

# Re-contact within the Justice System: Integrating Multiple Data Sources through Record Linkage

Alexander Reicker

Household Survey Methods Division, Statistics Canada,  
R. H. Coats Building, 100 Tunney's Pasture Driveway,  
Ottawa, Ontario, Canada, K1A 0T6

## Abstract

The size and scope of linkage projects are increasing as probabilistic record linkage becomes a standard method to integrate administrative and survey files based on personal identifiers. Thanks to recent improvements in computers and software, Statistics Canada is currently undertaking multiple, large record linkage projects, including the Justice Re-contact Project, which follows pathways through the justice system and measures the time between police contacts.

In this paper, we give an overview of the Re-contact project which involves a series of internal and external record linkages using data from police services, provincial courts, and provincial and federal correctional services. The files are linked probabilistically using the G-Link software package developed at Statistics Canada. Our results highlight the difficulties in integrating many administrative datasets in one common environment due to the variations in the data availability and data quality, as well as the technical challenges of managing a system where the number of records grows rapidly as new datasets are added.

**Key Words:** Record Linkage, Re-contact, Recidivism, Justice, G-Link, Administrative Data

## 1. The Re-contact Project

The Canadian Centre for Justice Statistics (CCJS, a division of Statistics Canada) has undertaken a multi-phase, multi-year project to demonstrate the feasibility of measuring re-contact with the justice system. No unique identifier for victims or accused persons is collected with the administrative data gathered on criminal incidents in Canada. The identifiers collected with courts and corrections data may or may not be unique depending on the jurisdiction. Therefore record linkage using non-unique identifiers, such as name and date of birth are necessary to track people through the justice system.

Understanding the different policy options and how they affect outcomes for individuals is critical to making informed decisions. Determining whether an

individual has previously been in contact with the one or more sectors of the justice system is a key step in this process. There may be differences due to age, sex or other variables (such as type of offence) in the prevalence and frequency of re-contact. Once these factors have been analyzed, the effects of court decisions and sentences, as well as information relating to their time in custody, if applicable, can be used to enrich the data and allow powerful analysis.

### **1.1 Definition of Re-contact**

The justice system in Canada consists of three sectors: policing, courts and corrections. At Statistics Canada, there are three corresponding administrative data collection programs: the Uniform Crime Reporting Survey (UCR), the Integrated Criminal Courts Survey (ICCS) and the Integrated Correctional Services Survey (ICSS). These three data sources measure official contacts with the justice system: incidents cleared by the police through the laying of charges or otherwise, cases tried in court, and community and custodial sentences overseen by correctional services.

A contact is defined as an official intervention by police, courts or corrections. A re-contact is defined as a subsequent contact signifying a new, official intervention. CCJS determined that measuring re-contact is more appropriate than attempting to measure recidivism given the nature of the available administrative data. Recidivism refers to the subsequent commission of a crime which may not come to the attention of police or for which the person may not be charged.

Re-contact may be measured in any of the sectors, such as time between correctional involvements, but may also be measured between sectors, such as time between the end of the correctional involvement and the next charge by police.

### **1.2 Types of Analysis**

There are four measures of re-contact which are addressed by this initiative:

- Prevalence of re-contact: The proportion of individuals having multiple contacts.
- Frequency of re-contact: The number of subsequent contacts for a given individual.
- Time to re-contact: The time elapsed between one contact and the next.
- Nature of re-contact: Comparison of the characteristics of the offences, e.g. an escalation in the seriousness of the offences being committed.

The analysis is performed by gender, age category and type of offence (violent/non-violent, etc.).

This project also allows CCJS to trace the trajectory of offenders through the justice system. There is currently no identifier in common between the different datasets allowing analysts to follow an offender's path from policing through courts to corrections. Linking police charges to court cases and correctional involvements using probabilistic record linkage allows CCJS to create an identifier which can be used in the analysis of these pathways through the justice system.

### **1.3 Project Phases**

Our justice partners in the federal, provincial and territorial governments, as well as the criminological research community in Canada have expressed interest and support for the analysis of re-contact. The CCJS Re-contact project was started in 2012 and is projected to continue until at least 2018, making it a large and ambitious project. It is being completed in phases, with CCJS having completed the first two phases as of May 2015.

