collections are supplied annually, and each supply can have only one set of values for the person-level components. However, where the person appears in multiple data supplies over a number of years, these values can be updated. Where there are multiple values for a component, these values contribute equally to the linkage (that is, no weighting is applied).

For example, a young person has 2 events in the 2012–13 youth justice supervision data for Tasmania (Table 2.2). Their usual place of residence was recorded as 'Wynyard' for their first event and as 'Smithton' for their second event (with corresponding changes to the postcode). In the 2013–14 youth justice supervision data for Tasmania, the letters of their family name have been updated from RON to MIH. These records are known to belong to the same person within the Tasmanian youth justice supervision data, because the person identifier is the same across all 3 records, even though the values for some of the linkage key components have changed.

**Table 2.2: Example of multiple sets of key component values for a single person identifier**

Collection year	Person identifier	F3	G2	D2	M2	Y4	S1	I1	TS	P4	P3	P2	P1
2012–13	6456890	RON	MI	02	02	1997	2	1	WYNYARD	7325	732	73	7
2012–13	6456890	RON	MI	02	02	1997	2	1	SMITHTON	7330	733	73	7
2013–14	6456890	MIH	MI	02	02	1997	2	1	SMITHTON	7330	733	73	7

**Step 2: Measuring the utility of each linkage key**

Once the possible linkage keys have been identified, the next step is to measure the utility of each key using the available child protection and youth justice supervision data. There are three measures of utility:

1. **The relative distinctiveness of each possible linkage key.** A key that mostly contains components with values that are shared by many in the study population (for example, sex) will not be useful in distinguishing between individuals. In contrast, a key that contains components with many values that are widely distributed among the study population (for example, letters of given name) will be more useful.
2. **The estimated average number of people for each possible linkage key.** A linkage key that results in fewer distinct people per key value will be more precise than one that results in more distinct people per key value.
3. **The estimated maximum number of people for each possible linkage key.** A linkage key that results in a lower maximum number of people per key value will be more precise than one that results in a higher maximum number.

The first measure, the relative distinctiveness of each possible linkage key, is obtained by first extracting all the distinct sets of values for each linkage key component (see Table 2.1) for each person identifier (as noted previously, these person identifiers are specific to the state or territory and the data collection). As discussed above, each person identifier may have multiple unique sets of values. Then, for each key component, a midpoint is obtained by calculating the smallest number of categories that account for at least 50% of the population. Only keys that were expected to have at least 10% as many different values in the population as a linkage key comprising letters of family name; letters of given name; day, month and year of birth; and sex were retained for further assessment of utility.

The second and third measures of utility are estimated by testing the possible linkage keys using data from the collections to be linked. Because the child protection and youth justice supervision data collections contain state-specific person identifiers (that is, within a state or territory in each of the data collections, it is possible to know which records belong to the same person), the average and maximum number of people for each possible linkage key can be estimated from the number of person identifiers attached to each key combination (see calculations for an example data set in Table 2.3).

**Table 2.3: Example of a data set for calculating the average and maximum number of person identifiers**

Person identifier	Values for key components					Combinations for selected linkage keys			
	F3	G2	D2	M2	Y4	F3Y4	F3G2Y4	G2D2M2Y4	F3G2D2M2Y4
6456890	MIH	MI	02	02	1997	MIH1997	MIHMI1997	MI02021997	MIHMI02021997
6866990	MIH	MI	10	09	1997	MIH1997	MIHMI1997	MI10091997	MIHMI10091997
6055443	DIL	MI	10	09	1997	DIL1997	DILMI1997	MI10091997	DILMI10091997
6155443	RON	LE	02	06	1999	RON1999	RONLE1999	LE02061999	RONLE02061999
6665491	MIH	AB	26	04	1997	MIH1997	MIHAB1997	AB26041997	MIHAB26041997
6889225	MIH	DE	18	08	1997	MIH1997	MIHDE1997	DE18081997	MIHDE18081997
<b>Total number of distinct combinations</b>						<b>3</b>	<b>5</b>	<b>5</b>	<b>6</b>
<b>Average number of person identifiers per combination</b>						<b>2</b>	<b>1.2</b>	<b>1.2</b>	<b>1</b>
<b>Maximum number of person identifiers per key</b>						<b>4</b>	<b>2</b>	<b>2</b>	<b>1</b>

### Step 3: Ordering the linkage keys

The three measures of the utility of each key are then used to generate an order for the linkage keys:

- The keys are first ordered by the average number of person identifiers per key value (lowest to highest)
- If this measure is the same for multiple keys, the maximum number of person identifiers per key value (lowest to highest) is then used to order the key
- If the relative precision of multiple keys can still not be distinguished, then the relative distinctiveness is used (highest to lowest).

