GHENT
UNIVERSITY

FACULTY OF
LAW AND CRIMINOLOGY

Forensic DNA databases as data sources for criminological research

Sabine De Moor

Supervisors: Prof. Dr. Tom Vander Beken
Dr. Christophe Vandeviver
Doctoral Guidance Committee: Dr. Marre Lammers
Dr. Stijn Van Daele

Doctoral thesis submitted to obtain the title Doctor in Criminology
Academic year 2018–2019

Forensic DNA databases as data sources for criminological research

Sabine De Moor

October 2018

ABSTRACT

Criminological research and forensic intelligence are often based on crime data registered in police or court records. However, such data only includes information on crimes that have been detected or reported. Moreover, only a small proportion of these crimes are cleared, and research suggests that offenders who are not arrested have different features from those who are. For instance, they may be more specialized, have more co-offenders, or spread their crimes over a larger geographical or temporal area. These two elements – a high number of unknown offenders and the possibility that these offenders may not randomly be missing from official crime statistics – pose a major challenge for criminological research. Several authors have already questioned the generalizability and applicability of research findings based on (known offenders registered in) official crime data and the consequential implications for theory and policy.

The development of DNA profiling techniques at the end of the 20th century was promising in light of these difficulties, since this resulted in the establishment of national forensic DNA databases in many countries. The significant added value of DNA and DNA databases for specific operational police investigations soon became clear. Forensic DNA databases also contain unique opportunities for criminological research, as they can extend current forensic intelligence to unknown offenders registered in police databases. After all, DNA databases not only store profiles from known offenders (e.g., convicted offenders), but also contain profiles obtained from DNA samples found at crime scenes that can differentiate offenders that have not (yet) been arrested. These DNA profiles from unknown offenders can be linked to different crime scenes, or to other co-offenders. This makes it possible to study unknown offenders' serial and co-offending behaviour, which cannot be done using police-recorded crime data, for example.

To date, very little criminological research has been carried out using DNA databases, and their validity as a data source has largely been unexplored. The present dissertation aims to fill this research gap. In particular, it reviews the specific weakness and strength of DNA databases as a source of criminological research, respectively the selectivity of the source (only a proportion of all registered crimes is stored in the DNA database) and whether it allows the (links between) unknown offenders and their offending behaviour to be studied. To answer these research questions, four separate studies were conducted.

First, the different actors and factors that could impact the selectivity of the Belgian National Genetic Database are identified via a review of legislation and literature. Observations are also made on the actors involved in the process that a DNA trace profile passes through before it reaches the database. It was found that only a minority of all registered crimes is represented in the DNA database. This representation is mainly influenced by the type of crime (e.g., contact crimes where the offender makes

physical contact with the victim or with an object at the crime scene are more likely to be present in the DNA database) and by local DNA retention policies (e.g., judicial districts can opt to examine the crime scene of only certain types of crime for DNA traces).

Second, the validity of the DNA database as a ‘sample’ of all police-recorded crime data is further evaluated in an empirical study. More specific, DNA data from the Belgian National Genetic Database and police-recorded crime data from the Belgian General Police Database were compared regarding the concentration and spread of unsolved crimes across the different Belgian judicial districts. The findings suggest that the DNA database is representative of the police-recorded crime data. However, some limitations were found. First, crime is concentrated in different districts in each dataset. Second, the similarity of the spatial pattern for violent theft, lethal violence and sexual offences, but not for aggravated burglary, is caused in several districts by a low crime count in the police-recorded crime data and a zero crime count in the DNA data. In other words, in these districts no unsolved crimes or unknown offenders are registered in the DNA database, which precludes carrying out further analyses. Therefore, only using DNA data may limit the possibilities for criminological research. However, the selectivity of the DNA data should not be overstated – this study only uses a restricted dataset consisting of crimes committed in 2014, and the added value of DNA databases lies in combining them with police databases.

Third, network analysis is used in an empirical study to investigate the advantages of using an integrated dataset, combining police-recorded crime data (i.e., crimes committed by known offenders) and DNA data (i.e., crimes committed by unknown offenders). In the first network study, performed at the crime level, the prevalence and characteristics of serial co-offences obtained from the integrated dataset are compared to those obtained from the police-recorded crime data. In the integrated dataset, the number of crime networks is higher and their size and spatial and temporal distributions are larger than in the police dataset. Such findings could, for instance, place research and theory on offender mobility in a different perspective.

Finally, in a second network study, performed at the offender level, the difference between known and unknown offenders is assessed. Networks in the integrated dataset, combining known and unknown offenders, are larger. Thus, more co-offenders can be identified in the integrated dataset compared to the police-recorded crime data. These networks also have a different structure than networks only including known offenders. For example, unknown offenders are more peripheral nodes in the offender network. With respect to criminological research, using integrated datasets can therefore extend the knowledge on roles and structures within criminal networks.

In conclusion, this dissertation demonstrates that a DNA database is a valuable source for criminological research when its selectivity is taken into account. Arguments are provided that support the concern about

the generalizability and applicability of research findings that are only based on known offenders' criminal behaviour. Both the network studies emphasize the importance of integrating unknown offenders and their crimes in criminological research. Disregarding unknown offenders and their offending behaviour is found to have a considerable impact on the insights into crime and criminal behaviour. The dissertation concludes with a number of policy implications and future research opportunities.

SAMENVATTING

Binnen criminologisch onderzoek maakt men vaak gebruik van data uit politiedatabanken of gerechtelijke dossiers. Dergelijke gegevens omvatten echter alleen informatie over gedetecteerde of aangegeven misdrijven. Bovendien wordt slechts een klein deel van deze misdrijven opgehelderd. Daarnaast suggereert onderzoek dat daders die niet worden gevat andere kenmerken hebben dan daders die wel worden gevat, zoals meer specialisatie, meer mededaders, of een grotere temporele en ruimtelijke spreiding van hun misdrijven. Deze twee elementen - een groot aantal ongekende daders en de mogelijkheid dat deze daders niet toevallig ontbreken in de officiële criminaliteitsgegevens - vormen een grote uitdaging voor criminologisch onderzoek. Verschillende onderzoekers hebben de generaliseerbaarheid en toepasbaarheid van onderzoeksbevindingen gebaseerd op (gekende daders van) officieel geregistreerde misdrijven en de daaruit voortvloeiende implicaties voor theorie en beleid al in vraag gesteld.

De ontwikkeling van DNA-profileringstechnieken aan het einde van de 20ste eeuw was veelbelovend in het licht van het hierboven geschetst probleem. In veel landen resulteerde dit immers in de oprichting van een nationale forensische DNA-databank. De meerwaarde van DNA en DNA-databanken voor politioneel en gerechtelijk onderzoek werd al snel duidelijk. Ook voor criminologisch onderzoek bieden forensische DNA-databanken echter unieke mogelijkheden. DNA-databanken kunnen bijvoorbeeld de huidige inzichten die enkel gebaseerd zijn op gekende daders uitbreiden. In DNA-databanken worden immers niet alleen DNA-profielen van gekende daders (zoals veroordeelden) opgeslagen, maar ook profielen verkregen uit sporen die werden aangetroffen op plaatsen delict. Aan de hand van deze sporenprofielen kunnen (nog) niet gearresteerde daders onderscheiden worden en kunnen hun DNA-profielen gekoppeld worden aan meerdere plaatsen delict of mededaders. Dit maakt het mogelijk om het veelplegen en mededaderschap van ongekende daders te bestuderen, wat niet mogelijk is met data uit politiedatabanken.

Tot op heden wordt in criminologisch onderzoek weinig gebruik gemaakt van DNA-databanken. Bovendien werd de validiteit van DNA-databanken als databron nauwelijks onderzocht. Het huidige proefschrift beoogt deze lacune op te vullen. De centrale onderzoeksvraag luidt als volgt: Hoe kunnen DNA-databanken bijdragen aan de studie van ongekende daders en hun crimineel gedrag? Daartoe werden de specifieke zwakte en sterkte van DNA-databanken als bron voor criminologisch onderzoek bestudeerd. Aangezien slechts een deel van alle geregistreerde misdrijven wordt opgeslagen in de DNA-databank, is de selectiviteit van de DNA-databank een potentieel nadeel voor criminologisch onderzoek. Het kenmerkende voordeel van de databank is dat er (linken tussen) ongekende daders en hun crimineel gedrag bestudeerd kunnen worden. Om de onderzoeksvraag te beantwoorden, werden vier afzonderlijke studies uitgevoerd.

In een eerste studie werden verschillende actoren en factoren geïdentificeerd die van invloed kunnen zijn op de selectiviteit van de Belgische nationale DNA-gegevensbank. Daartoe werden wetgeving en literatuur bestudeerd, alsook observaties uitgevoerd bij de verschillende actoren die betrokken zijn in het proces dat een DNA-sporenprofiel doorloopt vooraleer het de databank bereikt. Er werd vastgesteld dat slechts een minderheid van alle geregistreerde misdrijven is vertegenwoordigd in de DNA-databank. Deze selectiviteit wordt voornamelijk beïnvloed door het type misdrijf. Bijvoorbeeld, contactmisdrijven waarbij de dader fysiek contact met het slachtoffer of met een object op de plaats delict maakt, worden vaker in de DNA-databank opgeslagen dan bijvoorbeeld internetfraude waarbij geen duidelijke plaats delict is. Ook het DNA-beleid speelt een belangrijk rol. Zo kunnen gerechtelijke arrondissementen bijvoorbeeld opteren om enkel de plaatsen delict van bepaalde misdrijftypes op DNA-sporen te onderzoeken.

In een tweede studie werd de validiteit van de DNA-databank, als zijnde een 'steekproef' van alle misdrijven opgeslagen in de politiedatabank, empirisch geëvalueerd. De concentratie van onopgehelderde misdrijven opgeslagen in de nationale DNA-gegevensbank in de verschillende Belgische gerechtelijke arrondissementen, en de spreiding van deze misdrijven over deze arrondissementen heen, werden vergeleken met deze in de Algemene Nationale Gegevensbank van de politie. De bevindingen suggereren dat de DNA-data representatief zijn voor de politiedata. Hierbij dienen echter enkele kanttekeningen geplaatst te worden. Vooreerst verschillen de beide datasets op het vlak van de arrondissementen waarin de onopgehelderde misdrijven geconcentreerd zijn. Daarnaast is het gelijkaardig ruimtelijke patroon in meerdere arrondissementen veroorzaakt door een laag aantal misdrijven in de politiedataset en een afwezigheid van misdrijven in de DNA-dataset. Dit werd vastgesteld voor diefstal met geweld, levensdelicten en zedendelicten, maar niet voor zware diefstallen. Met andere woorden, de DNA-dataset bevatte geen onopgehelderde misdrijven of onbekende daders in deze arrondissementen, wat verdere analyses onmogelijk maakt. Dit kan uiteraard het gebruik van DNA-databanken voor criminologisch onderzoek naar onopgehelderde misdrijven beperken. De selectiviteit van de DNA-data kan hier echter overschat zijn, aangezien in dit onderzoek gebruik werd gemaakt van een dataset die enkel bestaat uit misdrijven gepleegd in 2014. Bovendien ligt de toegevoegde waarde van DNA-databanken in de integratie ervan met de politiedatabank.

In de derde studie werd een netwerkanalyse uitgevoerd om de voordelen van een geïntegreerde dataset, waarin politiedata (i.e., misdrijven gepleegd door gekende daders) en DNA-data (i.e., misdrijven gepleegd door ongekende daders) werden gecombineerd, te onderzoeken. In deze netwerkstudie, uitgevoerd op het niveau van het misdrijf, werden de prevalentie en kenmerken van netwerken van feiten gepleegd door veelplegers in groep uit de geïntegreerde dataset vergeleken met deze uit de politiedata. In de geïntegreerde dataset is het aantal en de omvang van deze netwerken hoger dan in de politiedata, evenals hun ruimtelijke en temporele spreiding. DNA-databanken kunnen zodus het criminologisch onderzoek naar, en theorievorming over bijvoorbeeld de mobiliteit van daders van nieuwe perspectieven voorzien.

Ten slotte werd een tweede netwerkstudie, op het niveau van de dader, uitgevoerd om het verschil tussen gekende en ongekende daders te bestuderen. Netwerken in de geïntegreerde dataset bestaande uit zowel gekende als ongekende daders zijn groter dan deze uit de politiedataset enkel bestaande uit gekende daders. Er blijken met andere woorden meer mededaders te zijn dan men op basis van de politiedataset met enkel gekende daders kan afleiden. Netwerken bestaande uit gekende en ongekende daders hebben bovendien een andere structuur dan de netwerken die alleen gekende daders omvatten. Ongekende daders blijken bijvoorbeeld eerder perifere actoren in het dadernetwerk te zijn. Het gebruik van een geïntegreerde dataset heeft bijgevolg belangrijke implicaties voor criminologisch onderzoek aangezien ze de kennis over daderrollen en structuren binnen criminele netwerken kunnen vergroten.

Concluderend toont dit proefschrift aan dat een DNA-databank een waardevolle bron is voor criminologisch onderzoek indien rekening gehouden wordt met de selectiviteit ervan. Het proefschrift biedt argumenten die aantonen dat de generaliseerbaarheid en toepasbaarheid van onderzoeksresultaten die enkel gebaseerd zijn op gekende daders uit politiedatabanken beperkt zijn. Beide netwerkstudies benadrukken het belang van het betrekken van DNA-data betreffende ongekende daders en hun misdrijven in criminologisch onderzoek. De dissertatie wordt afgesloten met het formuleren van beleidsimplicaties en toekomstige onderzoeksmogelijkheden.

TABLE OF CONTENTS

ABSTRACT I

SAMENVATTING V

LIST OF FIGURES XV

LIST OF TABLES XVII

ACKNOWLEDGEMENTS XIX

PART 1: INTRODUCTION 1

1. Background 3
2. Forensic intelligence 6
  2.1. Forensic science versus forensic intelligence 6
  2.2. Sources of forensic intelligence 8
  2.3. Forensic intelligence research based on DNA data 16
3. Research objective and research questions 20
  3.1. Selectivity 21
  3.2. Networks of unknown offenders and their crimes 22
4. Data and methods 24
  4.1. Data collection approval 24
  4.2. Data selection 25
    4.2.1. Selection criteria 25
    4.2.2. The DNA dataset 28
    4.2.3. Police dataset 29
  4.3. Data limitations 30
  4.4. Operationalization of the concepts used 32
    4.4.1. Unknown offenders and unsolved crimes 32
    4.4.2. Validity and representativity 33
    4.4.3. Serial co-offending 34
      4.4.3.1. Serial offending 34
      4.4.3.2. Co-offending 35

4.5. Methods 36
4.5.1. Descriptive statistics 36
4.5.2. Spatial point pattern test 37
4.5.3. Social network analysis 38
4.5.4. Observations and interviews 39
5. Structure of this report 42

PART 2: THE SELECTIVITY OF DNA DATABASES 45

6. Forensic DNA and DNA databases 47
6.1. Forensic DNA 47
6.2. DNA databases 49
6.3. The Belgian National Genetic Database 50
6.3.1. Legal framework 50
6.3.1.1. National legislation 50
6.3.1.2. Transnational context and the Prüm Convention 51
6.3.2. Selection mechanisms and actors 52
6.3.3. Content NGDB 55
6.4. Conclusion 57
7. The validity of DNA databases in studying the spatial behaviour of unknown offenders 58
7.1. Introduction 58
7.2. Data 59
7.2.1. Study area and spatial unit of analysis 59
7.2.2. Police-recorded crime data and DNA data 60
7.3. Methods 60
7.3.1. Lorenz curve and Gini coefficient 60
7.3.2. Spatial point pattern test 62
7.4. Results 65
7.4.1. Descriptive statistics 65
7.4.2. Crime concentration 65
7.4.3. Spatial similarity 68
7.5. Conclusion and discussion 69

PART 3: NETWORKS OF UNKNOWN OFFENDERS AND THEIR CRIMES 75

8. Serial co-offending and the network approach 77

9. Integrating police-recorded crime data and DNA data to study serial co-offending behaviour 79
9.1. Introduction 79
9.2. Data 80
9.3. Methods 81
9.3.1. Social network analysis 81
9.3.2. Jaccard similarity index 83
9.3.3. Normalized diversity index 83
9.4. Results 86
9.4.1. Descriptive statistics 86
9.4.2. Jaccard similarity index 87
9.4.3. Spatiotemporal distribution 88
9.5. Conclusion and discussion 91
10. Assessing the missing data problem in criminal network analysis using DNA data 94
10.1. Introduction 94
10.2. Data 100
10.3. Methods 101
10.3.1. Network analysis 101
10.3.2. Monte Carlo simulation 103
10.4. Results 104
10.4.1. Descriptives 104
10.4.2. The effect of missing data on degree and betweenness 105
10.5. Conclusion and discussion 110

PART 4: CONCLUSION 115

11. Context 117
12. Summary of research findings 119
12.1. Selectivity 119
12.2. Studying unknown offenders and their crimes 121
13. DNA as the silver bullet? 124
13.1. Methodological importance 124
13.2. Theoretical importance 125
13.3. Practical importance 126
14. Limitations 127
15. Future research 129

15.1. Criminal career research and SNA 129
15.2. Prüm and the international context 129
15.3. Inclusion and retention criteria of DNA databases 130
15.4. The impact of DNA policies and DNA databases 131
16. Policy implications 133
16.1. Reinforce information exchange ANG and trace databases 133
16.2. Link forensic science and forensic intelligence 134
16.3. Awareness and education 135
17. The ethics of forensic DNA and DNA databases 137