#### *1.3.1 Phase I (2012 – 2014)*

Phase I involved four independent record linkage projects. Each of the datasets was linked to itself to evaluate the quality of the personal identifiers and the keys used by the data providers. The data providers for the four datasets were:

- a. Toronto Police Service
- b. Waterloo Regional Police Service
- c. Nova Scotia Courts
- d. Saskatchewan Provincial and Federal Corrections

All linkages were done with the aim of evaluating the quality of the variables on the files. This includes the keys created by the data providers to identify re-contacts as well as the personal identifiers which are needed to do probabilistic record linkage. After the linkages, the datasets were also analyzed to produce measures of the prevalence, frequency and nature of re-contact.

#### *1.3.2 Phase II (2013 – 2015)*

Phase II involved the first cross-sectoral linkage of justice data using data from the province of Saskatchewan (approximately 1.1 million people). Five police data sets covering seven municipalities and the over 100 Royal Canadian Mounted Police (RCMP) detachments, as well as courts data and adult and youth corrections datasets were linked together. The police datasets covered more than 95% of the population; the courts dataset covered all cases except those tried in superior court; and corrections datasets excluded offenders serving sentences in a federal institution. This allowed the analysis of both the pathways through the justice system and measures of re-contact for any sector of the justice system. All

datasets covered the years 2009 – 2012 which allowed us to follow a cohort from 2009/2010 for two years.

### *1.3.3 Phase II Extension (2015 – 2016)*

Similar to Phase II, this phase involves the cross-sectoral linkage of data from Ontario, the largest province in Canada with more than 13 million people.

### *1.3.4 Phase III (2016 – 2017)*

Phase III concentrates on the analysis of the linked data by focusing on sector specific indicators of re-contact.

### *1.3.5 Saskatchewan Profiles Project (2016 – 2018)*

Crime in Canada is often analyzed and described in isolation of other pertinent social data. The Saskatchewan Profiles Projects builds on the linkages completed in the initial phases through the integration of other social domain data, such as data from health, education, social services, labour and income statistics. This part of the re-contact project will provide a more complete understanding of the factors associated with repeat contact with the justice system.

## **2. Probabilistic Record Linkage Using G-Link**

Record linkage is the process in which records or units from different sources are combined into a single file. There are two possible methods: deterministic and probabilistic record linkage. Deterministic methods rely on exact agreement between one or more fields while probabilistic methods estimate the likelihood that two records correspond to the same unit. As well, linkages can be internal (one-file) or external (two-file). The purpose of internal record linkages is to identify duplicates on the file, such as finding multiple police incidents involving the same accused person. External record linkages create matched pairs of records between two files, such as the courts and corrections files, linking the records to create a new record with information from both sources on it.

The output of a record linkage project is a partition of all the records into groups. Every record in a group belongs to the same unit in the population. In a two-file linkage, every group contains at least one record from both files. These linkages may also impose additional constraints, for instance that each group be 1-to-1 (contains records from one and only one unit from each file).

The units involved in the linkages done for the re-contact project vary. For the internal record linkages, the units are people. That is, records, be they police incidents, court cases or changes in legal hold status, for the same person, are linked together. For the external record linkages, the units are person-incidents.

Court cases and changes in legal hold status are considered as being derived from a specific incident and the records corresponding to the particular person-incident from the three different data sources are grouped together.

## **2.1 Deterministic Record Linkage**

Deterministic record linkage can be direct or hierarchical. Direct deterministic record linkage performs best when a unique, universal key, such as the Social Insurance Number (SIN) for people or Business Number (BN) for businesses, is available to identify matching units. The quality and accuracy of the linkage depends on the quality and accuracy of the data, since any error in the key will result in a missed or incorrect link. In addition, the key must be available for all units in the population and fixed over time. Direct deterministic linkage is used when the police service, or provincial court or corrections agency has already identified a re-contact and assigned such a key to an accused person.

When no such key is available, a key can be created by concatenating variables, such as name, date of birth and sex. This new key may not, however, be unique. Again, this type of linkage depends on the accuracy of the data. By requiring all values to agree exactly, this type of linkage allows relatively few false positive matches at the expense of missing many good links due to differences in variables such as name which are subject to typographical errors and change over time.

Hierarchical deterministic linkage involves performing multiple passes while varying the key. The first pass will use all variables, while each successive pass will relax the key, for instance requiring agreement on the middle initial instead of the full middle name. Each pass corresponds to accepting a particular agreement profile.