Using the linkage process described below, the linkage key order is tested within each state and territory, because it is possible to know which records belong to the same person within a state or territory in each of the data collections.

Test data sets are created by extracting all records for the state or territory within the data collection. In contrast to the data used to estimate the average and maximum number of people for each possible linkage key, which only contained one record per person identifier, these test data sets will contain multiple records for each person identifier where they have multiple values for a key component. These records are then duplicated to increase the number of possible links, and the duplicated records are randomly separated into 3 data sets (to approximate the actual linkage process). The 3 data sets are linked using the process described below ('Linking the data'). In the final linked data set, each link is considered correct if the person identifiers (which are specific to each state and territory within a data collection) match. This process is repeated for each state and territory in each of the data collections.

### Step 4: Determining the point at which linkage keys do not add any value to the linkage

For each linkage key, the harmonic mean is generated using the cumulative percentage of correct links (number of correct links divided by number of links) and the cumulative percentage of possible links (number of correct links divided by the maximum number of possible links). This test is replicated for each state and territory in each data collection, and

the harmonic means are graphed to identify the point at which additional keys do not increase the value of the harmonic mean. In practice, this value is generally consistent across states and territories.

### Linking the data

A series of steps are used to link the data one key at a time (Figure 2.1).

1. The individual state and territory data sets (a maximum of 16 data sets, 8 per collection, referred to as Data set 1, Data set 2... in Figure 2.1) are joined at the same time using the most precise key as determined by the testing described above.
2. Records with conflicts are identified and one record from each set of conflicts is randomly selected and inserted in the linked data set. A conflict occurs either because:
  - a particular combination relates to multiple person identifiers from the same state or territory within a collection (for example, 2 person identifiers in Source data set 2 share the same combination for a particular key, and this combination links these person identifiers with 1 person identifier in Source data set 1)
  - a person identifier in one data set has multiple combinations for a particular key and these combinations link to two or more different person identifiers in another data set (for example, 1 person identifier in Source data set 1 has 3 different combinations for that particular linkage key because they have 3 different days of birth recorded).
3. All data sets are then linked using Key 2, with updates made to previous links to allow for linkage information to accumulate, and new links are added to the linked data set. This process continues for the remaining keys.
4. The final linked data set is created by combining the records linked using the ordered list of keys with the records that remain unlinked (see example in Table 2.4). This final linked data set contains a project-specific identifier, which is used to count distinct people in the linked data collection.

**Table 2.4: Example of a final linked data set**

Project identifier	Source data set 1	Source data set 2	Source data set 3
1	123456	245996	3899447
2	133998	2665875	—
3	1889955	—	355778
4	—	—	3359221
etc			

Note: '—' indicates that record does not contain a person identifier from the source data set.