REFERENCES 141

APPENDIXES 161

LIST OF FIGURES

Figure 1: Fictitious example of eight recorded crimes 5
Figure 2: Proportion of the four selected crime types in the NGDB (Jeuniaux, 2014) 26
Figure 3: Percentage of registered crimes per judicial district in the DNA dataset and police dataset 30
Figure 4: Actors involved in the forensic DNA process 39
Figure 5: Typology of participant observation researcher roles (Saunders et al., 2009, p. 293) 41
Figure 6: Structure of DNA (Genetics Home Reference, 2017) 48
Figure 7: Schematic representation of the contents of the NGDB 53
Figure 8: Selection mechanisms and actors involved, from collection to storage of forensic DNA in the NGDB 55
Figure 9: Number and type of DNA profiles in the NGDB per year 57
Figure 10: Graphical representation of the Lorenz curve and the Gini coefficient 62
Figure 11: Illustration of an SPPT 64
Figure 12: Lorenz curve for the police dataset and DNA dataset 67
Figure 13: Rank order crime rates in 27 judicial districts in the police dataset plotted against rank order in the DNA dataset (high to low) 67
Figure 14: Local S-indices for violent theft (1), aggravated burglary (2), lethal violence (3) and sexual offences (4) 69
Figure 15: Hypothetical example of three crimes and two offenders 81
Figure 16: Bipartite graph of a hypothetical two-mode network of six offenders and six crimes 82
Figure 17: One-mode networks of offenders obtained from the bipartite graph in **Figure 16** 82
Figure 18: Illustration of the Jaccard similarity index (Fuxman Bass et al., 2013, p. 1174) 83
Figure 19: Fictional example of six crimes committed in four years and in three districts 85
Figure 20: Spatial dispersion of crimes at network level for the police dataset 90
Figure 21: Spatial dispersion of crimes at network level for the integrated dataset 91
Figure 22: Network illustrating the concepts size, geodesics and density (Wasserman & Faust, 1994, p. 111) 101
Figure 23: Network G(V,E) 102
Figure 24: Betweenness centrality of each node coloured from least (red) to greatest (blue) (Rocchini, 2007) 103
Figure 25: Forest plot 95% confidence interval for degree of simulated-reduced networks 108
Figure 26: Forest plot 95% confidence interval for betweenness of simulated-reduced networks 110
Figure 27: The relation between forensic intelligence and forensic science 126
Figure 28: Spectrum of possible target groups stored in DNA databases 130
Figure 29: Schematic representation of policy implications 136

Figure 30: Local S-indices for violent theft (1), aggravated burglary (2), lethal violence (3), and sexual offences (4) for 2010 167
Figure 31: Local S-indices for violent theft (1), aggravated burglary (2), lethal violence (3), and sexual offences (4) for 2011 169
Figure 32: Local S-indices for violent theft (1), aggravated burglary (2), lethal violence (3), and sexual offences (4) for 2012 171
Figure 33: Local S-Indices for violent theft (1), aggravated burglary (2), lethal violence (3), and sexual offences (4) for 2013 173
Figure 34: Local S-indices for violent theft (1), aggravated burglary (2), lethal violence (3), and sexual offences (4) for 2015 175

LIST OF TABLES

Table 1: Forensic science versus forensic intelligence 8

Table 2: Overview of advantages and drawbacks of different data sources for criminological research 14

Table 3: Overview of the four research questions 23

Table 4: Description of the four selected prevention codes 27

Table 5: Number of registered crimes, per year and per crime type, in the DNA dataset 28

Table 6: Number of registered crimes, per year and per crime type, in the police dataset 29

Table 7: Number and type of DNA profiles in the NGDB per year 56

Table 8: Description of datasets 65

Table 9: Generalized Gini coefficients for different crime types 66

Table 10: S-index proportion difference test 68

Table 11: Description of the three datasets under study 87

Table 12: Jaccard similarities between edges present in the three different datasets. 88

Table 13: Temporal and spatial spread of the three datasets under study 89

Table 14: Overview of main characteristics and results of studies on missing data in networks illustrated in the text 98

Table 15: Data selection of unknown offenders from the NGDB 100

Table 16: Centrality measures of network G in **Figure 23** 102

Table 17: Characteristics of the real-world network and the real-reduced network 105

Table 18: Degree of real-world network and reduced networks 107

Table 19: Betweenness of real-world network and reduced networks 109

Table 20: S-index proportion difference test 165

Table 21: Description of datasets for 2010 166

Table 22: Generalized Gini coefficients for 2010 166

Table 23: Description of datasets for 2011 168

Table 24: Generalized Gini coefficients for 2011 168

Table 25: Description of datasets for 2012 170

Table 26: Generalized Gini coefficients for 2012 170

Table 27: Description of datasets for 2013 172

Table 28: Generalized Gini coefficients for 2013 172

Table 29: Description of datasets for 2015 174

Table 30: Generalized Gini coefficients for 2015 174

ACKNOWLEDGEMENTS

As everyone knows, 'time flies when you're having fun.' Although not every day of the past four years have been great fun, they have flown by, and I have truly enjoyed being a Ph.D. student. I had the opportunity to work on a fascinating subject, observe crime scene investigations, attend international conferences and learn statistics from the best. Needless to say, this was a unique opportunity for an 'eternal student' like myself. A number of people have contributed to this unique opportunity and helped make this dissertation possible.

First of all, my supervisor Tom. Thank you for believing in me even at moments when I did not. Christophe, thank you for being my co-supervisor. I learned a lot from you and you managed to challenge me to do better and better. Stijn, as my initial co-supervisor, your door was always open to me, and I appreciate that very much. Marre, thank you for all the advice and guidance.

Dirk, the effort you put into making the police data available was essential for the success of this dissertation. Patrick, thank you for providing the DNA data and for all the time and energy you spent to convince me of the 'beauty of R'. Your passion for statistics and science was contagious.

Thank you to all the departments of the police, the NICC/INCC and the Ministry of Justice who gave me insight in the 'DNA story'.

Thank you to the BRAIN-be programme, 'Belgian Research Action through Interdisciplinary Networks' of the Belgian Science Policy Office for giving the financial support to conduct this dissertation in the framework of the Be-Gen project ('Understanding the Operational, Strategic and Political Implications of the National Genetic Database') [BR/132/A4/Be-Gen].

Thank you to my Be-Gen colleagues Bertrand, Paul, Caroline and Inès. It was a pleasure to work in a team on related subjects, and to be able to discuss my research findings in a wider context. Caroline and Inès, I really enjoyed our conference trips together.

Thank you to all the Ghent University colleagues, past and present, for the pleasant breaks, amusing analyses of highly scientific TV shows and many more. Ciska, your pep-talk helped me to continue when I lost track. Rebecca, Elias and Joke, we have one common concern that helped me escape from Ph.D. madness: what / where are we going to lunch today? Thanks for all the encouragement and laughter!

Finally, I am very gratefull to have family and friends who support me, not only during this Ph.D. adventure, but in every way in life. Kristof, you certainly top that list.

1

PART 1:

INTRODUCTION

The introduction to this dissertation starts with an explanation of the background to the study, and reviews why and how the research was conducted. The concepts of *forensic science* and *forensic intelligence* are introduced, and the research questions are defined. The data and the methods used are then presented. The structure of the remaining parts of the dissertation is given at the end of this introduction.

1. Background

The vast majority of criminological research is based on recorded crime data from police or court records (e.g., Bernasco, 2006; Felson, 2003; Reiss, 1988; Weerman, 2001, 2003). It seems logical to use such data, since they provide a rich source of information about crimes, offenders, modus operandi, etc., and are continuously updated. Much of our knowledge about crimes, offenders and their behaviour is therefore based on the data stored in police databases. However, police-recorded crime data only include information on *detected or reported* crimes. As a result, they may underestimate the true magnitude of crime. Many crimes are not reported by victims or members of the public, or are not detected or recorded by the police. In scientific literature, this unrecorded proportion of all crimes is described as the 'dark number' (Biderman & Reiss, 1967). In addition, only a small proportion of all registered crime are cleared – about 20% in most Western countries (De Wree, Vermeulen, & Christiaens, 2006; Lammers & Bernasco, 2013), and not all co-offenders of a crime are identified. In fact, the police are not always aware of the existence of co-offenders. The identification of one offender or even several offenders does not imply that there were no other offenders involved in the crime, and it seems unlikely that every offender will spontaneously inform on all his co-offenders (Alarid, Burton Jr., & Hochstetler, 2009) or confess to their other offences (Kocsis & Irwin, 1998). Kocsis & Irwin (1998, p. 199) state that as "*some serial criminals have committed more offences than those for which they are charged, it would be inappropriate to assume that a person found guilty of only a single offence could not be a serial offender*". As a result, undetected crimes by known offenders, and unknown offenders and their crimes, remain out of reach if recorded crimes are the sole data source when studying offending behaviour.

Researchers who use police-recorded crime data to study offenders' criminal behaviour often generalize their findings to 'offender behaviour', although clearly the data on which they base their study do not include all offenders (e.g., Townsley & Sidebottom, 2010). Researchers usually mention as a limitation the fact that only information about (offenders of) (solved) recorded crime was available (e.g., Vandeviver, Neutens, Van Daele, Geurts, & Vander Beken, 2015). Also, very few studies mention the possibility that there could be differences in the behaviour and characteristics between known and unknown offenders. Are known and unknown offenders alike? Do they display the same behaviour? If the answer to these questions is *yes*, it would not be necessary to study unknown offenders and their crimes, as including data about unknown offenders in research would only yield 'more of the same' results. The conclusions would be the same, regardless of whether known or unknown offenders were studied. The only exception to this would be estimates of the size of the offender population; these would require information about both known and unknown offenders. However, if the answer to the above questions is *no*, then both known and unknown offenders and their crimes must be included in the study in order to be able to make valid statements about the general offender population and their crimes. This also implies that current research gives a distorted picture of offenders and their crimes, since it is (mainly) based on known offenders registered in police-recorded crime data.

Several authors have already expressed concern about the generalizability and applicability of research findings based on registered crime data, and the implications this has for both theory and policy (Lammers, 2013, p. 2). It is possible that the effectiveness of crime prevention initiatives and law enforcement deployment strategies could be compromised if they are based solely on analysing known offenders' offending behaviour. Arguably, prioritizing law enforcement strategies that target known offenders is lowering detection and clearance rates, and making unknown offenders invisible again. It is therefore important that the offending behaviour of unknown offenders is also understood, in order to develop new and inclusive prevention and law enforcement strategies (De Moor, Vandeviver, & Vander Beken, 2018b). However, empirical research into unknown offenders is still limited, precisely because of the difficulty of studying these offenders.

How, then, can the unknown offenders responsible for a large number of crimes in police-recorded crime data be studied? Forensic DNA databases offer opportunities here. Generally, there are two types of DNA profiles in forensic DNA analysis: 'unknown' forensic profiles obtained from samples gathered at crime scenes (e.g., saliva or drops of blood), and 'known' reference profiles obtained from samples taken directly from a person (e.g., a buccal swab from an offender, suspect or victim) (Jeuniaux, Duboccage, Renard, Van Renterghem, & Vanvooren, 2016). These forensic profiles are very interesting when considering a study of unknown offenders. First, an unknown offender can be differentiated from another known or unknown offender using a DNA profile from the crime scene. Second, various crimes can be linked to the same offender by means of a DNA profile that is found at various crime scenes, even if this offender is (as yet) unknown. Similarly, third, different offenders can be linked to one crime, even though they are unknown, by means of their (different) DNA profiles found at the crime scene. This enables the serial and co-offending behaviour of unknown offenders to be studied. Fourth, DNA databases can supplement the police database with information about unknown offenders, and this for each individual crime.

Figure 1 illustrates these four advantages. The figure presents eight crimes that were committed between 2011 and 2015 by a total of five different offenders (A, B, C, D and E). It shows information about the eight crimes that is stored in both the police database and the DNA database. Offenders A, B and E are unknown offenders stored in the DNA database, offenders C and D are known offenders stored in the police database. The DNA profile of offender B, for example, was found at the crime scene of a theft in Ghent in 2011 (crime 2), a theft in Ghent in 2013 (crime 3) and finally a theft in Liège in 2014 (crime 5). Moreover, for two of the three thefts, this unknown offender B was accompanied by one or more co-offenders known to the police. In other words, integrating a police database with a DNA database makes it possible to study the serial co-offending behaviour of the eight offenders, regardless of whether they are known or not. Using only police-recorded crime data would only give a partial image of the total offending behaviour.

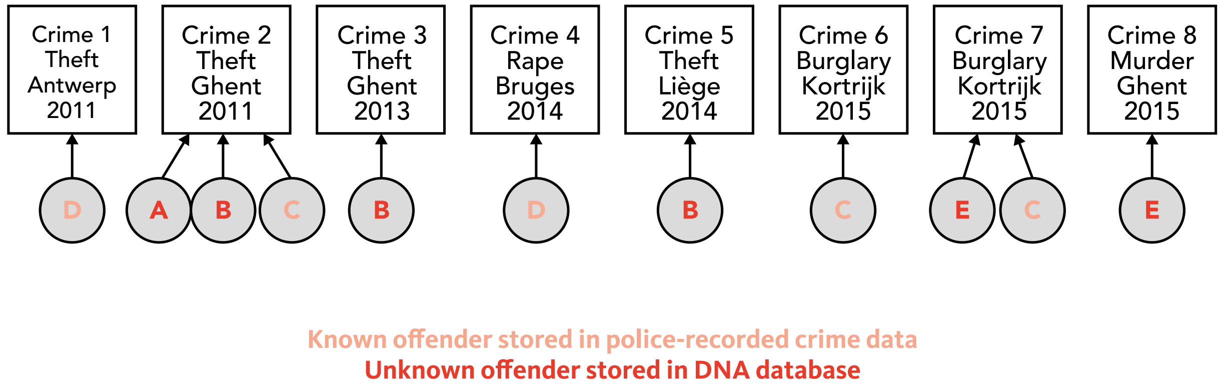


Figure 1: Fictitious example of eight recorded crimes

In theory, therefore, a DNA database is an ideal data source for studying the criminal behaviour of unknown offenders. But is that really the case? Can DNA databases be used as a meaningful source for criminological research, or are there disadvantages that would restrict or even prevent their use?

In contrast to the extensive research on the use of DNA to solve specific crimes, very little criminological research has made use of DNA databases. In addition, and more importantly, the validity of the data source has been largely unexplored. This dissertation therefore assesses the use of forensic DNA databases in the study of unknown offenders and their offending behaviour. The dissertation is of both of methodological and theoretical importance – methodological in the sense that a DNA database is evaluated as an alternative data source for criminological research; and theoretical in the sense that, by evaluating a DNA database, some substantive issues regarding unknown offenders are dealt with.

2. Forensic intelligence

Forensic DNA is a well-known and widely used form of evidence within police research. In this context, the use of DNA belongs to the field of forensic science, which is "*the application of scientific techniques and principles to provide evidence to legal or related investigations and determinations*" (Tilstone, Savage, & Clark, 2006, p. 1). In this dissertation, forensic DNA is used for forensic intelligence, which transcends the level of individual crime cases. A further explanation of both concepts is given below, and it will become clear that the boundary between the two is not strictly defined. The various sources that can be used for forensic intelligence, and their potential for studying unknown offenders, are evaluated, and existing scientific research using DNA databases is discussed. Once again, conscious attention is paid to the role that unknown offenders take in this research. This chapter on forensic intelligence is a further elaboration of De Moor, Vander Beken, and Van Daele (2017).

2.1. Forensic science versus forensic intelligence

DNA has a significant added value in specific police investigations (Martin, 2004; McCartney, Wilson, & Williams, 2011; Ribaux, Baylon, Roux, et al., 2010; Ribaux & Margot, 2003). The underlying principle is that matches can be made between different (unknown) crime scene samples mutually or between an unknown crime scene sample and a reference profile retrieved from a known offender. This comparative DNA analysis is intended to identify, directly or indirectly, offenders or to relate distinct cases. In addition to the purely reactive role of DNA profiles in confirming the evidence gathered by traditional investigative methods, DNA can also have a more proactive role (Williams & Johnson, 2005) as it sometimes provides the necessary evidence to solve a crime when the police have no leads and traditional investigative methods have failed to identify any (known) suspects (Doleac, 2016; Song, Patil, Murphy, & Slatkin, 2009; Tracy & Morgan, 2000; Weedn & Hicks, 1997; Wilson, Weisburd, & McClure, 2011). There are numerous anecdotal reports of success in matching a DNA profile from a crime-scene sample and a profile in a DNA database (i.e., cold hits), thereby identifying a formerly unknown potential suspect, and of the exoneration of wrongly convicted individuals on the basis of DNA evidence (Song et al., 2009; Weedn & Hicks, 1997). This exoneration may take place during the police investigation itself, as DNA analysis can help to eliminate innocent suspects. Of course, the mere presence (or absence) of a DNA profile from person X at a crime scene is not a sufficient condition to convict (or exonerate) person X. Further forensic casework research will have to evaluate the hypothesis.