## **2.2 Probabilistic Record Linkage**

Probabilistic record linkage involves estimating the likelihood that two records correspond to the same unit. Instead of determining the agreement profiles which will be accepted prior to the linkage, record pairs are compared and their agreement profiles are determined. Each profile is assigned a weight based on the probability of obtaining that profile given that the pair is a true match and the probability of obtaining that profile given that the pair is not a true match. In the Fellegi-Sunter approach [3], this weight is compared to a pair of thresholds to determine whether the pair is accepted, rejected or must be sent for clerical review. The calculation of the weight and the thresholds are therefore the critical steps when doing probabilistic record linkage.

## **2.3 G-Link**

G-Link is software for probabilistic linkage developed and supported by Statistics Canada for use in linkage projects throughout the agency. Statistics Canada has a

long history of research and development in probabilistic record linkage, including the development of GRLS (Generalized Record Linkage System, no longer supported) and G-Link. The latest version, G-Link 3.1, was released in 2014 and is a SAS-based system with a Graphical User Interface (GUI) [9].

### *2.3.1 Blocking Criteria*

The total number of pairs when linking two files is the product of the file sizes, which is extremely large even for relatively small files. Since most of the pairs created are not true matches, the first step of the record linkage is to eliminate obvious non-matches. That is, an efficient deterministic pass is completed to determine which pairs should be further evaluated using the rules and weights. This is known as the blocking step. In G-Link, the blocking criteria are determined by the user using a Proc SQL query in SAS.

### *2.3.2 Rule Application*

The rules are used to compare the fields on the different files. These rules generate an outcome, such as complete agreement, partial agreement, disagreement or missing, for each set of fields being compared. Built-in partial agreement outcomes include typos (for characters, numbers and dates) as well as nicknames and the Jaro-Winkler distance [10] between the two strings. In addition to the built-in rules, users can create their own customized rules using SAS code. This allows them the flexibility to compare complex data structures and combinations of fields.

Comparisons of equivalent fields, such as multiple given names, are implemented as matrix rules. These account for all possibilities of re-ordering the names while maintaining the ability to distinguish direct agreement (first given name agrees with first given name) from cross agreement (first given name agrees with second or third given name). In G-Link, rules can also be executed conditionally. For instance, checking the street address only if there was an agreement on the city name.

The rule application step is generally the time limiting step in G-Link projects and is multi-threaded to make maximum use of the available processing power. However, it involves writing out an outcome for each rule for each potential pair to disk and is limited by I/O constraints.

### *2.3.3 Weight Calculation*

G-Link uses an iterative process to determine the optimal linkage weights. At the start of each iteration, there is a set of links which are currently accepted as true matches (called the linked set,  $L$ ) and a set of unmatched links which are randomly generated (non-linked set,  $N$ ). The probability ratio of each rule

outcome profile  $r_i$  is compared in the two sets to define the pair weight:

$$10 \cdot \log_2 \left( \frac{P(r_i | L)}{P(r_i | N)} \right)$$

If this weight is large, then that profile is more likely to be found with true matches, if it is small, then it is less likely. These pairs with weights above the upper threshold define the linked set  $L$  for the next iteration. G-Link assumes that the rules are independent, since in this case the probability of the profile is the product of the probabilities of the different rule outcomes and the total weight is the sum of the weights for the individual rules. If the rules are not independent, then they should be re-written using conditional processing or by grouping fields together and using matrix or user rules.

As an alternative to calculating the weights, they can be entered manually. This gives consistent weights across different projects and also allows the user to correct for data quality errors and missing data. Missing data decreases the frequency of exact matches in the linked set which reduces the weight, even though exact agreement on a high quality, unique identifier should correspond with a very high weight.

G-Link also incorporates the concept of frequency weights, which refine the outcome weights to account for the agreement value, not just the type of agreement. That is, agreement on a common name such as SMITH should receive less weight than agreement on a rare name such as REICKER. The frequency weights can be imported from another project or generated using the data particular to the population being linked.

### 2.3.4 Grouping

Whether a project is a one-file or a two-file linkage, the output is a group bringing together all of the records which belong to the same unit. Links are transitive in the sense that if A is linked to B, and B is linked to C, then A and C belong to the same group even if they are not directly linked together. G-Link handles grouping for one-file and two-file linkages and groups can be reviewed and edited by the user.