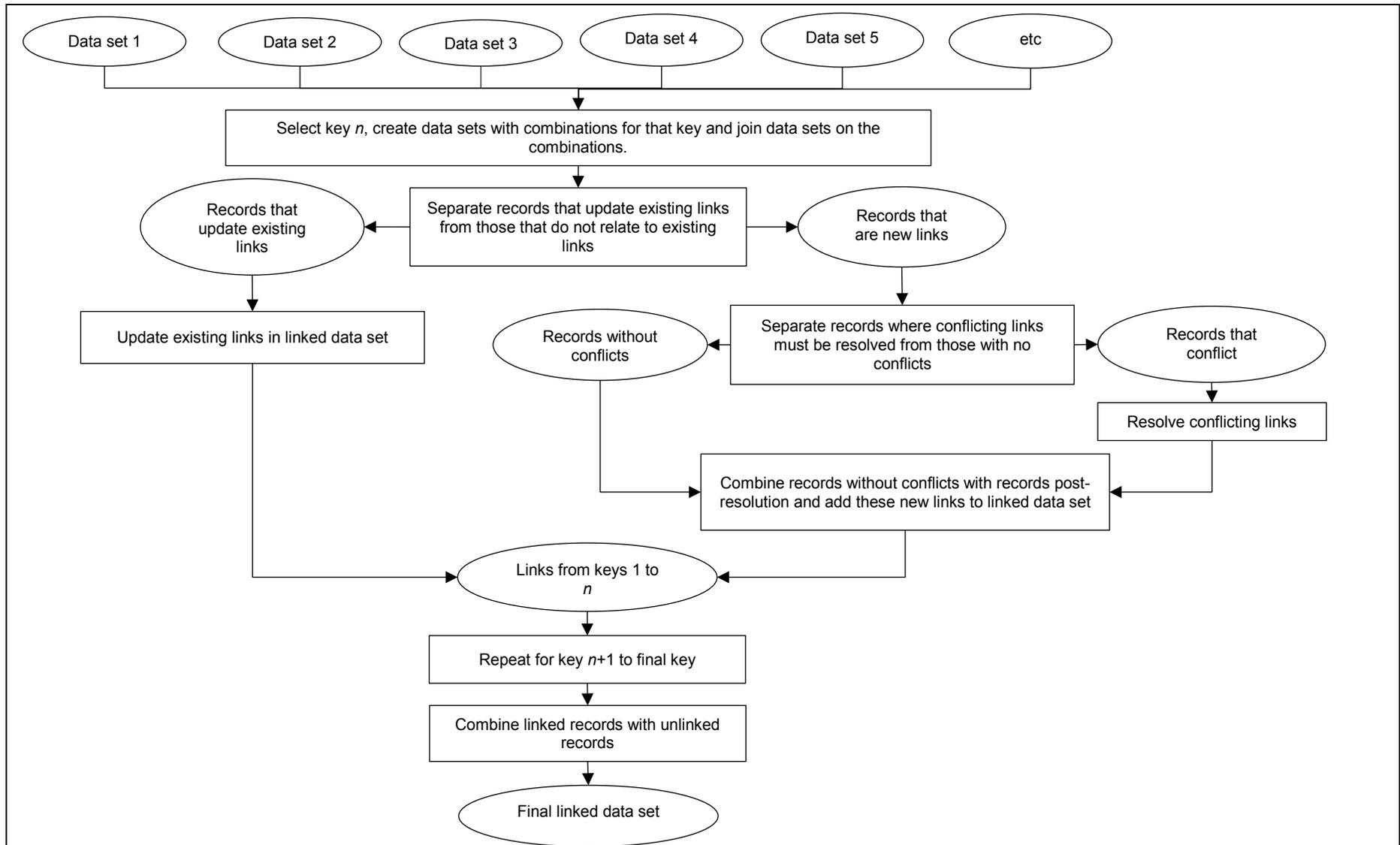


Figure 2.1: Linkage process

## **Creating data sets for analysis**

The final linked data set contains a concordance between a project-specific person identifier derived specifically for this project – and not derivable from the state-specific person identifiers – and the related state-specific person identifiers (with a maximum of 1 per state or territory per data collection). Using this concordance, relevant information about child protection services and youth justice supervision from the source data collections can be brought together for a range of analyses. To protect privacy, the only person identifier retained on an analysis data set is the project-specific person identifier.

In situations where the linkage has resulted in a person having multiple sets of demographic information (such as where date of birth has been recorded inconsistently across collections), the most common set is retained (if several sets of information are equally common, one set is randomly selected). Some of the information on child protection services and youth justice supervision may conflict, such as where a person appears to have been detained in multiple states at the same time. These inconsistencies can be used to measure of the quality of the linkage.

### **3 Exploring relationships between child protection and youth justice supervision**

Children and young people who are involved with both child protection and youth justice supervision have specific and complex needs, and are at risk of a range of health and welfare issues, both as children and young people and later as adults. Understanding the characteristics of these children and young people and their pathways through these sectors can assist support staff, case workers and policy makers to achieve optimal outcomes for children and young people, and their families.

The relationships between child protection and youth justice supervision can be explored in two ways. The first involves analysing the characteristics and patterns of events for children and young people who are involved in both child protection and youth justice supervision in the same period and comparing it with the characteristics of those who are only involved in one sector. This requires data from the 2 collections for the same period for young people aged 10–17, because children aged under 10 are not deemed to have criminal responsibility, while people aged 18 and over are processed by the adult criminal justice system. The second involves analysing the pathways of children and young people through the child protection and youth justice supervision systems.