In other words, DNA is used in specific police investigations to detect crime, and is criminal law driven (Burrows & Tarling, 2004; Jobling & Gill, 2004; Leary & Pease, 2003; Legrand & Vogel, 2012; Ribaux, Baylon, Roux, et al., 2010; Roux, Talbot-Wright, Robertson, Crispino, & Ribaux, 2015). This operational function of DNA traces (and other traces such as shoe marks, ear marks, tool marks, ballistics and fingerprints) is often referred to as 'forensic research' or 'forensic science'. Forensic science can therefore be defined as the study of unique crimes and the traces these crimes generate (Ribaux & Talbot Wright, 2014; Roux, Crispino, & Ribaux, 2012). DNA can also be helpful at a more strategic level. Beyond identifying individuals or relating cases to each other, the results of forensic research can also provide important information about the size, characteristics and evolution of different crime types. DNA databases are then seen as a tool to learn more about different types of crime (Bernasco, June 25 2014). Consequently, policy and other decision-making processes can be based on the retrieved information (Ribaux, Baylon, Roux, et al., 2010; Rossy, Ioset, Dessimoz, & Ribaux, 2013). This proactive approach to DNA databases is called 'forensic intelligence'. As in forensic research, in addition to DNA, other traces such as fingerprints, shoe marks, ear marks, tool marks and ballistics can contribute to forensic intelligence (Legrand & Vogel, 2014; Ribaux, Baylon, Roux, et al., 2010).

Both 'forensic science' and 'forensic intelligence' refer to specific aspects of the use of DNA traces and databases, although there is some confusion over the terminology used in practice. According to Ribaux, Baylon, Roux, et al. (2010), the concept of forensic intelligence is often wrongly used in the context of a forensic police investigation. The use of DNA sweeps in police investigations to confirm the identity of an offender (i.e., forensic science), for example, does not per se imply an intelligence-led perspective. This confusion may be reinforced by the many existing alternative concepts referring to the operational and strategic use of forensic DNA: 'forensic investigation' versus 'forensic analysis', 'database-led policing' versus 'intelligence-led policing' (Ribaux, Baylon, Roux, et al., 2010), 'crime analysis' versus 'criminal intelligence' (Ratcliffe, 2007) and 'tactical crime analysis' versus 'strategic crime analysis' (R. B. Santos, 2013, p. 61). However, forensic science and forensic intelligence are not completely distinct disciplines. On the contrary, both reinforce each other. On the one hand, information retrieved from a specific case by using forensic science (for example, via DNA sweeps) can be a source for forensic intelligence as it enables a better understanding of general crime patterns of offenders (i.e., the broader criminal context). On the other hand, patterns obtained from forensic intelligence can provide concrete guidance on how to find DNA and which traces should be looked for at different crime scenes (Ribaux, Baylon, Roux, et al., 2010). Without mentioning the two disciplines, Staley (2005, p. 38) clearly describes the interaction between forensic intelligence and forensic science as follows: "*A database that includes the people who are most likely to reoffend might help to ... catch them if they do reoffend. This would require the NDNAD* [the UK National DNA Database] *to reflect a better understanding of the crime patterns of offenders and their likelihood of reoffending*".

Table 1 summarizes the specific features of forensic science and forensic intelligence (Almog, 2014; Bieber, 2006; Burrows & Tarling, 2004; Cope, 2004; Harrison, 2006; Jamieson, 2004; Jobling & Gill, 2004; Leary & Pease, 2003; Legrand & Vogel, 2012, 2014; Morelato et al., 2014; Morelato et al., 2013; Ratcliffe, 2007; Ribaux, Baylon, Roux, et al., 2010; Ribaux, Crispino, & Roux, 2014; Ribaux et al., 2003; Ribaux & Margot, 2003; Ribaux & Talbot Wright, 2014; Robert, 2012; Rossy et al., 2013; Roux et al., 2012; Sheptycki, 2004; Tilley & Townsley, 2009).

Table 1: Forensic science versus forensic intelligence

Forensic science	Forensic intelligence
Operational	Strategic
Deductive	Inductive
What is happening / what happened?	Why it is happening / why did it happen?
Reactive, post-crime	Proactive
Crime detection	Crime prevention, reduction and control
Short term	Long term
Study of traces at crime scene	Study of series, patterns and case linking
Contextual and grounded	De-contextualized and de-personalized
Micro experiences	Macro picture
Single case approach	Broader criminal context
Particular offender(s)/offence(s)	General picture of characteristics of types of offender(s)/offence(s)
Technology and laboratory processes	Social science
Criminal investigation	Criminological research
Identification, attribution and arrest of offender(s)	Support decision-making and criminal policy
Criminal law driven	Life course and environmental criminology
Attempts represent less serious offences than committed offences	Attempts and committed offences both provide valuable information

2.2. Sources of forensic intelligence

Two types of data sources are generally used for assessing criminal behaviour in scientific research: (1)

recorded crime data (e.g., police-recorded crime data or conviction data); and (2) victimization surveys and self-report crime surveys[1] (D. S. Kirk, 2006; Maxfield, Luntz Weiler, & Spatz Widom, 2000; Thornberry & Krohn, 2000). With forensic DNA databases now established in many countries, a third instrument for 'forensic intelligence' is now available.

Recorded crime data, crime surveys and DNA databases have their own specific benefits and drawbacks as data sources. Numerous authors have studied the methodological aspects of data sources, although forensic DNA databases are hardly mentioned in their evaluations. Lammers and colleagues are the only authors to have comprehensively reviewed forensic DNA databases as a source for the study of criminal behaviour (Lammers, 2013, 2014a; Lammers & Bernasco, 2013; Lammers, Bernasco, & Elffers, 2012; Lammers, Bernasco, & van de Beek, 2011). **Table 2** presents an overview of the advantages and disadvantages of the different sources of information for criminological research. The first impression is that DNA databases will not be useful, since they have more disadvantages than advantages, and the same number of weaknesses as, or more than, the two other data sources. However, the discussion that follows suggests that the benefits DNA databases offer for the study of offending behaviour outweigh their drawbacks.

Recorded crime data

Official recorded crime data have two significant drawbacks when used as a data source for criminological research: selectivity of the registered data, and the fact that it relates to secondary data. Different elements are involved in the **selectivity** of recorded crime data. First, crimes must have been detected by the police (i.e., proactive activity of the police) or reported to the police by the public or by victims (i.e., reactive activity of the police) before they can be registered. However, the police only know about a small proportion of all crimes (Grapendaal & van Tilburg, 2002). Willingness to report a crime seems to depend on different factors concerning the victim, the crime and the social context (Goudriaan, Lynch, & Nieuwbeerta, 2003). Skogan identifies a variety of factors that influence reporting behaviour, including the seriousness of the offence, insurance payments, attitudes toward the police, feelings of guilt or culpability, and victim–offender relationships (Carcach, 1997; Skogan, 1984). The most consistent finding is that the seriousness of crimes has the greatest influence on reporting. Consequently, serious crimes that result in substantial financial or physical damage are generally reported more than petty crimes (Goudriaan et al., 2003; Skogan, 1984). Feelings of shame, guilt and a close relationship between the offender and the victim may also hinder the reporting of crimes. For example, sex crimes such as

---

[1] The term 'self-report surveys' usually refers to 'self-report offending surveys' (Tilley & Townsley, 2009, p. 15); however, it is also sometimes used in the context of 'self-report surveys of victimization' (Cantor & Lynch, 2000). Thus, offending and victimization measures may be surveyed using common methodological features (Goethals, Ponsaers, Beyens, Pauwels, & Devroe, 2002). In this dissertation the two types of survey will therefore be discussed together, but with any differences highlighted.

child sexual abuse, rape and prostitution are definitely under-reported by the public, although incidents of prostitution are often discovered by the police themselves (R. B. Santos, 2013). Similarly, domestic violence is reported much less than violence by strangers, although the difference in the willingness to report such crimes has reduced due to an increase in the public's awareness of the crime (Gartner & Macmillan, 1995). Victims with a lower social-economic status are less likely to report a crime than those with a higher status (Wittebrood, 2005). Younger victims and people who have been the victim of multiple crimes are also less willing to report offences to the police (Carcach, 1997). However, even if a crime is detected or reported, the offender often remains unidentified. In many Western countries clearance rates are low (De Wree et al., 2006; Lammers & Bernasco, 2013). The US Federal Bureau of Investigation (FBI)'s 2016 annual report mentions clearance rates of 45.6% for violent crimes and 18.3% for property crimes, for example (FBI, 2017). The second issue regarding selectivity is that not all crimes are officially recorded, even if they have been detected by the police or reported by the public. The police often have discretionary power to decide whether or not to register an offence (Brooks, 2015; Gilleir, 2013). Third, recorded crime data also reflect the priorities of the police. For example, the Belgian National Security Plan defines the police's principal security concerns over a timespan of four years. Radicalization, violent extremism and terrorism, smuggling and trafficking in human beings and drugs are the top three (of ten) priorities for the period 2016 to 2019 (Ministerraad, 2016). This is not an exhaustive list of the crimes that the police respond to; however, it obviously has an impact on the official crime rates. Fourth, several researchers claim that minorities are over-represented in recorded crime data, and consequently also in DNA databases (Greely, Riordan, Garrison, & Mountain, 2006; M. E. Smith, 2006). According to critics, this form of discrimination represents a broader social issue where minorities are targeted more often by the police (Rothstein & Talbott, 2006). To conclude, selectivity can lead to a substantial over- and under-reporting of certain criminal behaviours and offenders in recorded crime data.

The second significant drawback of recorded crime data and DNA databases is that both databases contain **secondary data**. The data are not collected by the researchers themselves (Boslaugh, 2010) – the databases are set up to serve the judicial system and the police in their operational research. Criminological research may require different or additional information than that gathered to solve specific crimes. Furthermore, they will have different stakeholders and will be set up using their own type of database, have different registration methods, etc. This makes the exchange or aggregation of the information stored in the two databases more complex. Thorough data preparation and data-cleaning procedures are needed to exclude cases with missing, invalid or incongruent information from further analysis.

Self-reports and victimization surveys

Self-reports and victimization surveys such as the International Crime Victims Survey (ICVS) have been introduced in an attempt to obtain more information on the relatively large number of unreported crimes (Coleman & Moynihan, 1996). Indeed, according to the results of the fourth ICVS conducted in 2000, only about 50% of six offence types (theft from cars, car vandalism, bicycle theft, burglary with entry, attempted burglary and theft of personal property) were reported to the police in 17 industrialized countries, including Belgium (van Kesteren, Mayhew, & Nieuwbeerta, 2000). However, crime surveys have significant drawbacks, too. Self-report and victimization studies remain an infrequent registration, in contrast with recorded crime databases and DNA databases, which are based on more regular, permanent documentation (Devroe, Beyens, & Enhus, 2006). Self-report studies and victimization surveys of criminal events also share the same pitfalls as any other questionnaire on any topic, with the reliability and validity of the surveys' methodology the main concern. The sensitive topics of crime and victimization cause additional difficulties, such as respondents providing socially desirable answers.[2] Nevertheless, there is an important difference between the two types of surveys where criminal behaviour is concerned. In contrast to victimization surveys, self-report surveys are more focused on minor offences and often do not include many of the more serious crimes for which people are arrested. This may lead to inconsistent results between victimization, self-report and recorded crime data, as Hindelang, Hirschi, and Weis (1979) discovered when comparing the characteristics of offenders (sex, race and social class) based on these three data sources (see also Thornberry & Krohn, 2000). Related to this finding, Babinski, Hartsough, and Lambert (2001) concluded that several high-frequency but less serious crimes such as vandalism and some low-frequency but serious crimes such as assault were less evident in the self-reports than in recorded crime data. On the other hand, a substantial number of individuals reported that they had committed crimes without being arrested, particularly public disorder crimes such as drug crimes or carrying a weapon.

The types of crime being studied and the target population of a survey both have an impact on the study's validity. Respondents must be willing to reveal a crime honestly and accurately, as the offender or as the victim (Haen Marschall, 1996). Their motivation and the overall balance of perceived costs (e.g., loss of time, disapproval, etc.) and benefits (e.g., incentive, appreciation, etc.) of completing the survey will influence their inclination to accurately and truthfully participate in a self-report survey. The perceived costs and rewards differ between diverse social groups (Haen Marschall, 1996). Hindelang, Hirschi, and Weis (1981, p. 213) stated that *"the self-report method can produce reliable and valid results within the populations to which it is generally applied"*, referring in their case to white, in school and generally not seriously delinquent groups. Thirty years later, Kivivuori (2011, p. 3) seemed to come to a similar conclusion, referring to the large amount of methodological literature on self-report crime surveys: *"self-report surveys are a reliable and fairly valid means of measuring the prevalence and incidence of crime especially in juvenile and young adult populations"*.

---

[2] For a detailed discussion of the problems typical of survey research, see Gideon (2012), de Leeuw, Hox & Dillman (2008) and Dillman, Smyth & Christian (2014).

Hindelang et al. (1981, p. 16) stated that the fact that *"the findings of research based on official data are consistent with the predictions of traditional theories of delinquency"* is one of the reasons for preferring recorded crime data to self-report data. However, this could be interpreted as a self-confirming theory, as policy-makers often rely on recorded crime data based on known offenders, and the priorities of the police are a derivate of the policy, which has repercussions for the data. On the other hand, these authors also stated that researchers interested in the aetiology of criminal behaviour prefer self-report studies as *"the findings of self-report research are clearly more consistent with modern theories, many of which were actually formulated before the results of self-report research became available"* (Hindelang et al., 1981, p. 15). Self-report surveys allow information to be collected on a broad range of personal and environmental factors that can be linked to crime and the criminal career: income, personal attributes and personality, social roles, socialization contexts (e.g., school, work, family and peers), etc. Recorded crime data such as police records generally do not include this kind of information (Huizinga, Esbensen, & Weiher, 1991; van Kesteren et al., 2000).

DNA databases

While many of the restrictions that are applicable to official police or judicial data are also (partially) valid when using DNA traces (Lammers et al., 2011), DNA has a significant added value for scientific research in the field of criminology. The four advantages mentioned at the beginning of this chapter are repeated and further explained here. First and foremost, DNA traces enable crimes carried out by unidentified offenders to be studied: an anonymous DNA profile is all that is required. In addition, DNA traces involving the same offender found at different crime scenes, and at different points in time (i.e., a serial offender), can be linked even if this offender has not previously been arrested or has not confessed to all their offences. This second advantage of forensic DNA is very important in the light of the firmly established characteristic of criminal behaviour that only a relatively small proportion of offenders are responsible for a large share of all crimes (Everson, 2003; Farrington et al., 2006; Farrington et al., 2003; Ferwerda, 2013; Grapendaal & van Tilburg, 2002; Home Office, 2004; Staley, 2005; Struijk, 2009; Thornberry & Krohn, 2000; Versteegh, Janssen, & Bernasco, 2003). The size of this 'small proportion' and 'large share' depend, *inter alia*, on the research method used, the definition of chronic offenders, the crimes taken into account and the population under study (Grapendaal & van Tilburg, 2002; Meijer, van Panhuis, Siero, & Smit, 2002). According to a 2004 report from the UK Home Office, for example, every year around 100,000 offenders are responsible for 50% of all offences in England and Wales (with a total population of more than 52 million people) and only 5,000 people commit around 9% of all crimes, accounting for about a million offences in total (Home Office, 2004). The third advantage of forensic DNA is that DNA traces involving different offenders found at one single crime scene can be linked even if these offenders have not previously been arrested. Moreover, it is also possible to identify networks of (unknown) co-offenders by relations of transitivity (Legrand & Vogel, 2014; Ribaux, Baylon, Roux, et al., 2010; Ribaux, Walsh, & Margot, 2006). For example, if person A shares case 1 with person B, and if person B shares case 2 with person C, then persons A, B and C are part of the same 'network'. Fourth, a DNA database can add further information about unknown offenders to crimes stored in the police database. This allows links to be made between unknown offenders and known offenders in each of the databases separately, and also between known and unknown offenders mutually.

Thornberry and Krohn (2000) cited Thorsten Sellin when they stressed the advantage of self-report studies in obtaining information as close to the source of criminal behaviour as possible: *"the value of a crime rate for index purposes decreases as the distance from the crime itself in terms of procedure increases"* (1931, p. 346). Although at the time of Sellin's publication in 1931 the structure of DNA had not even been discovered, his observation may be even more relevant when using DNA as a data source because DNA is collected at the crime scene itself and does not rely on information from victims or witnesses. However, DNA techniques are not infallible. Many researches have questioned the uniqueness of DNA profiles and the match probability.[3] The fact that DNA profiles are extremely rare does not make them unique. The problem of false positive matches (i.e., a match is found where no match should be found), and even false negative matches (i.e., a match that should be found, is not found), fits in this context. The number of false positive matches can be reduced by increasing the number of required matching loci (van der Beek, 2011) (see 6.1). Also, DNA matching may be impeded by mixed profiles (from offender(s), victims and persons not of interest), the availability of only small amounts of DNA, and contamination at the crime scene and/or later on (Balding, 1999; Bramley, 2009; Jobling & Gill, 2004; Lee, Lee, Park, & Hwang, 2001; Meulenbroek, 2009; National DNA Database, 2004; Penacino, Sala, & Corach, 2003; Staley, 2005). Of course, the problem of mismatches, either positive or negative, is more significant when DNA is used as evidence than when it is used to study criminal behaviour.

In addition to the fact that DNA databases only store registered crimes, they may also suffer from another form of selectivity. DNA profiles can only be obtained at a minority of crime scenes. Offenders who do not shed DNA easily (i.e., 'bad' shedders), or crimes without a real crime scene, such as money laundering, may limit or prevent the retrieval of suitable DNA profiles, for example. Moreover, DNA profiles are not always collected at crime scenes, even if they are likely to be present. Different elements influencing this selectivity are further discussed in Chapters 6 and 7 of this dissertation.