### 2.2.5 Mapping

In a two-file linkage, if the input files have already been unduplicated, then each group should contain records belonging to exactly one unit on each file, known as a one-to-one linkage. Similarly, if one file has been unduplicated the linkage is one-to-many or many-to-one. These conditions are not always satisfied, but application of the mapping algorithm ensures that this is the case for the final output. G-Link selects the link with the highest weight and breaks any links in conflict with that one. This process is repeated until the resulting groups satisfy

the required mapping condition. In the case where two or more links have exactly the same weight, one is chosen randomly.

### 3. Linkage Projects

#### 3.1 Data Sources

##### 3.1.1 Police

The Uniform Crime Reporting Survey (UCR) measures police reported incidents of crime in Canada and is used to calculate the official crime rate in Canada [8]. A criminal incident may involve several victims, accused persons and violations of the law. Two or more violations of the law (and their related victims and accused persons) are grouped into the same incident if and only if they are committed by the same person or group of persons and if they are either:

- i) part of simultaneous or sequential actions that occur at the same place; or
- ii) part of interrelated actions over a short period of time; or
- iii) a violent action that is repeated over a period of time and that comes to the attention of the police at one time; or
- iv) a series of similar crimes, committed at the same location by the same individual(s), that comes to the attention of the police at one time.

There are several exceptions; in particular, traffic and non-traffic violations must be reported in separate incidents. Note also that an incident may occur over a long period of time. While each incident may involve any number of violations, only the four most serious violations are collected by Statistics Canada.

There are more than 1,200 separate police detachments responding to the survey, comprising 204 different police forces. The UCR collects incident-level microdata from police services covering more than 99% of the population of Canada. Data for the Incident-based Survey are collected directly from police records management systems. There are two widely used records management systems, Versaterm and Niche, which simplifies the standardization of the data prior to linkage. Data is collected in three files, the incident file, the accused file and the victim file, with each file containing a key (the incident file number) which is used to link the files for analysis. The unit of analysis for re-contact is either the person or the person-incident and incidents that involve multiple accused persons are split up.

Names are not collected by the UCR. Instead, the last names and given names are concatenated and the Soundex algorithm is applied; the result is submitted to Statistics Canada via the UCR. Special extracts were obtained from the police services to fill in this data gap. These special extracts included many variables, including full name, as well as known aliases and other names (e.g. maiden



names) and an identifier for each accused person to link them between incidents. The record linkage projects search for cases where a single person was assigned multiple identifiers; that is, their different incidents had not been linked together by the police service. Because the personal identifiers are used as primary keys in the police data bases, it was impossible for CCJS to evaluate when the accused persons involved in two incidents may have been linked together erroneously.

### *3.1.2 Courts*

The Integrated Criminal Court Survey (ICCS) maintains a national database of statistical information on appearances, charges, and cases for youths and adults appearing in criminal court in Canada [7]. The survey is intended to be a census of pending and completed federal statute charges heard in provincial-territorial and superior courts in Canada; however some superior courts are not covered. Appeal courts, federal courts (e.g., Tax Court of Canada) and the Supreme Court of Canada are not covered by the survey. Data is collected at the appearance level and then aggregated to the case level as part of the standard processing of the data at CCJS. A case is defined as one or more charges that were processed at the same time and can involve only one accused person.

Names are already collected with the ICCS data. Some, but not all, of the provincial-territorial courts attempt to identify accused persons who have been involved in multiple cases. However, even those jurisdictions that do track accused persons miss a significant number of links. For this project, an auxiliary file containing the warrant execution date was used in addition to the ICCS data. This date was used to link the court case to a change in legal hold status of the correctional involvement allowing us to avoid false positive matches involving the same offender.

### *3.1.3 Corrections*

The Integrated Correctional Services Survey (ICSS) collects microdata on adults and youth under the responsibility of the federal and provincial/territorial correctional systems. The collected data include admissions and releases by legal hold status (e.g. remand, sentenced, probation). Legal hold statuses are aggregated to derive a correctional involvement which is an uninterrupted period of time when the offender is in sentenced custody, under probation, or with another community-based program.

The federal Correctional Service of Canada (CSC) is responsible for offenders serving sentences of two years or more while the provincial correctional services oversee offenders serving shorter sentences. However, all offenders are overseen by the provincial correctional systems while they are on remand and prior to being transferred to the CSC.

Names are collected on the ICSS. In addition, correctional services do an excellent job of identifying people in their datasets, and very few people are issued multiple identifiers within either the adult or the youth subsets of the data. However, the identifier does not always carry over between the youth and adult datasets, so if someone is sentenced as a youth and an adult, these must be linked together using name, date of birth, and other non-unique identifiers.