While analysing the full pathway from birth until the age of 18 for a given cohort of children requires at least 18 years of child protection data, and so is not currently feasible, with current data it is possible to explore the sequence of events within a particular timeframe, such the sequence of events immediately prior to being placed on remand or supervised bail.

#### **3.1 Young people in both sectors in the same financial year**

Typically, children and young people who are involved with both child protection and youth justice supervision, and their families, often access a range of services from a number of different agencies and departments. Some move between child protection and youth justice supervision in the same period or have an ongoing involvement with both departments over a period of time. For example, some young people are serving sentences supervised by a youth justice department while they are also the subject of an investigation of suspected abuse or neglect conducted by a child protection department. Others may be on child protection orders when they commit offences and are placed under youth justice supervision.

Having a greater understanding of the characteristics of the children and young people who are involved in both systems within a year, and the types of child protection services and youth justice supervision they receive, would assist support staff and case workers in different agencies to work together to produce the best outcomes for everyone involved (for example, by using early intervention to divert the child from criminal activity). It would also allow policy makers to improve practices and strengthen the responses provided to young people and their families at critical periods of transition, such as on leaving care. A comprehensive approach to the needs of young people who are leaving out-of-home care

and who are also involved in the youth justice system would reduce the risk of the young person later becoming involved in the adult justice system.

Analyses could include:

- comparing the number, rate and demographic characteristics of young people involved with both child protection and youth justice supervision in the same financial year with those who were involved with only one of these sectors
- exploring the relationships between different types of involvement (for example, whether children in out-of-home care are more likely to be under youth justice supervision than children with a substantiated investigation, but not in out-of-home care).

### **3.3 Pathways through child protection and youth justice**

While a number of young people will be involved in both the child protection system and youth justice supervision in the same year, for some, a number of years will elapse between the abuse or neglect and entering the youth justice system.

Learning about the pathways through these sectors will assist policy makers in identifying both the characteristics of children and young people who are more likely to be involved in child protection and youth justice supervision, and the possible points of intervention. This information can then be used to provide more effective support and to produce better outcomes for these children and young people.

Once a sufficient number of years of data have been linked, analyses of the pathways that children and young people take through child protection (from birth) and youth justice supervision (from the age of 10) could include exploring the common pathways and identifying whether pathways differ by demographic characteristics, or if particular events (such as type of abuse or neglect and age at first notification) are correlated with later youth justice supervision.

In the short-term, these pathways could also be explored using restricted age groups, such as those aged 15–17.

In addition, more focused analyses on the relationships between particular events would also provide valuable information on potential points of intervention. These could include:

- exploring the relationships between stability of out-of-home care placement and involvement in youth justice supervision
- exploring the relationships between particular child protection and youth justice supervision events, such as being in out-of-home care or on a care and protection order and being placed in unsentenced detention.

### **3.4 Data availability and reporting**

Both child protection services and youth justice supervision are only relevant to specific age groups. For child protection, the age range is from birth to 17 years, while youth justice supervision applies only to those aged 10 to 17. Consequently, the years for which data are available determine which analyses are possible. Each year of data contains a cross section of people of different ages, and as data accumulate over time and the same children and young

people appear in data sets for a number of years, it becomes possible to do longitudinal analyses.

Youth justice supervision data are currently available from 2000–01 to 2013–14 and each year of data contains a cross section of young people aged 10–17 (Table 3.1). This collection currently has longitudinal data covering all possible youth justice supervision from the age of 10–17 for 7 consecutive birth cohorts (see shaded cells in Table 3.1). That is, all possible youth justice supervision data for young people born in 1990–91 will be included in the data for 2000–01 to 2007–08; all possible youth justice supervision data for young people born in 1991–92 will be included in the data for 2001–02 to 2008–09; and so on.

Child protection data are currently available from 2011–12 to 2013–14 and each year of data contains a cross section of young people aged 0–17 (Table 3.2). Longitudinal data covering all possible child protection services for those aged 0–17 in 2011–12 will only be available from 2029–30 (although longitudinal data could be available before then if data from before 2011–12 were provided). Currently, an age-based subset of each of the 3 years of available child protection data overlaps with the youth justice supervision data (see shaded cells in Table 3.2).