---

[3] Match probability refers to the chance of two unrelated people sharing the same DNA profile.

Table 2: Overview of advantages and drawbacks of different data sources for criminological research

	ADVANTAGES	DRAWBACKS
Recorded crime data: arrest and conviction data	- Extensive geographical area covered - Extensive time period covered: no reference period - Data recorded regularly - Instrument validity: consistent crime categories within juridical districts - Findings are consistent with traditional theories of delinquency - Absence of responder bias	- Crime must be reported to/detected by the police - Higher-frequency or more serious crimes are more likely to involve police contact and registration - Rely on accurate and consistent record keeping - Reflect the working of the official agencies instead of the actual crimes being committed in society - Changes to laws or crime categories may impede temporal comparisons - Restricted access to the data
Crime surveys: self-report and victimization studies	- Ability to study unreported/undetected crimes - Ability to study the circumstances, motives and aetiology of the offence: information on individual, familial, environmental and other aspects can be explored - As the general population is usually the target population, a random sample from home addresses, schools, National Insurance numbers/Social Security numbers, etc. gives easy access to the target population - Findings are consistent with modern theories of delinquency - International crime surveys make comparisons between nations possible, by transcending national differences in criminal law and recording practices - Ability to work with longitudinal designs	- Measurement error caused by incorrect responses to survey questions: a) Dependent on the individual's willingness to truthfully/accurately report criminal activity/victimization b) Socially desirable responding c) Memory decay and telescoping d) Respondents' errors in classification of the delinquency items - Sample design and selection (i.e., sampling error and non-coverage error) - Participation and response rate (i.e., non-response error) - Validity/reliability of survey instrument - Survey content: a) Not always clear what 'crime/delinquency' represents b) Too few types of crime are covered by questions c) Victimless crimes like drug abuse and economic or organizational crimes are not surveyed (victimization survey)

		- Generally low comparability of surveys (with other surveys and recorded crime data): a) No standard of delinquency items that are surveyed b) Varying reference period - No permanent registration - Expensive and time-consuming data collection method
DNA database	- Extensive time period covered: no reference period - Data recorded regularly - Ability to detect unknown offenders - Ability to study the criminal career of unknown offenders (i.e., serial offending) - Ability to link unknown offenders with (un)known offenders (i.e., co-offending) - Ability to compare identified offenders/networks with unidentified offenders/networks - DNA data are collected early in the investigative process and therefore suffer less attrition - DNA data are 'objective'	- Crime must be reported to/detected by the police - DNA can only be obtained at a minority of crime scenes: • Not always an obvious crime scene • Useable DNA not always left at crime scenes • Offenders can be 'good' or 'bad' DNA shedders • Interval between leaving DNA behind and recovery of DNA - DNA not always collected at crime scenes, even if available in all probability - Not all DNA samples collected at crime scenes are analysed (i.e., no DNA profile available) - Reflects the working of the official agencies instead of the actual crimes being committed in society - Uniqueness of profile and match probability, false positives/negatives (e.g., mixed samples, contamination, etc.) - Is the retrieved DNA from the offender or from another person? - Offender characteristics cannot be studied (e.g., age etc.) - Forensic awareness amongst offenders may lead to fewer DNA traces at crime scenes - Restricted access to the data - Retention period of DNA profiles stored in the database is limited

(Source: Babinski et al., 2001; Balding, 1999; Beauregard & Bouchard, 2010; Beaver, 2014; Bernasco, 2008; Bramley, 2009; Brantingham & Brantingham, 1993; Cantor & Lynch, 2000; Dillman, 1991, 2000; Eck & Weisburd, 1995; Farrington, 2001; Farrington & Ttofi, 2014; Forensic Science and Pathology Unit, 2005; GeneWatch UK, 2006; Goethals et al., 2002; Gottfredson & Hindelang, 1977; Hindelang et al., 1981; Jobling & Gill, 2004; Johnson, 2013; Junger-Tas & Haen Marshall, 1999; Kazemian, Pease, & Farrington, 2011; D. S. Kirk, 2006; Kivivuori, 2011; Lammers, 2013; Lammers et al., 2011; Lee et al., 2001; Lovell et al., 2017; Maxfield et al., 2000; McGloin, Sullivan, Piquero, & Pratt, 2007; Meulenbroek, 2009; National DNA Database, 2004; Nazaretian & Merolla, 2013; Penacino et al., 2003; Raymond, van Oorschot, Gunn, Walsh, & Roux, 2004; Roman et al., 2008; Thornberry & Krohn, 2000, 2003; Tilley & Townsley, 2009; Townsley, Smith, & Pease, 2006; van Kesteren et al., 2000; van Kesteren, van Dijk, & Mayhew, 2014; Wallace, 2006; Wiles & Costello, 2000)

From the analysis above it should be clear that every data collection method has its own qualities and specifications. Whatever data source is used, keeping its individual drawbacks in mind during the research process remains a critical research approach. The choice of a specific data source must be weighted in the context of the research goal. Self-report studies seem to be more appropriate for the study of the aetiology of crime, whereas DNA databases seem to be extremely suitable for the study of unknown serial co-offenders. Of course, self-report surveys can also reveal information on offenders of registered but unsolved crimes stored in the police database. As a researcher, however, you have no assurance that these crimes are actually registered (and under which qualification). Moreover, this information cannot be linked to the specific crime registered in a police database. These are disadvantages that do not apply to DNA databases.

2.3. Forensic intelligence research based on DNA data

There are four types of scientific research that use DNA as a data source or as a research topic:

(1) Studies on the technique of DNA: loci, PCR, DNA reports (e.g., Malsch, Taverne, Elffers, de Keijser, & Kranendonk, 2013); different sampling techniques (e.g., Barash, Reshef, & Brauner, 2010); match probability and false positives (e.g., Kloosterman, Sjerps, & Quak, 2014; Meulenbroek, 2009; Penacino et al., 2003; Thompson, 2008); mixed profiles (e.g., J.-A. Bright, Curran, & Buckleton, 2014); etc.

(2) Studies on the legal and ethical issues surrounding the use and exchange of DNA: privacy, legitimacy, protection of human rights, prevention of misuse, function creep, civil liberties, criteria for inclusion, social and ethical costs (e.g., Bramley, 2009; De Gorgey, 1990; Forensic Genetics Policy Initiative; GeneWatch UK, 2006; Levitt, 2007; McCartney, 2004; McCartney et al., 2011; Prainsack & Toom, 2013; Staley, 2005; Tracy & Morgan, 2000; Voultsos, Njau, Tairis, Psaroulis, & Kovatsi, 2011; Wallace, 2006; Wallace, Jackson, & Thibedeau, 2014; Williams & Johnson, 2004, 2005); public attitudes (e.g., Hochschild & Sen, 2012; Machado & Silva, 2014; McCartney et al., 2011); etc.

(3) Studies on forensic science: the use of DNA at case level compared to other more traditional forms of investigation, to learn more about resolving specific cases, clearance rates and (cost-) effectiveness (e.g., Bond, 2007; Burrows & Tarling, 2004; Doleac, 2016; Martin, 2004; Ribaux, Baylon, Lock, et al., 2010; Ribaux & Margot, 2003; Roman et al., 2008; Rossy et al., 2013; Townsley, Smith, & Pease, 2005; Tracy & Morgan, 2000; Wilson et al., 2011); hit rates (e.g., Campbell, Pierce, Sharma, Feeney, & Fehler-Cabral, 2016); etc

(4) Studies on forensic intelligence: the use of DNA for criminological research to learn more about criminal behaviour and the broader criminal context.[4]

In this dissertation, category (4) is of specific interest. A review of the scientific literature reveals that there is not much research available that uses a DNA database to study crime and offending behaviour. Below, three studies that can be seen as the beginnings of DNA-based criminological research are discussed. The three studies made use of the UK National DNA Database (NDNAD), which was set up in 1995 and was the first national forensic DNA database in the world. Only five years later, Wiles and Costello (2000) published the first study (partially) based on data from the NDNAD.

- Research on offender mobility
  Wiles and Costello (Costello & Wiles, 2001; Wiles & Costello, 2000) integrated three different data sources in their study of the mobility of high-volume offenders in and around Sheffield: geocoded police-recorded crime data, the NDNAD, and offender interviews. They explained the use of the NDNAD to study the travel patterns of offenders, irrespective of whether they are known to the police or not. The main finding of the researchers was that most offender movements are relatively short.

- Research on criminal careers
  Leary and Pease (2003) used data from the NDNAD over a 19-month period from one police force area (West Midlands) to study criminal careers. They stated that the number of crime scene samples submitted to the NDNAD increased in that period, whereas the number of resulting matches with offenders already in the database stagnated. They concluded from this that the population of active offenders is a fast-evolving group. Many offenders recorded in the NDNAD no longer offended, whereas many of the active offenders may have been ‘newcomers’ not yet present in the database.

  Taking this finding a step further and integrating the criminal career literature stating that offenders are more generalists than specialists, implications for DNA retention policies arise, according to Leary and Pease: in a fast-changing, versatile offender population, DNA sampling from offenders as well as from crime scenes should not only be expanded to a wide range of offences but should be taken at the first available opportunity. Subsequent studies by Tseloni and Pease (2011) and Kazemian et al. (2011) also focused on DNA retention. These authors did not use a DNA database to study criminal behaviour as such; instead, they used findings from criminal career research to devise an evidence-based policy on DNA profile retention.

---

[4] Although the study of how (certain types of) criminal behaviour could be influenced by a person’s genes is part of the fourth type of research, the study of the genetic links to criminal behaviour is beyond the scope of this dissertation.

- Research on offender specialization

  Townsley, Smith and Pease (Townsley et al., 2005, 2006) used the NDNAD to study the criminal careers of offenders, and more specifically offender specialization. A central research question is 'What other types of crime do offenders committing a serious offence also commit?' For this purpose they compared crime scene samples from all solved serious cases ((attempted) murder, manslaughter, sexual offences, rapes, and various types of burglary) within the Metropolitan Police Service jurisdiction for the year 2003 with all profiles in the NDNAD. Eleven per cent generated a match with an offender already stored in the NDNAD, indicating offender versatility that is in accordance with criminal career research.

It is notable that other researchers referring to (one of) these studies have paid little (e.g., Fox, 2010) or no (e.g., Westerberg, Grant, & Bond, 2007) attention to the fact that the researchers used DNA profiles as a data source. In all probability, this is due to the fact that the researchers making use of the NDNAD studied the entire offender population, without making a specific distinction between known and unknown offenders. The use of DNA had no impact on the interpretation of the results, as the researchers made statements about the entire offender population. Two other studies do make this distinction: Lammers and colleagues (Lammers, 2013, 2014a, 2014b; Lammers & Bernasco, 2013; Lammers et al., 2012), who used the Dutch DNA database; and Jeuniaux and colleagues (Jeuniaux et al., 2018; Jeuniaux et al., 2016), who used the Belgian National Genetic database (NGDB). These researchers highlighted the specific 'added value' that DNA databases can bring to a study of unknown offenders, and its importance for criminological research.

- Research on probability of arrest

  Using the Dutch DNA database to compare arrested and non-arrested offenders, Lammers and colleagues (Lammers, 2013, 2014a, 2014b; Lammers & Bernasco, 2013; Lammers et al., 2012) studied the characteristics of offending behaviour that could influence the probability of being arrested. Five crime types were analysed: violent crimes, sex offences, burglary, theft and theft of/from a car. The researchers concluded that: "*The more crimes an offender commits, the greater the probability that he will be arrested, specialized offenders have a smaller probability of being arrested than generalists, and as the number of police regions in which an offender commits his crimes increases his probability of being arrested decreases*" (Lammers, 2013, p 98).

- Research on co-offending networks

  Jeuniaux et al. (2016) used the NGDB to study co-offending networks. They performed a network analysis on the more than 60,000 genetic profiles stored in the database in 2014. Both the forensic profiles and the reference profiles were integrated in the analysis. A total of 445 networks were identified and described in terms of the number of people, the judicial district and the types of crime involved. The operational use of network intelligence is also discussed. Networks of particular interest are those that are neither completely identified (i.e., only containing reference profiles of identified suspects or offenders) nor completely unidentified (i.e., only containing forensic profiles), because the likelihood of identifying unknown offenders is increased when they are linked to known offenders.

In other research based on the NGDB, Jeuniaux et al. (2018) studied the criminal career of 7,535 offenders. In total 17,887 crimes were involved. Again, both the unknown forensic profiles and the known reference profiles were integrated in the analysis. Although the researchers acknowledged in both studies the selectivity of the DNA data and suggested that future research should make a comparison with other data such as police-recorded crime data, they clearly emphasized the unique feature of DNA databases – that both known and unknown offenders can be studied.

The empirical studies conducted for this dissertation – on the validity of DNA databases (De Moor, Vandeviver, & Vander Beken, 2018a), on serial co-offending behaviour (De Moor et al., 2018b) and on missing data in offender networks – also belong in this overview. These three studies will be extensively discussed later in this dissertation.

3. Research objective and research questions

Forensic DNA databases offer the opportunity to extend forensic intelligence to unknown offenders of crimes registered in police databases. In the previously mentioned study by Wiles and Costello (Costello & Wiles, 2001; Wiles & Costello, 2000) forensic DNA data was used to supplement other (police) data. The researchers used both reference profiles of known offenders and forensic profiles of unknown offenders. They did not distinguish between the two types of profile. The researchers therefore made statements about the entire offender population, without being able to prove any difference between known and unknown offenders. Lammers (2013) explicitly made this distinction, and in her research on probability of arrest found a (limited) difference between known and unknown offenders. However, she only used DNA data to make statements about both known and unknown offenders. DNA data represent only a proportion of all offenders, whether known or unknown, because not every crime registered in the police database provides DNA profiles that are stored in DNA databases. Therefore, DNA databases only contain data on certain crimes, and thus do not include all offenders. Using only a DNA database to study offenders and their offending behaviour can therefore give a distorted picture.

In the (limited) research that has so far used a DNA database as data source, too little attention has been paid to the question of whether such databases are suitable for criminological research. The selectivity of the data source is usually mentioned, but not examined further. However, it is important to establish which unknown offenders and which crimes are actually stored in a DNA database, before starting research that examines the difference between the behaviour and characteristics of known and unknown offenders.

This brings us seamlessly to the goal of this dissertation, which is to assess the usability of DNA databases as a data source for criminological research. The resulting central research question is:

How can DNA databases contribute to the study of unknown offenders and their criminal behaviour?

The research question is operationalized by evaluating the specific characteristic weakness and strength of DNA databases as a source of criminological research: the selectivity of the source versus the ability to study (links between) unknown offenders and their offending behaviour. This results in four research questions that are explained below and are schematically presented in **Table 3**.

3.1. Selectivity

Generally, all the crimes in a DNA database are also included in police-recorded crime data. Conversely, however, this is not the case. As already mentioned, only a minority of all registered crimes is represented in a DNA database, as not every crime scene is swabbed for DNA, and not every crime type provides suitable DNA profiles. Crimes such as money laundering, for example, are generally not represented in DNA databases. Before a DNA profile can be stored in a DNA database, it must pass various selection processes. The answers to the first research question provide an overview of these selection processes.

Q 1: Which actors and factors have an impact on the selectivity of DNA databases?

Because of its selectivity, a DNA database is only a 'sample' of all crimes stored in a police dataset. Before a DNA database can be used as a criminological data source, it is therefore important to question the representativeness of this sample. Criminologists frequently use police-recorded crime data and, although it is also selective, researchers generally regard it as a valid data source. Therefore, the second research question weighs the validity of DNA databases against the validity of police-recorded crime data. Since the potential of DNA databases lies specifically in the study of unknown offenders and their behaviour, the research question is limited to unsolved crimes.

Q 2: Are DNA databases less valid than police-recorded crime databases for the study of unsolved crimes?

Both of these research questions are answered in the second part of this dissertation. The first chapter, which reviews legislation, literature and information on the actors involved, explores which actors and factors have an impact on the selectivity of the DNA database used in this study (i.e., the Belgian National Genetic Database). In the subsequent chapter these findings are tested in an empirical study in which the validity of this DNA database is evaluated by comparing the spatial distribution of unsolved crimes in the DNA data with police-recorded crime data from the Belgian General Police Database.

3.2. Networks of unknown offenders and their crimes

Although DNA data contain fewer crimes than police-recorded crime data, DNA databases have specific advantages that can make a fundamental contribution to the study of crime and criminal behaviour: the ability to study unknown offenders and their (serial and co-)offending behaviour. Moreover, by integrating police-recorded crime data and DNA data, links between known and unknown offenders can be identified, which is not possible with any other data source. This also offers new research possibilities, as it is very unlikely that known and unknown offenders are two completely separate groups.

The question that arises from these opportunities is whether a different picture of crime emerges if the study is based on both police-recorded crime data and DNA data compared to the crime picture only arising from police-recorded crime data. As these two types of database store information on both crimes and offenders, the picture can be studied on both these levels, resulting in two different research questions:

Q 3: To what extent does the crime picture change when unsolved crimes are included in a network analysis?

Q 4: To what extent does the offender profile change when unknown offenders are included in a network analysis?

These two research questions are answered in the third part of this dissertation, using an integrated dataset in which police-recorded crime data are supplemented with DNA data. Both research questions are answered using a network analysis of two case studies, at the level of the crime and the level of the offender.