### 3.2 Linkage Projects

Phase I involved four independent internal (one-file) record linkages, two from police (Toronto Police Service, Waterloo Regional Police Service), one from courts (Nova Scotia) and one from corrections (Saskatchewan). These four datasets were linked and analyzed separately.

Phase II involved a series of 13 related internal and external record linkages using data from Saskatchewan, including 5 police datasets (Regina, Saskatoon, Prince Albert, Moose Jaw and the RCMP), the courts dataset and adult and youth corrections datasets. The linkages were performed as follows:

- i. Each of the five police datasets was linked back to the UCR data to ensure that only incidents which had been counted in the official statistics were included. This linkage was done deterministically using the incident file number.
- ii. The police datasets were independently linked to the courts dataset using G-Link, creating five output separate output datasets. One-to-one mapping was applied so that the final output links one police incident with one court case.
- iii. These five datasets were appended and unduplicated in a single internal record linkage using G-Link, resulting in a single Police-Courts dataset.
- iv. The Adult and Youth Corrections datasets were linked together using probabilistic linkage. No mapping step was done which allowed us to identify the (very few) duplicates
- v. Police-Courts dataset was linked to the Corrections dataset.

Since both inputs to the final file had been unduplicated in earlier steps, no internal record linkage was necessary.

The linkage units in Phase I were people, meaning that we were bringing together all of the records (police incidents, court cases, correctional involvements) belonging to the same person. In Phase II, the units were the more detailed person-incidents. That is, we distinguished between different incidents, court cases and correctional involvements involving the same offender. In order to correctly measure re-contact, we have to distinguish between the multiple contacts one person may have with the justice system. It is critical to associate each police incident with the correct court case and correctional involvement.

### 3.3 Variables

Different linking variables were used for the different linkages. Table 1 shows the availability of the different variables used in Phase II.

**Table 1: Availability of Linkage Variables**

<b>Variable</b>	<b>Policing</b>	<b>Courts</b>	<b>Corrections</b>
<b>Full Name</b>	96%	100%	100%
<b>Date of Birth</b>	85%	100%	100%
<b>Address</b>	2%	60%	61%
<b>Fingerprint ID</b>	10%	0%	39%
<b>Driver's Licence</b>	28%	36%	0%
<b>Social Insurance Number</b>	9%	0%	33%
<b>Incident Date</b>	100%	100%	N/A
<b>Incident Violation Code</b>	100%	98%	N/A
<b>Warrant Execution Date</b>	N/A	100%	100%

Most of the available variables were person-level variables, such as names and dates of birth, which were sufficient for the internal record linkages where the units were people. The most discriminating variables, such as fingerprint ID (FPS), driver's licence and social insurance number (SIN), were not available for the entire population. Therefore the linkages depended on the quality of the name and date of birth in particular.

The incident date and incident violation code were necessary when linking police incidents to court cases, and the warrant execution date was used to link court cases to correctional involvements. These variables are of high quality as they are critical to the processing of the data by the police, courts and correctional services, however they are not particularly discriminating without additional information. Multiple police incidents and court cases may be related to a single correctional involvement since the offender may commit an offence while still under supervision. Having access to these supplementary date variables was critical to our success in linking across sectors to follow the pathway through the justice system.

## 4. Results

### 4.1 Phase I

The primary goal of Phase I was to determine the quality of the existing identifiers being used by the police, courts and correctional services. As expected, the quality of the identifiers varied greatly across each of the datasets.

The identifiers used by both police services missed a significant number of matches which we were able to complete through probabilistic record linkage. We were unable to evaluate the number false matches due to the way the

databases are structured; however we did not see any evidence to suggest that this was due to a problem with the identifiers.

The quality of the identifiers used by the courts and corrections datasets we examined was very good. However, it is known that this is not the case for all jurisdictions in Canada. Producing national-level re-contact statistics will require evaluating the identifiers for each province and territory separately.

The secondary goal was to measure re-contact at the police service level. This was done and the results were shared with the police services themselves as well as with the wider police community through a presentation made to the Police Information and Statistics Committee (POLIS) [6].

In addition to these goals, Phase I evaluated the use of deterministic linkage using the sex, Soundex and date of birth (which are all regularly collected as part of UCR) instead of the full name to calculate measures of re-contact. The TPS data was used and the results of the deterministic and probabilistic linkages were compared. While good deterministic rules were able to approximate the results of the probabilistic linkage, deterministic matching missed some re-contact events. Importantly, re-contact will be underestimated for chronic offenders.