Annual reports on the linked data collection would be able to explore the number and characteristics of young people both receiving child protection services and under youth justice supervision in the same year. It would also be possible for these reports to contain longitudinal analyses that report on a subset of possible pathways. Using currently available data, it is possible to explore the 3-year pathways for children from the ages of 8 to 10 through to young people from the ages of 15 to 17. Future reports would be able to incorporate longer pathways; for example, a report using data for 2016–17 and earlier would be able to explore 6-year pathways, from children aged 5–10 through to young people aged 12–17.

**Table 3.1: Availability of youth justice supervision data: age of young person in a collection year by year of birth**

Year of birth (1 July to 30 June)	Collection year (1 July to 30 June)													
	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14
	Age of young person in the financial year													
1990-91	10	11	12	13	14	15	16	17						
1991-92		10	11	12	13	14	15	16	17					
1992-93			10	11	12	13	14	15	16	17				
1993-94				10	11	12	13	14	15	16	17			
1994-95					10	11	12	13	14	15	16	17		
1995-96						10	11	12	13	14	15	16	17	
1996-97							10	11	12	13	14	15	16	17
1997-98								10	11	12	13	14	15	16
1998-99									10	11	12	13	14	15
1999-00										10	11	12	13	14
2000-01											10	11	12	13
2001-02												10	11	12
2002-03													10	11
2003-04														10

Note: Shading indicates birth cohorts for which complete longitudinal data are available.

**Table 3.2: Availability of child protection data: age of young person in a collection year by year of birth**

Year of birth (1 July to 30 June)	Collection year (1 July to 30 June)		
	2011–12	2012–13	2013–14
	<b>Age of young person in the collection year</b>		
1994–95	<i>17</i>		
1995–96	<i>16</i>	<i>17</i>	
1996–97	<i>15</i>	<i>16</i>	<i>17</i>
1997–98	<i>14</i>	<i>15</i>	<i>16</i>
1998–99	<i>13</i>	<i>14</i>	<i>15</i>
1999–00	<i>12</i>	<i>13</i>	<i>14</i>
2000–01	<i>11</i>	<i>12</i>	<i>13</i>
2001–02	<i>10</i>	<i>11</i>	<i>12</i>
2002–03	<i>9</i>	<i>10</i>	<i>11</i>
2003–04	<i>8</i>	<i>9</i>	<i>10</i>
2004–05	<i>7</i>	<i>8</i>	<i>9</i>
2005–06	<i>6</i>	<i>7</i>	<i>8</i>
2006–07	<i>5</i>	<i>6</i>	<i>7</i>
2007–08	<i>4</i>	<i>5</i>	<i>6</i>
2008–09	<i>3</i>	<i>4</i>	<i>5</i>
2009–10	<i>2</i>	<i>3</i>	<i>4</i>
2010–11	<i>1</i>	<i>2</i>	<i>3</i>
2011–12	<i>0</i>	<i>1</i>	<i>2</i>
2012–13		<i>0</i>	<i>1</i>
2013–14			<i>0</i>

Note: Italics indicate the current overlap with available youth justice supervision data.

### 3.5 Linkage with other data collections

Children and young people who both receive child protection services and are supervised by youth justice departments are at risk of a range of health and welfare issues, both as children and young people and as adults. These include mental health conditions, alcohol and other drug problems and homelessness. They are also more likely to be involved in the criminal justice system as adults than people who do not have a history of child protection or youth justice supervision.

The previous AIHW linkage project linking youth justice supervision data with selected child protection data also included data from an earlier version of the homelessness data collection (the Supported Accommodation Assistance Program data collection). The results showed that both children receiving child protection services (substantiated notifications) and young people under youth justice supervision were more likely to have received homelessness services in both the year before and the year after their most recent substantiated notification or youth justice supervision, compared with the general population.

The AIHW manages a number of collections that have, or will be expanded to have, data suitable for linkage. These data collections contain information on:

- people who have died, including age at death and cause of death
- the health of adults in prisons
- people who receive disability support services, including people with mental health issues and acquired brain injury
- people who receive treatment for alcohol and other drugs
- people who receive homelessness services.

Expanding the child protection and youth justice supervision linked data collection to include other health and welfare data collections would provide valuable information that could be used to improve the lives of disadvantaged children and their families.

# Glossary

**harmonic mean:** a type of mean that is used to find the average of rates.

**key-based linkage:** a method that uses a series of linkage keys that contain sufficient information to link records for statistical purposes.