Table 3: Overview of the four research questions

	Research question 1	**Research question 2**	**Research question 3**	**Research question 4**
Research question	**Which actors and factors have an impact on the selectivity of DNA databases?**	**Are DNA databases less valid than police-recorded crime databases for the study of unsolved crimes?**	**To what extent does the crime picture change when unsolved crimes are included in a network analysis?**	**To what extent does the offender profile change when unknown offenders are included in a network analysis?**
Unit of analysis		Registered crime: Unsolved crimes	Registered crime: Solved and unsolved crimes	Registered offenders: Known and unknown offenders
Data sources	Literature, interviews and observations	DNA data and police-recorded crime data	DNA data and police-recorded crime data	DNA data and police-recorded crime data
What?	Validity of DNA as data source compared with police-recorded crime data	Validity of DNA as data source to study unsolved crimes	Difference/change in crime picture	Difference/change in offender profile
How?		Police-recorded crime data as a standard against which the validity of DNA is weighed	Police-recorded crime data are supplemented with information from the DNA database that is only available in the DNA database	Police-recorded crime data are supplemented with information from the DNA database that is only available in the DNA database
Part of dissertation	Part 2 – Chapter 6	Part 2 – Chapter 7	Part 3 – Chapter 9	Part 3 – Chapter 10

4. Data and methods

Two types of databases are used in this dissertation: police-recorded crime data and DNA data. The police-recorded crime data are extracted from the Belgian General Police Database (ANG – in Dutch: Algemene Nationale Gegevensbank). This database is the main source of nationwide crime statistics in Belgium (Van Daele & Vander Beken, 2011). All reported and detected crimes are registered in this database. The Belgian General Police Database contains, inter alia, information on crimes, offenders, modus operandi and victims. The DNA data are extracted from the Belgian National Genetic Database (NGDB), and includes a proportion of all crimes stored in the Belgian General Police Database. The NGDB will be discussed further in Chapter 6.

4.1. Data collection approval

The fulfilment of this dissertation relied on access to both the NGDB and the Belgian General Police Database. Requests to obtain and analyse DNA data and police-recorded crime data were submitted to the relevant judicial authorities, the federal police, and to the DNA Index System service (DIS) of the National Institute for Forensic Science and Criminology (NICC/INCC – in Dutch: Nationaal Instituut voor Criminalistiek en Criminologie), which manages the NGDB.

The 1999 DNA Law[5] requires the DIS service to comply with strict data access control of the NGDB. Because of this, and in order to comply with the law of 8 December 1992 on the protection of privacy, the advice of the Belgian Commission for the Protection of Privacy (CPP) was sought. An authorization application to the Sectoral Committee for the Federal Government did not appear to apply. A charter of trust between Ghent University (UGent) as receiver of the data and the NICC/INCC as provider of the data was drawn up in which the UGent declared it would use the provided DNA data only in the framework of the Be-Gen project and in the context of this dissertation that falls within the bounds of the Be-Gen project. The DIS staff extracted all crimes that met the selection criteria, which are presented in the next section.

Regarding the police-recorded crime data, a declaration of subsequent processing of coded personal data was submitted at the CPP. Permission was received from the General Commissioner of the Federal Police to access the Belgian General Police Database. The Directorate of Police information and ICT resources (in Dutch: DRI – Directie van de politionele informatie en de ICT-middelen) defined the practical modalities of the data access. An internship at the Central Directorate of the Fight Against Serious and

---

[5] Belgian DNA law of 1999, discussed in Chapter 6.

Organized Crime (in Dutch: DGJ-DJSOC – Directie van de bestrijding van de zware en georganiseerde criminaliteit) was completed during the period 1 October 2015 to 31 January 2016. The internship offered the opportunity to obtain insight into the content and structure of the Belgian General Police Database. The Directorate's staff extracted all the crimes that met our selection criteria (see 4.2.1).

Before the data from the NGDB and the Belgian General Police Database was handed over to the researcher in the first half of 2016, an agreement between DIS and DGJ–DJSOC was made to provide the data in a coded manner. All crimes have a unique number (in Dutch: PV-nummer), and are registered in the NGDB and/or in the Belgian General Police Database by means of this unique number. The same key was used to encode this number in both datasets. This ensured that, although encoded, the information on a crime registered in both the NGDB and the Belgian General Police Database could be combined. Links could then be made between the two databases based on offenders' presence at shared crime scenes.

4.2. Data selection

4.2.1. Selection criteria

The same selection criteria were used for both the DNA data and the police-recorded crime data. Although both crimes and offenders are studied in this dissertation, the initial selection of data was based on registered crimes. Both the datasets used in this research link offenders to crimes, and each crime is registered using a specific PV number.

The crimes (i.e., PV numbers) selected for this study were committed in Belgium and registered in one of 27 judicial districts.[6] Two important selection criteria were applied: the year in which the crime was committed (i.e., the year it was registered, derived from the PV number) and the type of crime. The NGDB was established in Belgium in 2002, and of course police-recorded crime data have been available for much longer, but this study was limited to crimes registered in 2010 to 2015 inclusive.

In the context of this dissertation, crime types were selected according to whether DNA traces could be found at the crime scene. Crimes such as, for example, money laundering or libel were not included, as they don't produce DNA traces and therefore would not add value to this study. Four crime types were selected from the list of 90 official prevention codes used to register crimes in Belgium: aggravated burglary, violent theft, sexual offences and lethal violence **Table 4**. They were chosen because they were four of the top five most frequently occurring crimes in the national DNA database (representing,

---

[6] Until the reform of the judicial landscape in 2014, Belgium had 27 judicial districts. An increase in scale then reduced the number of districts from 27 to 12. In this dissertation, the old structure is used.

respectively, numbers 1, 2, 3 and 5[7]) at the time of the data request (**Figure 2**). Most of the profiles that are stored in relation to crimes with the prevention code 11 and 37 are of convicted offenders; while prevention codes 17 and 30 are mainly crime scene profiles (Duboccage, 2008; Jeuniaux et al., 2016).

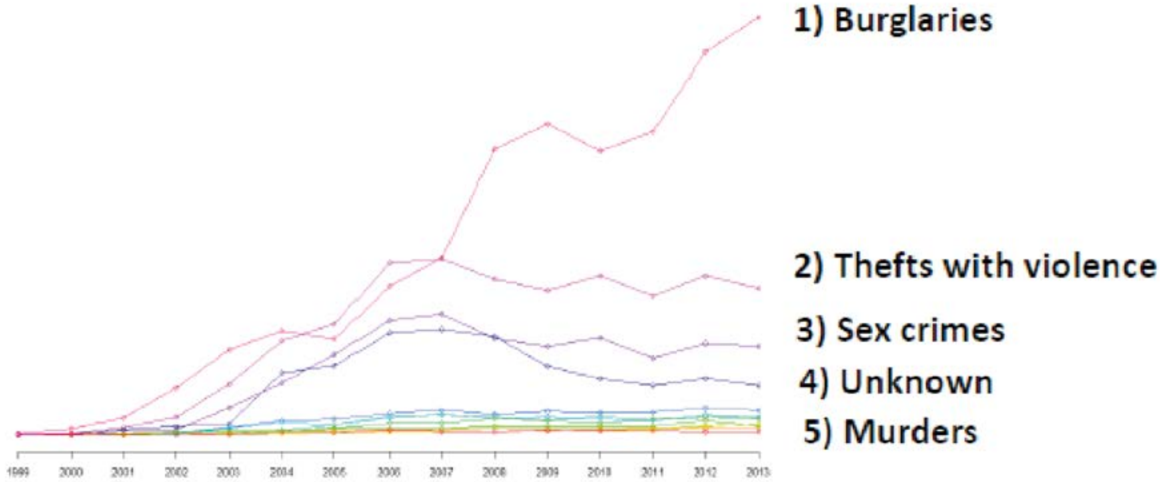


Figure 2: Proportion of the four selected crime types in the NGDB (Jeuniaux, 2014)

---

[7] Number 4 contains the group 'unknown'.

Table 4: Description of the four selected prevention codes

Prevention code	Description
11	Violent theft: Theft through violence or threats Theft where weapons are shown or used Extortion Carjacking Home jacking
17	Aggravated burglary: Theft through burglary, climbing in or false keys Theft through burglary of domestic residence Attempted theft through burglary, climbing in or false keys Car theft through breaking in, climbing in or false keys Theft of bicycle or motorbike through burglary, climbing in or false keys Theft from garage Raid Theft from a car
30	Lethal violence: Murder Manslaughter Manslaughter to facilitate theft Attempted murder or attempted manslaughter Genocide law – law of universal jurisdiction Poisoning (Articles 392, 394 and 397 of the Belgian Criminal Code) Intentional administration of harmful substances resulting in accidental death (Articles 402 and 404 of the Belgian Criminal Code)
37	Sexual offences: Rape Indecent assault Public indecency Voyeurism To provoke fornication Fornication of minors Fornication Prostitution Pimping of adults Obscene movies, prints, objects or books Paedophilia Trafficking in human beings – sexual exploitation (Article 433quinquies §1, 1 ° SWB) Pimping of minors Child pornography Publicity about prostitution and fornication Incest Unwanted intimacy

The DNA data and the police-recorded crime data were used to answer three of the four research questions (Q 2–4). A specific data selection was made for each research question, and these are explained in more detail in Parts 2 and 3. This section provides a general overview of the available data, after data cleaning. The DNA dataset and the police-recorded crime dataset (hereafter 'police dataset') described below contain all recorded crimes between 2010 and 2015 for the four selected crime types. In other words, no distinction was made between solved and unsolved crimes and the offenders involved.

4.2.2. The DNA dataset

The provided DNA dataset contains 25,492 crimes. The crime rate in 2015 appears to be notably lower than in the other years in both datasets. Although there was a generally determined crime drop in most advanced countries in that year (Aebi & Linde, 2010; Farrell, Tilley, & Tseloni, 2014), another explanation for this phenomenon is that, at the time of the data extraction in March 2016, not all of the 2015 crimes had been processed by the NGDB and the Belgian General Police Database.

Aggravated burglary is by far the most common category in the DNA dataset (62.36%). The second, much smaller, category is violent theft (22.26%), followed by sexual offences (11.51%) and lethal violence (3.87%) (**Table 5**).

Table 5: Number of registered crimes, per year and per crime type, in the DNA dataset

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**Total**	**%**
11	1,160	1,235	1,255	957	745	323	**5,675**	**22.26**
17	2,410	3,160	3,807	3,255	2,190	1,076	**15,898**	**62.36**
30	166	224	160	187	159	90	**986**	**3.87**
37	670	676	613	513	322	139	**2,933**	**11.51**
Total	**4,406**	**5,295**	**5,835**	**4,912**	**3,416**	**1,628**	**25,492**	
%	**17.28**	**20.77**	**22.89**	**19.27**	**13.40**	**6.39**		**100**

There is a big difference between judicial districts in terms of the crime level. Hasselt has the highest crime level at 17.10% and Arlon the lowest at 0.26% (mean = 3.70, standard deviation (S.D.) = 4.42). Hasselt (17.10%), Antwerp (15.19%), and Brussels (11.51%) are the only judicial districts in the DNA dataset with a crime rate of at least 10%. Verviers, Ypres, Veurne, Eupen, Neufchâteau, Dinant, Marche-en-Famenne and Arlon each account for less than 1% of the total crime count in the DNA dataset (**Figure 3**).

4.2.3. Police dataset

The police dataset contains far more crimes (N = 1,186,061) than the DNA dataset (N = 25,492) (**Table 6**). As in the DNA dataset, the crime rate in 2015 appears to be notably lower than in the other years.

Aggravated burglary is also the biggest crime category (83.07%) in this dataset, followed by violent theft (11.43%) and sexual offences (5.02%). Lethal violence accounts for only 0.48% of all registered crimes.

Table 6: Number of registered crimes, per year and per crime type, in the police dataset

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**Total**	**%**
11	22,892	26,442	25,395	22,218	21,008	17,555	**135,510**	**11.43**
17	173,229	182,127	179,742	169,549	163,892	116,770	**985,309**	**83.07**
30	904	995	988	951	980	849	**5,667**	**0.48**
37	10,236	10,651	10,098	10,679	10,459	7,452	**59,575**	**5.02**
Total	**207,261**	**220,215**	**216,223**	**203,397**	**196,339**	**142,626**	**1,186,061**	
%	**17.47**	**18.57**	**18.23**	**17.15**	**16.55**	**12.03**		**100**

As in the DNA dataset, there is a big variance between the crime rates in the different judicial districts. The highest crime level is 26.39% (Brussels) and the lowest is 0.33% (Eupen) (mean = 3.70, S.D. = 5.32). Brussels (26.39%) and Antwerp (10.90%) are the only judicial districts with a crime rate of at least 10%. Oudenaarde, Arlon, Marche-en-Famenne, Neufchâteau, Ypres, Veurne and Eupen each account for less than 1% of the total crime count (**Figure 3**). There is also a difference in ranking between the crime rates in the different districts in the DNA dataset and the police-recorded crime dataset. This difference will be illustrated further in Chapter 7.

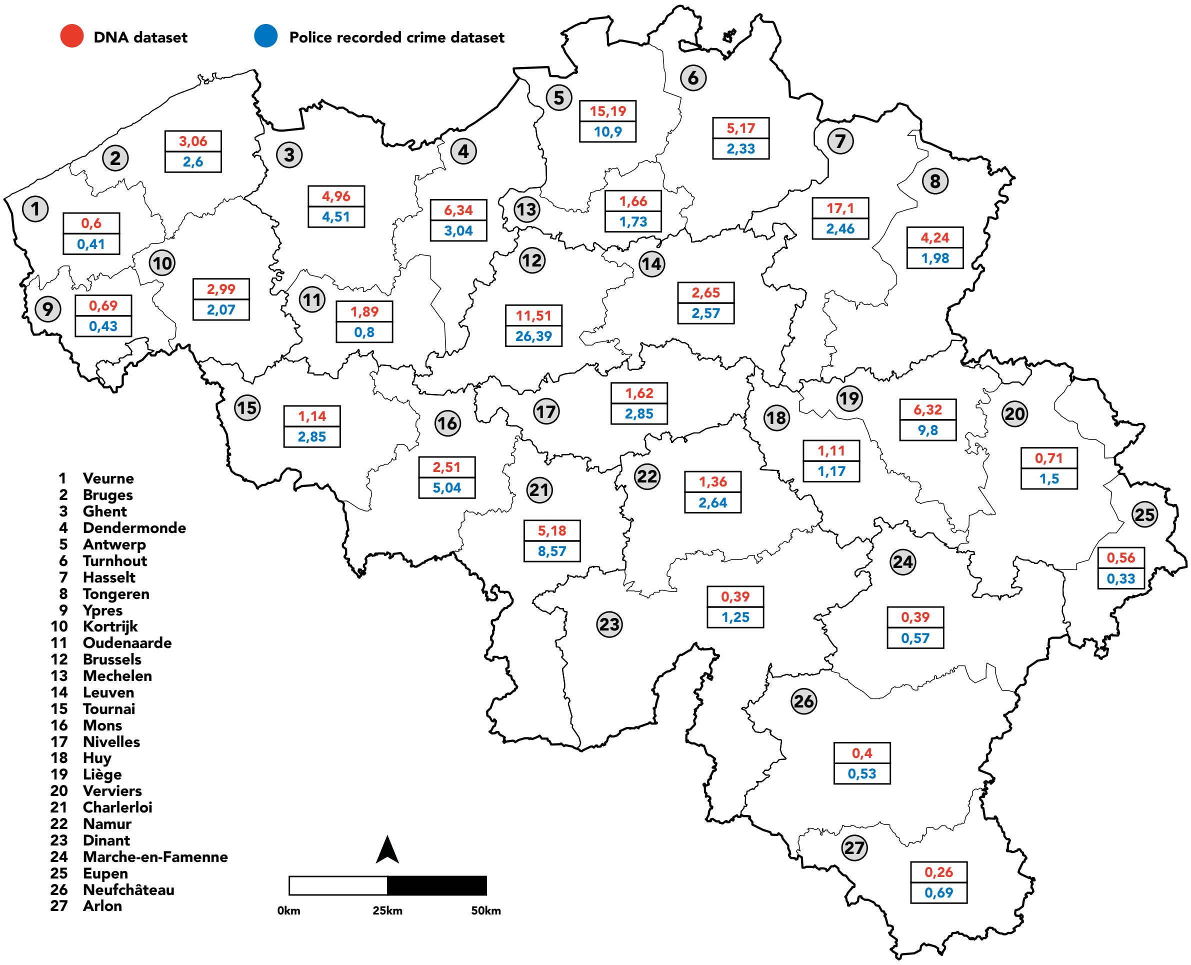


Figure 3: Percentage of registered crimes per judicial district in the DNA dataset and police dataset

4.3. Data limitations

As discussed above (see 2.2), each type of data source has particular advantages and disadvantages. However, there are also other restrictions that apply specifically to the data used in this dissertation.