## **4.2 Phase II**

Phase II evaluated the feasibility of linking multiple datasets from police, courts and corrections in Saskatchewan to create pathways through the justice system. This was successful and allowed CCJS to develop the techniques and methods which will be necessary to extend this work to Ontario.

One issue that arose when linking the police datasets was that false positive matches were introduced when the 5 different datasets were linked independently to the courts datasets. This meant that police incidents from different jurisdictions could be linked to the same court case. While this was a relatively rare occurrence (~50 cases), each one had to be resolved manually. To avoid this situation when linking Ontario, the police files may be appended prior to linkage with the Courts file. This way the one-to-one mapping step in G-Link will automatically link the correct police incidents and court cases.

Another problem that we encountered when working with the cross-sectoral linkages of Saskatchewan data was the files sizes increased rapidly as more datasets were brought together. This was due to the variation within each dataset and between datasets. For instance, the police datasets record many names and aliases of offenders because of typographical errors and because people may be trying to disguise their identity from the police. For the courts datasets, the personal identifiers were recorded for each appearance, and slight variations in spelling caused a new record to be generated. Even substituting the middle initial for a middle name, or leaving it out entirely, caused new records to be created.

We were successful in linking the data from the different sectors together, but these linkages necessarily involve incident, court case and correctional involvement level information. Most of the identifiers on the data files were person level identifiers. If an offender was involved in multiple police contacts and court cases, all of the links between these receive very high weights, especially if there is a high quality identifier (FPS, SIN, DL Licence) on the files. This increases the risk of creating false positive links by matching the police incident to the wrong court case, etc. Matching on dates was critical to our decision rules surrounding these cases. The police-courts linkage used the incident date, which is unique, but the linkage to the corrections file used the warrant execution date which is not unique.

The secondary goal was to measure re-contact in the different sectors. The results of this work were shared with the data providers and with the Liaison Officers Committee (LOC) [2].

## 5. Conclusion

The Re-contact project is an ambitious, multi-phase, multi-year project involving all three sectors of the justice system: policing, courts and corrections. The initial phases have shown that probabilistic record linkage can be used to link files across the sectors to follow an offender's pathway through the justice system and to measure the prevalence and frequency of re-contact with the police, courts or corrections.

This work has also demonstrated that probabilistic record linkage using full names provides a significant improvement over deterministic linkage using the Soundex of the name, sex and date of birth. Accurate re-contact statistics cannot be produced using the data collected by the UCR as missed links will cause an underestimate of the number of contacts for repeat offenders.

The quality of the identifiers also varies across the sectors and across jurisdictions in the case of the provincial and territorial courts. Police services have the lowest quality identifiers making record linkage a necessary step in the measurement of re-contact.

The remaining phases of this project will extend our work on re-contact by linking larger datasets and incorporating additional variables and data sources to enrich our analysis. Re-contact is an area of interest to researchers and politicians, and Statistics Canada is proud to be able to provide estimates of re-contact in all sectors of the justice system.

## References

1. Brennan, S., S. Franklin, A. Matarazzo, M. Radulescu, A. Reicker, D. Tapper, D. Towns. 2012. Re-contact with the Justice System: Nova Scotia Courts Evaluation. Statistics Canada internal documentation.
2. Brennan, S. 2016. Re-contact with the Justice System: Phase II results. Statistics Canada internal documentation.
3. Fellegi, I. P. and A. B. Sunter. 1969. A Theory for Record Linkage. In *Journal of the American Statistical Association*. 1183-1210.
4. Franklin, S., A. Reicker, D. Tapper. 2012. Identifying Multiple Contacts with the Toronto Police (2006-2010). Statistics Canada internal documentation.
5. Law, D., A. Reicker, L. Veilleux. 2014. Re-contact with the Saskatchewan Justice System: Police to Courts Linkages. Statistics Canada internal documentation.
6. Matarazzo, A. 2013. Re-contact with the Toronto Police Service. Statistics Canada internal documentation.
7. Statistics Canada. 2015. Integrated Criminal Court Survey Description.
8. Statistics Canada. 2015. Uniform Crime Reporting Survey Description.
9. Williamson, K., M. Montreuil. 2015. G-Link User Guide. Statistics Canada internal documentation.
10. Winkler, W. 1990. String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association.