**linkage keys:** a string of characters and numbers that can be used to link records for statistical purposes.

**person identifier:** a series of alphanumeric characters that is unique to a person within a state or territory in either the child protection or youth justice supervision data collections. A person who received child protection services or youth justice supervision in more than one state or territory, or who received child protection services and youth justice supervision in the same state or territory, will have multiple person identifiers.

**project-specific identifier:** a series of alphanumeric characters that is unique to a person in the linked data set. A person who, following the data linkage, is considered to have received child protection services or youth justice supervision in more than one state or territory, or who is considered to have received child protection services and youth justice supervision in the same state or territory, will have one project-specific identifier.

# References

- AIHW (Australian Institute of Health and Welfare) 2012a. Children and young people at risk of social exclusion: links between homelessness, child protection and juvenile justice. Data linkage series no. 13. Cat. no. CSI 13. Canberra: AIHW.
- AIHW 2012b. Linking SAAP, child protection and juvenile justice data: technical report. Data linkage series no. 14. Cat. no. CSI 14. Canberra: AIHW.
- AIHW 2015a. Child protection Australia 2013–14. Child welfare series no. 61. Cat. no. CWS 52. Canberra: AIHW.
- AIHW 2015b. Youth justice in Australia 2013–14. AIHW bulletin no. 127. Cat. no. AUS 188. Canberra: AIHW.
- Cashmore J 2011. The link between child maltreatment and adolescent offending: systems neglect of adolescents. *Family Matters* 89:31–41.
- Currie J & Tekin E 2006. Does child abuse cause crime? Cambridge, MA: National Bureau of Economic Research.
- Indig D, Vecchiato C, Haysom L, Beilby R, Carter J, Champion U et al. 2011. 2009 NSW Young People in Custody Health Survey: full report. Sydney: Justice Health and Juvenile Justice.
- Johnson G, Natalier K, Mendes P, Liddiard M, Thoresen S, Hollows A et al. 2010. Pathways from out-of-home care. Melbourne: RMIT Research Centre: Australian Housing and Urban Research Institute (AHURI).
- Jonson-Reid M & Barth RP 2000. From maltreatment report to juvenile incarceration: the role of child welfare services. *Child Abuse & Neglect* 24(4):505–20.
- Ryan JP, Herz D, Hernandez PM & Marshall JM 2007. Maltreatment and delinquency: investigating child welfare bias in juvenile justice processing. *Children and Youth Services Review* 29(8):1035–50.
- Stewart A, Dennison S & Hurren EH 2005. Juvenile offending trajectories: pathways from child maltreatment to juvenile offending, and police cautioning in Queensland. Natha, Qld: School of Criminology and Criminal Justice, Griffith University.
- Weatherburn D & Lind B 2006. What mediates the macro-level effects of economic and social stress on crime? *The Australian and New Zealand Journal of Criminology* 39(3):384–97.
- Widom CS 1989. Child abuse, neglect and violent criminal behavior. *Criminology* 27:251–71.

# List of tables

- Table 2.1: Linkage key components .....9
- Table 2.2: Example of multiple sets of key component values for a single person identifier .....10
- Table 2.3: Example of a data set for calculating the average and maximum number of person identifiers .....11
- Table 2.4: Example of a final linked data set .....12
- Table 3.1: Availability of youth justice supervision data: age of young person in a collection year by year of birth .....18
- Table 3.2: Availability of child protection data: age of young person in a collection year by year of birth .....19

# List of figures

Figure 2.1: Linkage process.....13

## Related publications

This report can be downloaded free of charge from <http://www.aihw.gov.au/youth-justice-publications/>.

More information about children and young people in the child protection system in Australia and under youth justice supervision in Australia is available from the AIHW website at <http://www.aihw.gov.au/families-and-children/>.

In addition to the publications listed in the References, the following AIHW publications may also be of interest:

- AIHW 2014. Youth detention population in Australia 2014. Juvenile justice series no. 16. Cat. no. JUV 53. Canberra: AIHW.
- AIHW 2014. Pathways through youth justice supervision. Juvenile justice series no. 15. Cat. no. JUV 40. Canberra: AIHW.

Using available national data to understand the characteristics of children and young people who are both in the child protection system and under youth justice supervision, and their pathways through these systems, would assist support staff, case workers and policy makers to achieve optimal outcomes for children and young people and for their families. This report describes how these data collections can be linked and how the relationships between child protection and youth justice supervision can be explored.