First, the available data has a **high degree of aggregation on two levels**, which can be attributed to the fact that the research is based on secondary data. The first level applies to the temporal and spatial aggregation of the data. Partly because information about the location and the time of a crime is not important for the operation of the NGDB, the information stored about these features is limited (Jeuniaux et al., 2018). The NGDB only stores the year and judicial district in which the crime was registered. The Belgian General Police Database stores much more detailed spatiotemporal information, often referring to the exact time and location of the crime. However, due to privacy concerns, only aggregated spatiotemporal information was received from the Belgian General Police Database: the year and the judicial district[8] in which the crime was registered. Criminological research has demonstrated that the same dataset aggregated at a different spatial and/or temporal level can lead to different results (Linning, Andresen, & Brantingham, 2016; Ratcliffe, 2004b; Steenbeek & Weisburd, 2015). Especially high aggregated levels, as applied in our research, may hide variation between months or census tracts, for example. At the spatial level, this is called the modifiable areal unit problem (MAUP) (Openshaw, 1984; Ratcliffe, 2004b), which can be interpreted as a geographic manifestation of the ecological fallacy (Waller & Gotway, 2004, p. 104). However, the aggregated data may not be an issue in this dissertation for two reasons. First, this study focuses on the spatiotemporal variation between, rather than within, the DNA and police datasets – i.e. the potential difference between the two datasets, irrespective of the level of aggregation, rather than the spatiotemporal distribution of crime. Second, the judicial district level is a particularly suitable spatial unit of analysis as DNA retention policies and police priorities are (partly) made at this level. This will become clear in the next part.

A large degree of aggregation also applies at the level of the crime types being studied. Not all of the crimes that fall under a selected prevention code are eligible for inclusion in the DNA database. For example, the category of 'sexual offences' includes voyeurism or offensive films, prints, objects or books. Therefore, the relative importance of DNA may be underestimated, as several crimes may be integrated in the analyses for which no DNA profiles can exists.

Second, our analysis is restricted to a **small number of crime types**, albeit these are the four most frequently recorded crime types in the NGDB. Expanding the research to all crime types included in the NGDB, such as drug-related crimes, could reveal other results. Nevertheless, as mentioned before, the crimes stored in the NGDB are limited to particular types. Crimes without a real crime scene (e.g., financial or internet crimes) or obvious crime scene (e.g., crimes in public places, such as shoplifting), for example, cannot be represented in a DNA database. This third limitation is in the same line as the second limitation mentioned here. Both point to the selectivity of the DNA database, the impact of which is studied in the second part of this dissertation.

The networks that are identified may have been even larger and more dispersed if other crime types, such as drug use or organized crime, had been integrated into the analysis. This seems likely, as most offenders are generalists and do not limit themselves to committing only one type of crime (Leary & Pease, 2003; Nieuwbeerta, Blokland, Piquero, & Sweeten, 2011; Piquero, Farrington, & Blumstein, 2003).

Third, DNA profiles extracted from crime scene samples are stored in the NGDB on the assumption that the DNA profile comes from (one of) the offender(s). However, it is not inconceivable that some

---

[8] One of the 27 former judicial districts.

crime scene samples come from **third parties**, such as the victim or individuals who have entered the crime scene (e.g., police, medical personnel, etc.). However, police officers are trained in crime scene awareness and skills, and will collect only relevant DNA samples and prevent contamination (Lammers, 2013; UNODC, 2009). Moreover, when searching for crime scene samples, the police make use of their knowledge and experience in order to reconstruct a crime (i.e., crime scene reconstruction or CSR). In CSR the central question is not what traces are present, but what could have happened. Different scenarios are taken into consideration that guide the police in their search for traces of the offender(s). Taking a reference sample from a third party in order to differentiate any traces of this non-suspect from those of the alleged offender(s) is expressly provided for in the Belgian DNA Law of 2011 (Art. 44sexies SV, Belgisch Staatsblad, 2011a), although only limited use is made of this facility. In this dissertation, an extra measure against this limitation was taken by studying serial offending (Jeuniaux et al., 2016; Lammers, 2013). The probability that a person's DNA profile is found at multiple crime locations, without this person being the offender, is very small indeed.

Finally, throughout this dissertation, the individuals stored in the NGDB and the Belgian General Police Database are systematically referred to as '**offenders**'. Except for the reference profiles stored in the NGDB, however, their status is 'suspect', since at the time of registration they will not have been convicted. Some may never be convicted because they are not proven guilty (i.e., individuals stored in the Belgian General Police Database) or because they are never caught and their guilt cannot be proved or disproved (i.e., individuals stored in the NGDB). However, this limitation is not unique to this dissertation and is shared by a lot of other criminological research into offenders and their offending behaviour based on police-recorded crime data.

Despite the weaknesses and uncertainties surrounding the (DNA) data, the use of both databases provides a unique opportunity not only to study offending patterns of both known and unknown offenders, but also to make links between the offenders and their crimes, irrespective of their status (known or unknown).

4.4. Operationalization of the concepts used

4.4.1. Unknown offenders and unsolved crimes

In this dissertation, reference is often made to *unknown offenders* (as opposed to *known offenders*) and *unsolved crimes* (as opposed to *solved crimes*). Before explaining these terms, it is important to consider the distinction between the concepts *identification* and *individualization*, which are often used in forensic science. Identification refers to placing a trace in a restricted class (e.g., identifying a fluid as blood). Individualization refers to the person (or object) leaving this trace behind. According to Kirk, identification is not a goal as such, but it is a preliminary step to individualization (P. L. Kirk, 1963, p. 236).

Taking this distinction into consideration, the concept of the *unknown offender* may have a different meaning depending on the reference frame: the NGDB or the Belgian General Police Database. A crime registered in the Belgian General Police Database committed by an unknown offender implies that the police could not individualize a single offender of that crime. The crime is *detected* by the police, but the offenders are not. As a result, no information is available about the offender(s) and the crime is *unsolved*. In other words, the name of the offender is not available. As soon as one offender has been detected and thus individualized, the crime will be referred to as a *solved crime*, even though there may be other unknown offenders involved.

Where in the Belgian General Police Database an unknown offender is the equivalent to the absence of an offender in the database, the NGDB contains information on *unknown offenders*. A DNA profile extracted from a sample found at a crime scene makes it possible to individualize an offender. This forensic profile is used to infer the presence of an undetected offender (Rossy & Morselli, 2018). So, even though the profile does not match with a reference profile of a previously convicted offender, for example, this unknown offender can still be individualized by his unique DNA profile. Although these unknown offenders can be linked to one or more crimes, these crimes remain *unsolved* as no offender can be arrested and prosecuted for the crime.

4.4.2. *Validity and representativity*

Is the NGDB a valid source for studying unknown offenders and their crimes? Are the (unsolved) crimes stored in the NGDB representative of the (unsolved) crimes stored in the Belgian General Police Database? These are central questions in this dissertation. Therefore, it is important to clarify the concepts ‘validity’ and ‘representativity’.

Validity refers to the extent to which research actually measures what it intends to measure, and only that. It refers to the absence of systematic errors or bias (Golafshani, 2003; Pauwels, 2017). Validity can be seen as a measure of accuracy (Bloor & Wood, 2006). Translated to our research, the NGDB is a valid data source when it captures the unsolved crimes stored in the Belgian General Police Database. As will be demonstrated in the next part, different processes and actors influence the amount and type of forensic profiles stored in the NGDB. As a result, bias is not inconceivable.

Representativity is related to validity. The (unsolved) crimes stored in the NGDB are a sample of the (unsolved) crimes stored in the Belgian General Police Database. The question is whether the smaller sample (i.e., the NGDB) accurately reflects (represents) the larger population of interest (i.e., the Belgian General Police Database).

4.4.3. Serial co-offending

Part 3 evaluates the NGDB's usefulness for criminological research by studying serial co-offenders and their behaviour. Although one may intuitively have an idea of what serial offending and co-offending represents, a more detailed description is needed.

4.4.3.1. Serial offending

Research on serial offending is related to research on the *criminal career*. Criminal career literature approaches four components of offending behaviour longitudinally: the *prevalence* or the proportion of the population who commit crime, the *frequency* or the number of crimes an offender commits, the *duration* or the length of time between the first and last crime, and the *seriousness* of the crimes (Blumstein & Cohen, 1987; Piquero et al., 2003). Serial offenders are also called *repeat offenders*, *persistent offenders* or *chronic offenders* in the literature. Most research on serial offending relates to serial murder, serial rape and serial arson (Kocsis & Irwin, 1998).

Serial offending may be defined in relation to the *number of victims* of an offender, or the number of targets in the case of arson, for example. The minimum number of victims required varies across research, but three or four victims are common. Another way to define serial offending is by taking into account the *number of crimes* committed by a single offender (Kocsis & Irwin, 1998). Strictly speaking, an individual can be defined as a serial offender as soon as he has committed more than one crime. The higher this limit is, the lower the prevalence of serial offenders will be. Both these number-based definitions are influenced by a temporal criterion. Criminal career research indicates that a time lapse of even ten years without committing a crime is not exceptional, so the number of serial offenders will increase as the research period increases (Metcalfe & Baker, 2014). Short study periods may therefore underestimate the number of serial offenders. Many studies use a number-based definition of serial offending and conclude that only a relatively small proportion of offenders are responsible for a large share of all crimes. Finally, serial offending can be defined using a *propensity* approach. An offender may have psychological characteristics that drive him to re-offend. Therefore, offenders can be defined as serial offenders even if there is no evidence that they are responsible for more than one crime (Grapendaal & van Tilburg, 2002; Kocsis & Irwin, 1998; R. B. Santos, 2013).

In the context of this dissertation, a number-based definition of serial offending is applied: individuals committing more than one registered crime. These crimes may be limited to only one of the four selected crime types or may be a variety of these four types, committed in the time span of six years.

4.4.3.2. Co-offending

According to van Mastrigt (2014), co-offending is not only a general but also a well-documented phenomenon, although most of the knowledge on co-offenders seems to be based on research on juvenile offenders (Felson, 2003). In the academic literature, there are several synonyms used for 'co-offending', including accomplice offending, concurrent felonies, multiple offender crimes, companionate offences and joint offending (see for example Carrington, 2014; Cheatwood, 1980; DuRant, Cadenhead, Pendergrast, Slavens, & Linder, 1994; Juodis, Woodworth, Porter, & Ten Brinke, 2009; Reiss & Farrington, 1991; Tontodonato, 1996; Wiles & Costello, 2000). Even 'gangs', 'organized crime' and 'group offending' are often cited in the context of co-offending, although there are some conceptual differences between those items (Carrington, 2014; Francis, Humphreys, Kirby, & Soothill, 2013; Reiss, 1988; Warr, 1996).

The most simplistic definition describes co-offending as more than one offender jointly conducting a crime at the same time and place (Carrington, 2014; Felson, 2003; Paternoster, 2014; Reiss, 1988; Weerman, 2003; Weerman & Kleemans, 2002). Tremblay (1993, p. 20) defines co-offending more broadly: "*The term co-offenders, ... refers not only to the subset of an offender's pool of accomplices but rather to all those other offenders he must rely on before, during and after the crime event in order to make the contemplated crime possible or worthwhile*". For example, advisers helping to prepare a burglary or buyers of the stolen goods are also co-offenders of the thief, according to Tremblay's definition (Weerman, 2014). Furthermore, offenders do not always co-offend with all their delinquent acquaintances. Therefore, a distinction must be made between active 'offending groups' who have already committed crimes together and the broader 'accomplice networks' consisting of potential co-offenders for future crimes (Warr, 1996).

In addition to the number of offenders, several other characteristics also define co-offending. Weerman described eight characteristics of co-offending (Weerman, 2003; pp. 399-401). First, there are three forms of offending: exclusive solo offending; mixed solo and co-offending; and exclusive co-offending. Most offenders can be classified in the second type, as they commit both joint and lone offences (Goldweber, Dmitrieva, Cauffman, Piquero, & Steinberg, 2011; Reiss, 1988; Reiss & Farrington, 1991; van Mastrigt, 2014). However, being a mixed offender does not mean they commit an equal amount of solo and co-offences. These offenders seem to have a preference, as they generally commit crimes either by themselves or with accomplices. Second, although most crime types can be committed by multiple offenders, the level of co-offending varies according to the type of offence. For example, burglary and robbery are typical co-offending crime types, in contrast to sex crimes (Carrington, 2014; Reiss & Farrington, 1991). Third, the prevalence of co-offending decreases with age, as it does for offending in general. This may be due to a change in the criminal population, with solo offenders carrying on and co-offenders stopping offending. Another explanation, which is confirmed by the research of Reiss & Farrington (1991), is that it is caused by a change in individual criminal careers: co-offenders act alone more as they get older. Fourth, co-offending is usually incited by one of the co-offenders. Fifth, co-offending groups can range from clearly defined criminal groups with a division of labour and different roles, to relatively loose alliances and networks (Weerman & Kleemans, 2002). However, more complex co-offending groups seem to be made up of more professional offenders. Sixth, the vast majority of co-offences are committed by offenders working in pairs (Carrington, 2014). Associations of four or more offenders is not rare in late childhood and early adolescence, but these groupings become smaller as the offenders get older (Warr, 1996). Seventh, co-offenders are relatively homogenous with regard to age, gender and crime experience (Carrington, 2002; Reiss, 1988; Reiss & Farrington, 1991; van Mastrigt & Carrington, 2014; Warr, 1996). Eighth, the partnerships between different offenders can be temporary (only for one crime, for example) or more permanent in nature (Weerman & Kleemans, 2002). Except for more specialized groups, offenders do not usually commit multiple offences with the same co-offender (Reiss & Farrington, 1991). It is therefore not surprising that co-offenders often belong to multiple offending groups at the same time (Warr, 1996).

In this dissertation the basic definition of co-offending is used. A co-offender is defined as an individual who committed at least one registered crime together with one or more other offenders at the same time and place. The co-offenders in our research can therefore exhibit mixed solo and co-offending behaviour, or exclusive co-offending behaviour. Both the offending group and the accomplice network are taken into account. Taking both concepts – serial offending and co-offending – together, a serial co-offender is defined in this dissertation as an individual who has committed more than one registered crime, of which at least one crime was committed with another offender at the same time and place.

4.5. Methods

The various methods used in this dissertation are briefly explained below. All statistical analyses were performed using the R software environment (R Core Team, 2018). A detailed description of the statistical methods is provided in the relevant chapters of Parts 2 and 3. However, the method of 'observations and interviews' is only discussed in this introduction because of its importance throughout the entire dissertation.

4.5.1. Descriptive statistics

Basic descriptive statistics such as the sample size (in subgroups) and measures of dispersion and central tendency such as the mean, minimum, maximum and standard deviation were computed in this dissertation. In addition, several other descriptive statistics were computed: the Jaccard index, normalized diversity indexes, the Lorenz curve and the Gini coefficient.

Both the **Lorenz curve** and the **Gini coefficient** were developed to measure inequality of income or wealth but can be applied to any distribution (Bernasco & Steenbeek, 2017), including crime concentration across judicial districts, as in this dissertation (Chapter 7). The Gini coefficient is derived from the Lorenz curve, which graphically presents the concentration of crime. A Gini coefficient of 0 indicates that every judicial district has the same crime level (i.e., complete equality) and a value of 1 indicates that all crime is concentrated in only one district (i.e., complete inequality) (Bernasco & Steenbeek, 2017; Eck, Lee, O, & Martinez, 2017).

Normalized diversity indexes were computed to assess the spatial and temporal variability within the crime networks obtained in Chapter 9. The indexes range between 0 and 1. A value of zero indicates that there is no diversity as all crimes are committed within the same district/year. A value of one indicates that all crimes are evenly spread over the different possible districts/years (Mazerolle, Brame, Paternoster, Piquero, & Dean, 2000, pp. 1153-1154).

The **Jaccard index** measures similarity between two datasets. In this dissertation, the index is used to measure the similarity between crime networks present in the police-recorded crime data and crime networks present in an integrated dataset containing both police-recorded crime data and DNA data (Chapter 9). The minimum index is 0 (or 0%), which means that both datasets have no data points in common. The index reaches a value of 1 (or 100%) when there is perfect overlap of the two datasets (Simpson, Lyday, Hayasaka, Marsh, & Laurienti, 2013).

4.5.2. Spatial point pattern test

The non-parametric spatial point pattern test (SPPT) was developed by Andresen (2009) specifically to test the similarity between the spatial point patterns in two datasets – a base dataset and a test dataset – within specified spatial units of analysis. In criminology, an SPPT is used to study the similarity of spatial crime levels at different spatial units (e.g., de Melo, Matias, & Andresen, 2015), over different time frames (e.g., Andresen, Linning, & Malleson, 2017), and/or between different crime types (e.g., Andresen & Linning, 2012). Several authors have also used an SPPT to evaluate spatial similarity between different data sources. Hibdon, Telep, and Groff (2016), for example, used an SPPT to study the spatial concentration and stability of drug activity in Seattle in police and emergency medical services data. Tompson, Johnson, Ashby, Perkins, and Edwards (2014) used an SPPT to study the spatial accuracy of open crime data compared with police-recorded crime data. In this study, the SPPT is used to examine spatial similarity between the crimes registered in the DNA dataset and the police dataset over the 27 judicial districts (Chapter 7). For every judicial district, a pairwise comparison is made of the percentages of crimes in the police dataset (i.e., the base dataset), and the DNA dataset (i.e., the test dataset). Therefore, the SPPT allows a validation of the representativeness of the unsolved crimes stored in the NGDB.

A fictional example illustrates the SPPT: 35% of all crime recorded in the police dataset was committed in Antwerp. According to the DNA dataset, 12% of all crime was committed in Antwerp. The SPPT compares these two relative frequencies with each other and indicates by means of a local S-index whether the difference between the two frequencies is significant. A local S-index is calculated for every judicial district or spatial unit. These local S-indices indicate whether the point count in a specific judicial district is lower in the base dataset (S-index = +1), higher (-1) or similar (0) compared to the test dataset. A global S-index provides insight into the similarity across all spatial units. The global S-index represents the percentage of judicial districts that have a similar spatial point pattern (local S-index = 0).

4.5.3. Social network analysis[9]

In their seminal work, Wasserman and Faust (1994, p. 20) defined a social network as "*a finite set or sets of actors and the relation or relations defined on them*". These actors can be very diverse: individuals like school children or members of an activist group, but also companies, countries, websites, etc. The relations between these actors can be any type of tie that connects these actors: friendship, kinship or business transactions, for example. Translated to criminological research, the actors can be offenders linked by a co-offending relationship. In social network analysis (SNA) the focus is rather on the study of the relationships between the actors than on the characteristics of the actors themselves (e.g., gender or age).

Social network analysis can be used in two distinct ways: a formal approach or a structural approach (Rossy & Morselli, 2018; Vlaemynck, 2014). In the formal approach, the network is the dependent variable. The aim of the analysis is to describe the network structure and the position of the nodes in the network. Most network researchers apply this formal approach by describing the structure of various criminal groups, such as street gangs (McGloin, 2005) or the Hell's Angels motorcycle club (Morselli, 2009), for example. A formal social network analysis can also be applied to assess the characteristics that distinguish criminal networks from non-criminal networks (Papachristos, 2011).

In the structural approach, the network is the independent variable. The starting point of this approach is the interdependencies among actors, which have a significant impact on the behaviour of the actors. As such, the aim is to explain how groups relate to crime and criminal behaviour (Papachristos, 2011). The occurrence of crime can be explained by the characteristics of the individual actors (e.g., gender or age), and also by the relationships between the actors (Rossy & Morselli, 2018, p. 192). Or, as Papachristos (2011, p. 107) stated: "*Gender might exert a direct effect on crime simply because women commit less crime than men, but the network approach would also be interested in how gender influences patterns of social relationships that, in turn, are responsible for the gender effect in crime*". The impact of groups on crime is also reflected in a number of criminological theories, such as the differential association theory by Sutherland (Sutherland & Cressey, 1960) and the self-control theory by Gottfredson and Hirschi (1990). In a nutshell, the differential association theory states that criminal attitudes, behaviours, values and motives are learned from associations with criminal others. According to the self-control theory, self-

---

[9] For a general overview of SNA, see Wasserman and Faust (1994) and Scott (2013). Bouchard and Malm (2016) and McGloin and Kirk (2010) described the use of network approaches in criminology.

control restrains people from committing crimes, and is learned from others.

In this dissertation, the formal approach to social network analysis is applied, as the aim is not to explain crime. Social network analysis is used to study the changes crime networks undergo when DNA data are integrated with police-recorded crime data (Chapters 9 and 10).

4.5.4. Observations and interviews

Knowledge about forensic DNA and the role that DNA plays in police research was limited at the start of this study. Studying the available (foreign) scientific literature, legislation or any other form of public information was not enough to fully understand the subject. For example, it soon became clear that the 'theory' that is publicly available is not always in line with what happens in 'practice' (for example, because of limited resources and time), or that there are differences between judicial districts regarding DNA retention policies. For these reasons, various observations and interviews were carried out.

A (descriptive) observation is extremely suitable for collecting information about a topic for which little knowledge is available or for collecting information that is not available in any other form (Mason, 2003; Zaitch, Mortelmans, & Decorte, 2009). In addition, an observation allows a *multi-method strategy* to be used: the observer can conduct interviews during the observations. These interviews can be scheduled but also spontaneous (Mason, 2003). A spontaneous interview allows researchers to respond to what they observe and offers the opportunity to immediately respond to issues that are unclear to them. Interviews can also be applied as a separate technique to gain insight into the knowledge, experiences and interpretations of people. As with the observation, by conducting an interview, information can be obtained that is difficult or impossible to obtain via another technique, such as surveys.

Various actors operating within the domain of forensic DNA were identified and contacted, and were asked to provide an explanation of their organization and operation, and the possibilities and limitations they experience during the performance of their duties, through an interview and/or observation. Five different DNA actors were selected who were involved in the process from the inclusion of DNA traces to the registration of DNA profiles in the NGDB (**Figure 4**).

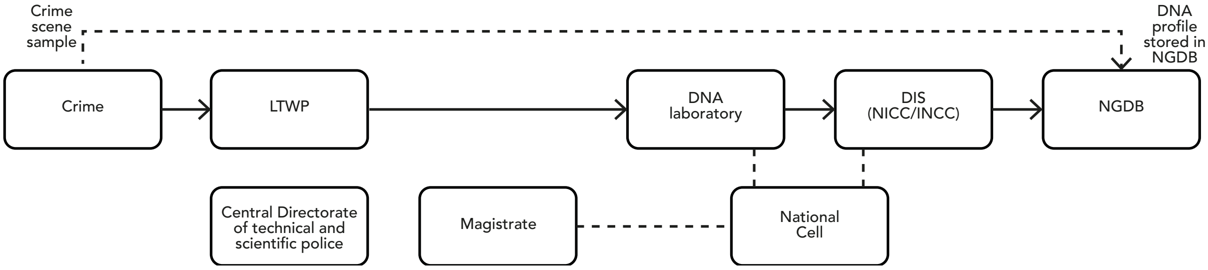


Figure 4: Actors involved in the forensic DNA process

The Central Directorate of the Technical and Scientific Police (DJT), the National DNA Cell, and the DIS service of the NICC/INCC were questioned using *unstructured interviews* (between October 2014 and January 2015).[10] Unstructured interviews are more informal than (semi-)structured interviews. The interviewers had no predetermined list of questions.[11] Each interview started with the interviewee introducing their organization. The interviewee was given the opportunity to talk freely about what they thought was important about their organization (Saunders, Lewis, & Thornhill, 2009). The interviewers could ask additional questions during and after the presentation.

Two *participant observations* were conducted: one at the NICC/INCC's DNA laboratory (27, 28 and 30 October 2014), which is one of the seven accredited forensic DNA laboratory in Belgium,[12] and one at the Laboratory of the Technical and Scientific Police (LTWP) of the judicial district of Ghent (2–15 March 2015). A participant observer can adopt four different roles during an observation (**Figure 5**). These roles differ depending on whether or not the identity of the researcher is concealed and whether or not the researcher takes part in the activities of the setting that is being observed (Gold, 1958; Saunders et al., 2009; Zaitch et al., 2009). During the observations at the DNA laboratory and the LTWP Ghent, the researcher took the role of *observer as participant*. The true identity and purpose of the researcher was known to all concerned. However, the role of the researcher was limited to observing the activities of the DNA laboratory and the LTWP Ghent, without taking part in the same way that the staff of these organizations did. In other words, the researcher did not carry out any crime scene investigation, or any other analysis in the DNA laboratory, for example.

---

[10] The DJT provide support to the LTWP, the magistrates and the police. For a description of the National Cell and the DIS service, see Chapter 6.

[11] As these actors were relevant to all the partners of the Be-Gen project, the researchers of the VUB and the NICC jointly conducted the interviews.

[12] There are seven DNA laboratories responsible for analysing reference profiles of suspects and crime scene profiles (traces) in Belgium. A German laboratory is responsible for analysing DNA from convicts and internees (Belgisch Staatsblad, 2017b).

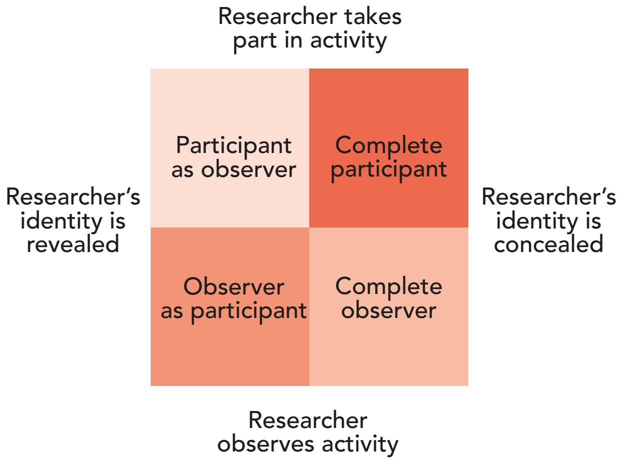


Figure 5: Typology of participant observation researcher roles (Saunders et al., 2009, p. 293)

The observations were limited to *descriptive observations* (as opposed to focused observations and selective observations, see Spradley, 2016). The short-term observations were no more than a first acquaintance with the DNA laboratory and the LTWP Ghent. It was important to understand their operation and organization as comprehensively as possible (Mortelmans, 2007; Spradley, 2016).

During the observation at the NICC/INCC laboratory, the researcher was able to observe the activities of the forensic experts: from investigating forensic physical evidence, to determining the DNA profile and its representation in a report. The observation at the LTWP Ghent offered insights into the work of the police during the following phases: (1) planning, organization and coordination of the work at the crime scene; (2) preservation of the crime scene and the evidence; (3) documentation of the crime scene, the evidence and activities at the scene; (4) recognition, recovery and preservation of physical evidence; and (5) transportation, storage and submission to the DNA lab (UNODC, 2009).

Although these observations are not recorded in this dissertation, these observations and interviews are considered necessary for answering the research questions. Both the interviews and observations broadened the knowledge of the researcher, and offered the context and background that are necessary for a thorough interpretation of the findings of this dissertation.

5. Structure of this report

This dissertation has four parts: an introduction, two empirical parts addressing the four research questions, and a conclusion. **Part 1** (Chapters 1–5) has provided a description of the background of the study, the relevance of forensic intelligence, the resulting research objective and research questions, and the data and methods used in the dissertation.

The first two research questions, regarding the selectivity of the DNA data, are answered in **Part 2** (Chapters 6–7). ***Chapter 6*** starts with some background information on forensic DNA and DNA databases before explaining the legal framework and other mechanisms and actors that all have an impact on the content, and thus selectivity, of the NGDB. ***Chapter 7*** deals with the validity of forensic DNA data in the study of the (spatial) behaviour of unknown offenders, given the fact that DNA data are only a subset of the crimes committed by unknown offenders stored in police-recorded crime data. Therefore, a study is made of the concentration and spatial similarity of detected but unsolved crimes in police-recorded crime data (N = 181,483) and DNA data (N = 1,913) over 27 Belgian judicial districts for the four selected crime types. Spatial similarity is established for certain crime types in some districts. This offers opportunities for DNA data to be used to study unknown offenders' spatial offending behaviour.

Part 3 (Chapters 8–10) focuses on the advantages of using DNA databases to study and link unknown offenders and their crimes. The importance of a network approach in criminology is explained in ***Chapter 8***. The central question in ***Chapter 9*** is whether unsolved registered crimes have different features from solved registered crimes. Therefore, both unsolved crimes from the DNA database and solved crimes from the police-recorded crime data are integrated to study serial co-offending behaviour. The focus is on the changes the networks of crimes that are obtained from police-recorded crime data undergo when integrated with data from unknown offenders in the DNA database. The results show that an integrated dataset reveals more and larger networks of crimes with a larger spatiotemporal spread, compared to the police-recorded crime data only.

Chapter 10 studies the possible difference between unknown and known offenders by assessing the missing data problem in criminal network analysis. Missing data are pertinent to criminal networks due to the hidden nature of crime. If one offender (i.e., 'node') or link between two offenders (i.e., 'edge') is missing, the picture of the network may be distorted. Generally, researchers evaluate the impact of incomplete network data by extracting or adding nodes and/or edges from a known network. Statistics on this reduced or completed network are then compared with the statistics from the true network. In this chapter, an important advancement over previous missing data studies is made by integrating police-recorded crime data with DNA data. As a result, the network only containing known offenders from the police-recorded crime data is extended with unknown offenders from the DNA database. Statistics from the integrated dataset (i.e., ‘true network’) are compared with the statistics from the police-recorded crime data, which only contain known offenders (i.e., ‘reduced network’).

The **fourth and final part** (Chapters 11–17) presents the conclusions of the dissertation. The answers to each of the four research questions are summarized, and a broader discussion on the results is provided. Policy implications and future research opportunities are formulated, and the specific strengths and limitations of the dissertation are outlined.

2

PART 2:

THE SELECTIVITY OF DNA DATABASES

"Nonetheless, it is a fact that forensic data are collected at a minority of crime scenes, as shown in Williams (2004). To what extent could an accurate description of a population be based on such a small sample? Of themselves, small samples are not much of a problem if the sample is representative [...]."
(Tilley & Townsley, 2009, p. 375)

Part 2 reviews the impact of the selectivity of DNA databases on scientific research. The first chapter examines the actors and factors that may influence the composition of the NGDB. This chapter has a theoretical approach. These findings are empirically tested in the second chapter, and the validity of the NGDB as a source for scientific research is evaluated on the basis of a case study.

6. Forensic DNA and DNA databases

This chapter briefly explains forensic DNA and DNA databases. It gives an overview of the elements that make up DNA databases, and in particular the NGDB, both in terms of the nature and the number of the stored profiles. The aim is to answer the following research question:

Q 1: Which actors and factors have an impact on the selectivity of DNA databases?

6.1. Forensic DNA[13]

> *"Nul ne peut agir avec l'intensité qui suppose l'action criminelle sans laisser des marques multiples dans son passage, tantôt le malfaiteur a laisse sur les lieux des marques de son activité, tantôt par une action inverse, il a emporté sur son corps ou sur ses vêtements les indices de son séjour ou de son geste."* (Locard, 1923)

Edmond Locard's famous exchange principle dates from long before Alec Jeffreys discovered in 1984 that each individual has a unique DNA pattern, a *genetic fingerprint* (Jeffreys, Wilson, & Thein, 1985). Nevertheless, DNA appears to be a textbook example for applying Locard's principle, as everywhere people go they unknowingly and unintentionally shed hair, skin cells, saliva, and thus DNA. DNA (deoxyribonucleic acid) comprises the hereditary material present in human cells. DNA is arranged in two connected chains (i.e., a double helix) containing four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T) (**Figure 6**). The sequence of the bases determines how the human body is built and maintained. Except for identical twins, the genetic material is unique to each individual and remains unchanged throughout their life. Only a small percentage of DNA is responsible for hereditary properties (i.e., the genotype) such as hair and eye colour or disease patterns. The externally observable characteristics (e.g., black hair, blue eyes) corresponding to this genotype are called the phenotype.

In order to protect the privacy of those involved, only non-coding DNA or 'junk DNA' can be examined as part of a criminal investigation in Belgium. Non-coding DNA, according to current knowledge, does not determine hereditary properties. The only external feature that may be identified by DNA analysis is gender. The European Union recommends that member states use the European Standard Set (ESS) in forensic DNA analysis. This consensus makes it easy to compare DNA profiles at both national and international levels. The ESS comprises 12 specific particles of non-coding DNA that have to be tested,

---

[13] For more information on forensic genetics, see Meulenbroek (2009) (in Dutch) and Goodwin, Linacre, and Hadi (2007).

called loci.[14] An analysis of these loci results in a combination of numbers and letters, an alphanumeric code unique to a particular individual (Goodwin et al., 2007; Meulenbroek, 2009; Stappers et al., 2016).

If two genetic profiles are compared and the profiles are different, then it is certain that the profiles originate from different people. However, when two samples result in the same profile it is not absolutely certain that they originate from the same person. As more loci are examined, the probability that two individuals have the same DNA profile decreases rapidly (and also the number of false positive matches (van der Beek, 2011)). The probability is statistically calculated using population studies. The quality of the sample will influence the interpretation. For example, a match between a reference profile and a forensically simple profile (i.e., coming from only one offender) will yield a higher probability than when a match is found between a reference profile and a forensically mixed profile (i.e., coming from at least two individuals) (NICC/INCC, 2014). When a match is found between a reference profile and a simple profile, it is generally at least one billion times more likely that the profiles come from the same individual, than that the trace is from an unknown person randomly taken from the European population unrelated to the donor of the reference profile (Jeuniaux et al., 2015).

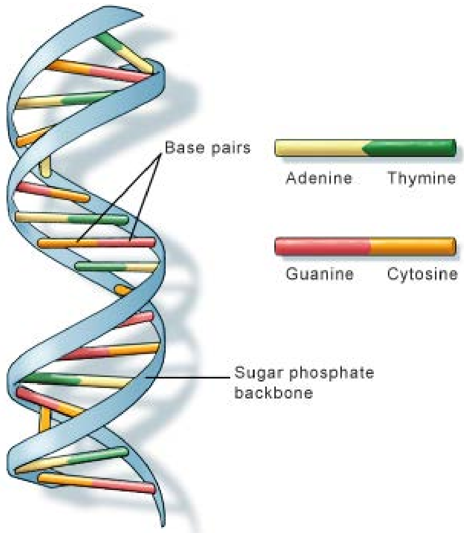


Figure 6: Structure of DNA (Genetics Home Reference, 2017)

---

[14] Council Resolution C296/1 of 30 November 2009 on the exchange of DNA analysis results.

6.2. DNA databases

DNA profiling techniques were first developed in the mid-1980s (Lynch, 2003), and in 1992 the European Council published a non-binding recommendation on the use of DNA analysis within the framework of the criminal justice system (Council of Europe, 1992). In 1995 the United Kingdom National DNA Database (NDNAD) was established, becoming the first national forensic DNA database in the world (Martin, Schmitter, & Schneider, 2001; Wallace et al., 2014). Since then, many countries have created their own national DNA database (Voultsos et al., 2011).[15] According to the DNA Interpol Survey published at the end of 2012, however, only 63 countries had a national database for genetic profiles, whereas 135 were already using DNA for judicial investigation (Hallauer, 2014). National legislation regarding the collection, storage and use of DNA samples and DNA profiles in criminal cases varies considerably (Bramley, 2009; F. Santos, Machado, & Silva, 2013). For example, the NDNAD is one of the largest and most comprehensive DNA database, while the legislation of other countries such as Belgium, France or the Netherlands have more restrictive effects on their DNA databases (F. Santos et al., 2013).

The establishment of DNA databases means that, in addition to collecting DNA, the profiles are also stored for a specific time.[16] The reasons for collecting DNA profiles may be different from the reasons why they are stored. GeneWatch UK[17] explains this as follows: "*The purpose of entering an individual's DNA profile on the Database is to see if they are a potential suspect for a past crime. ... The purpose of retaining an individual's DNA profile on a database is to treat them as suspects for any future crime*" (GeneWatch UK, 2006, pp. 2-3). When creating and expanding a DNA database, it is therefore important to always weigh up the rights and privacy of the individual against the need for crime prevention, crime detection and public protection (GeneWatch UK, 2006; Home Office, 2009; Kazemian et al., 2011; McCartney, 2004; Staley, 2005; Tseloni & Pease, 2011). In doing so, one should consider "*the presumption of innocence; data protection principles; ethical issues; human rights arguments including the right to privacy, as well as concerns about overbearing surveillance powers and the 'database state'* " (McCartney, 2012, p. 257). Although the debate about the storage and retention regime mainly relates to reference profiles, similar questions can be asked for all types of profiles. Most DNA databases only store forensic profiles and reference profiles. However, other types of profiles may be included for different reasons. The DNA of those who are directly or indirectly involved in the detection, analysis and processing of traces found at crime scenes may be

---

[15] For an overview of the origins of the different national DNA databases, see the publications of Martin (Martin, 2004; Martin et al., 2001; Schneider & Martin, 2001).

[16] For further information and discussion on the debate about whose profiles (i.e., which specific crimes) should be stored in a DNA database and for how long, see for example Kazemian et al. (2011), Tseloni and Pease (2011), M. E. Smith (2006) and Chapter 15 discussing future research possibilities.

[17] GeneWatch UK is a not-for-profit policy research and public interest group that monitors developments in genetic technologies. See: http://www.genewatch.org.

stored in order that they can be excluded as donor of a forensic profile. Profiles of missing persons or their blood relatives and unidentified bodies may be included in a (separate) DNA database in order to try to identify bodies and remains.

6.3. The Belgian National Genetic Database

Before the Belgian National Genetic Database (NGDB) can be evaluated as a data source, it is important to understand the relevant legal framework, how it operates and what it contains (see for details Jeuniaux et al., 2015; Renard, Duboccage, Jeuniaux, & Vanvooren, 2013).

6.3.1. Legal framework

This section elucidates both national and international legislation that influence the creation, operation and content of the NGDB.

6.3.1.1. National legislation

Although DNA was used in criminal cases in Belgium during the 1990s, DNA testing was only legally regulated on 22 March 1999 by the law on the identification procedure via DNA research in criminal cases. This law adds a new article to the Code of Criminal Procedure stating that comparative DNA research may only compare the DNA profiles of detected or acquired human cell material in order to directly or indirectly identify persons involved in a crime (Art. 44ter. § 1., Belgisch Staatsblad, 1999).

It is noteworthy in the framework of this comparative DNA research that the law established two DNA databases within the NICC/INCC: the ‘criminalistics’ database, and the ‘convicted offenders’ database (Belgisch Staatsblad, 1999, 2011a). Together they form the NGDB. Traces found when a crime is being investigated are stored in the criminalistics database. The convicted offenders database contains the DNA profiles of individuals convicted or interned for the offences defined by law (Belgisch Staatsblad, 2011b). Therefore, the database includes both identified profiles (i.e., profiles of offenders) and unidentified or unknown profiles (i.e., profiles of traces found at crime scenes).

In 2011[18] the legislation was comprehensively amended to improve and simplify the use of DNA in criminal cases. In addition, the conditions under which DNA profiles can be stored and compared were extended. First, the list of crimes for which an offender’s DNA profile can be stored in the database was supplemented with a number of new crimes, such as terrorist crimes and genocide. Second, this list was expanded to include crimes that regularly lead to links with the profiles in the criminalistics database. Examples include crimes related to criminal organizations or gang formation. Following these

---

[18] This new DNA law came into effect on 1 January 2014 (“Omzendbrief DNA: identificatieprocedure via DNA-onderzoek in stafzaken,” 2013).

changes, and under certain circumstances,[19] the profiles of suspects could be drawn up and systematically forwarded to the NGDB for a one-off comparison. When a match is obtained, suspect profiles can also be stored in the criminalistics database (Belgisch Staatsblad, 2011b; Jeuniaux, Duboccage, Renard, Van Renterghem, & Vanvooren, 2016; Renard et al., 2013). The identity of the suspect or convict is stored by means of a DNA code number. The National Cell, established under the 2011 Act and part of the Federal Prosecutor's Office, assigns, manages and centralizes DNA code numbers. This unique code links a DNA profile to a suspect or convicted person without storing any information about the person's identity in the NGDB (Belgisch Staatsblad, 1999, 2017a).

A third database, for missing persons, was established under the Act of 21 December 2013 amending the Code of Criminal Procedure and the Act of 22 March 1999 concerning the identification procedure for DNA research in criminal cases (Belgisch Staatsblad, 2013b, 2017c). This database stores DNA profiles of the remains of unidentified bodies, DNA profiles of samples (e.g., teeth, blood) and/or from personal items (e.g., hairbrush) belonging to a missing person, and finally reference profiles of blood relatives of missing people. Previously, these traces were stored in the criminalistics database.

In 2017 a new law created a fourth database, the 'elimination' database (Belgisch Staatsblad, 2017d). Individuals who come into contact with crime scenes and the broader forensic tracing investigation, such as emergency services, the police or public prosecutors and lab technicians, can unknowingly leave their DNA at the scene and thus contaminate DNA traces. Storing these DNA profiles allows the authorities to identify possible contamination at an early stage and prevent the investigation from following a false lead. This dissertation could not take into account possible contamination that was brought to light by the introduction of the elimination database, as the data extraction was carried out before it was introduced.

6.3.1.2. Transnational context and the Prüm Convention

In the last decade, not only have the number of national DNA databases increased but the exchange of DNA data and the cooperation between different countries has also been strengthened. In 2005 seven European countries[20] signed the Prüm Convention on the improvement of cross-border cooperation, in particular to combat terrorism, cross-border crime and illegal migration. The parties stated that they would create and manage national DNA databases for the purpose of detecting criminal offences (Council of the European Union, 2005, Chapter 2, Article 2). This network of national DNA databases should make it possible (and easier) to identify offenders who are responsible for multiple crimes in different countries (Martin et al., 2001). The goal is to link unsolved crimes in one country with (un)solved

---

[19] Art. 44quinquies SV and Art. 90undecies SV

[20] Belgium, Germany, Spain, France, Luxembourg, the Netherlands and Austria.

crimes in another country (Prainsack & Toom, 2013).

In 2008 the Convention was converted into EU legislation (Council of the European Union, 2008a, 2008b) stating that all EU members must create a database for DNA, fingerprints and traffic data. These databases must be searchable by other EU members via automated searches on a 'hit/no hit' basis – meaning that anonymous data are compared, and the relevant data about the case and the person can only be exchanged between the countries involved if there is a match (van der Beek, 2011; Voultsos et al., 2011). Therefore, the Prüm Treaty does not create a central European DNA database, but works with national DNA databases and a national contact point per EU member. The NICC/INCC is the national contact point for Belgium (Duboccage, 2008).

In addition to the 28 EU member countries, Iceland, Liechtenstein, Norway and Switzerland were also admitted to the convention. In 2014 only 20 countries complied with the convention. Belgium integrated the implementation of the Prüm Treaty into the DNA law of 7 November 2011. By June 2018, Belgium was automatically exchanging DNA data daily with 10 countries: the Netherlands, France, Germany, Luxembourg, Sweden, Austria, the Czech Republic, Romania, Estonia and Spain. This will soon be extended to include Bulgaria, Lithuania and Poland. The ultimate goal is the automatic exchange of DNA data with all Prüm member states.

6.3.2. Selection mechanisms and actors

Storing DNA profiles in a database is based on the assumption that offenders will commit other crimes in the future, so that links can be made between offenders / suspects and traces, between different traces or between different offenders / suspects, eventually leading to identification. Two central questions arise: (1) Which profiles are included in the database? In other words, which (offenders of which) crimes are included in the database? (2) How long are the profiles stored in the database?

The content of the database at a given moment, in terms of type and number of profiles, is thus determined by, on the one hand, the profiles that are stored in the database and, on the other hand, the profiles that are removed from the database (**Figure 7**). Moreover, every actor within the criminal justice chain has an impact on the content of the database, whether consciously or not.

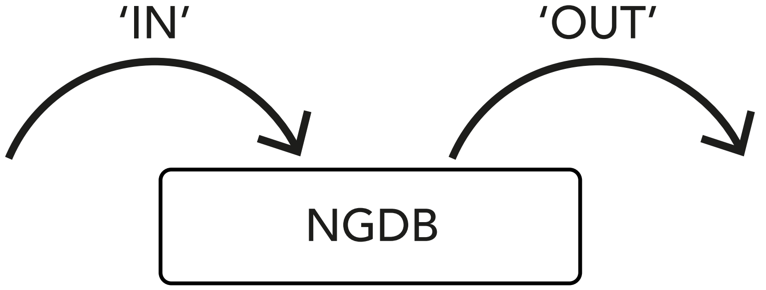


Figure 7: Schematic representation of the contents of the NGDB

Profiles stored in the criminalistics database

The criminalistics database[21] contains (unidentified) DNA profiles collected at crime scenes. In contrast to the convicted offenders database, the legislator has not drawn up an exhaustive list of crimes that can be included in the criminalistics database. In principle, every crime type is eligible. In practice, however, it includes crimes that have a specific crime scene and where it is likely that there has been physical contact between the offender and the victim, or with the crime scene, making it possible that the offender has left DNA behind.

A number of actors are significant when storing DNA profiles of traces in the criminalistics database. First of all, there is the **offender**. Not all offenders leave their DNA, for example in the form of skin cells, hair or saliva, at the scene of a crime (Develtere, 2014). Offenders are often aware of the traces they may leave behind during their offence. This 'forensic awareness' ensures, for example, that offenders wear gloves to limit DNA traces (Beauregard & Bouchard, 2010).

The offence must, of course, be reported to, or discovered by, the **police**. In view of the large 'dark number' of unreported crimes, this is perhaps the most important factor in determining which profiles are stored in this database (Biderman & Reiss, 1967). Moreover, the Laboratory of the Technical and Scientific Police (LTWP) or another police service collect these traces not simply because a crime is known about and is likely to contain DNA traces. Four levels of knowledge play a role in the decision of the police to visit a crime scene and to collect traces: the strategic, the criminal, the immediate and the physical environment (Ribaux, Baylon, Lock, et al., 2010). The strategic environment is shaped by the priorities set out in the Belgian National Security Plan, for example, as well as (limited) budgets and resources. The criminal environment is constituted by knowledge about previous similar cases or criminal phenomena such as organized crime (i.e., forensic intelligence). This knowledge is important to be able to detect

---

[21] Only the criminalistics database is discussed, as it is directly relevant to this research. For further information about the missing persons and elimination databases see Renard et al. (2018).

all possible (relevant) traces. Knowledge about previous, similar crimes, for example, can increase the chance of finding traces, and the importance of the case can also be higher, given that the offender is possibly a multiple offender. The immediate environment will also determine whether traces are collected. Where, when and how exactly did the crime happen? The modus operandi and the seriousness of the crime are important when assessing the immediate surroundings. For example, a first study by the medical examiner on the crime scene could provide information about whether or not the victim died a natural death. In the event of a natural death, the crime scene investigation (and thus, also, the safeguarding of possible DNA carriers) will be much more limited. Crime scene investigators are often most familiar with the physical environment of the crime scene. For example, public places such as bars are less suitable for collecting DNA traces because of the high number of DNA traces that will be present and the risk of contamination (Ribaux, Baylon, Lock, et al., 2010). Of course, in addition to these four levels of knowledge, the experience and training of police officers will also influence the number and type of traces collected at crimes scenes.

Once the DNA traces have been collected, the authorized **magistrate** must mandate a certified DNA laboratory to examine them. The traces that have been collected are not all analysed by default; the magistrate often makes a selection. In addition to operational motives such as whether or not the police advise or request that they be tested (using model 14[22]), budgetary restrictions are often the decisive factor for whether certain DNA samples are analysed. DNA analyses are expensive, which means that some screening takes place. At the end of 2015 both the rates and the description of DNA research in criminal cases were determined in accordance with regulations defined by Royal Decree (Belgisch Staatsblad, 2015). At the same time, the cost of DNA analysis has considerably reduced. On average, the analysis of a trace profile now costs 273 euro.[23] Before the 2015 Royal decree it was twice as much. Not only has the cost fallen in recent years, it is also striking that magistrates often do not know how much analysis costs, or greatly overestimate the cost (Stappers, 2018).

The process that recognized **DNA laboratories** use has three phases. The first phase investigates whether the pieces of evidence will contain sufficient DNA traces to be useful for carrying out DNA analysis. So far, it is not known whether useful DNA traces are present on the pieces of evidence. Assuming they are, the second phase involves extracting and quantifying the DNA and the creation of the DNA profile. The third and final phase consists of interpreting the DNA profile that has been obtained, and drawing up an

---

[22] Standard table for the selection and arrangement of trace carriers for possible DNA research, added to the Royal Decree of 2013 (COL21/2013, revised version 8 June 2017). See **Appendix 1.**

[23] This price includes: examination of pieces of evidence and taking samples: 30 euro (Art. 4); extraction and quantification of DNA (Art. 6): 49 euro; creating the genetic profile (Art. 7): 194 euro. The price for drawing up a genetic profile of a reference sample has been 60 euros since the 2015 Royal Decree, which is also a decrease compared to the old rate of 430 euros.

expert report.[24]

The DNA profiles are then forwarded to the NICC's **DNA Index System** (DIS) service. DIS is the final step before the profiles are stored in the database. The only samples stored in this database are forensically simple profiles originating from one person, or forensically mixed profiles of a maximum of two people. DIS uses the Combined DNA Index System (CODIS) developed by the US FBI to manage, store and compare DNA profiles.

The number and quality of the DNA profiles obtained can be affected at each stage by human error (Ribaux & Talbot Wright, 2014). **Figure 8** summarizes this process schematically.

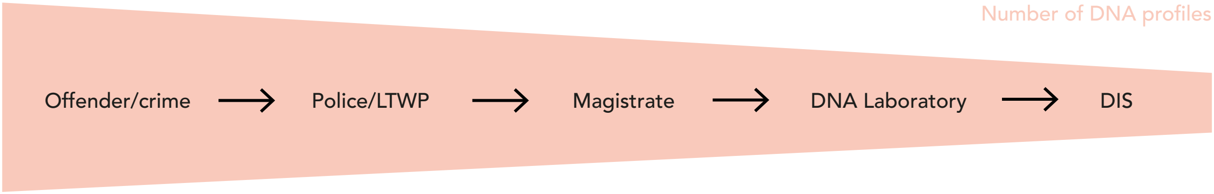


Figure 8: Selection mechanisms and actors involved, from collection to storage of forensic DNA in the NGDB

Profiles removed from the database

For **practical and financial** reasons it is likely that some DNA profiles will need to be removed from the database. The management of a database in which profiles are added and never deleted would be a heavy burden. In addition, an unlimited storage period would not seem to comply with the aforementioned **human rights. Substantive arguments** also play a role in limiting the retention period. For example, Walsh, Curran, and Buckleton (2010) state that solving old offences has a lower priority and is often more difficult than solving more recent offences. Moreover, offenders will no longer be criminally active in the course of time, whether that is because they stop committing crime, die or are taken prisoner; keeping track of their profiles will then no longer be useful. Another reason can be found in Belgian law, which limits the term for which data can be stored to thirty years for both databases (Belgisch Staatsblad, 2011b). This period corresponds to the statute of limitation for each crime type (the maximum period of time during which a conviction can be obtained), after which it is no longer appropriate to keep the data.

6.3.3. Content NGDB

At the end of 2017 the NGDB stored about 100,000 profiles, crime scene profiles and reference profiles

---

[24] For more information on DNA analysis, see Butler (2010).

(i.e., convicted offenders and suspects) in total (**Table 7**). **Figure 9** shows that the annual number of stored profiles has steadily increased since the NGDB database was implemented in 2002.

Table 7: Number and type of DNA profiles in the NGDB per year

Year	Crime scene profiles	Convicted offenders	Suspects	Total
< 2002	1,020			1,020
2002	1,938	4		1,942
2003	3,563	448		4,011
2004	5,398	2,001		7,399
2005	7,752	4,536		12,288
2006	9,945	8,570		18,515
2007	12,671	1,4442		27,113
2008	15,933	15,596		31,529
2009	19,032	18,003		37,035
2010	22,493	20,774		43,267
2011	26,016	22,861		48,877
2012	30,480	25,333		55,813
2013	34,725	27,710		62,435
2014	39,339	30,724	636	70,699
2015	45,534	34,695	1,330	81,559
2016	47,374	39,538	1,975	88,887
2017	50,788	44,698	2,553	98,039

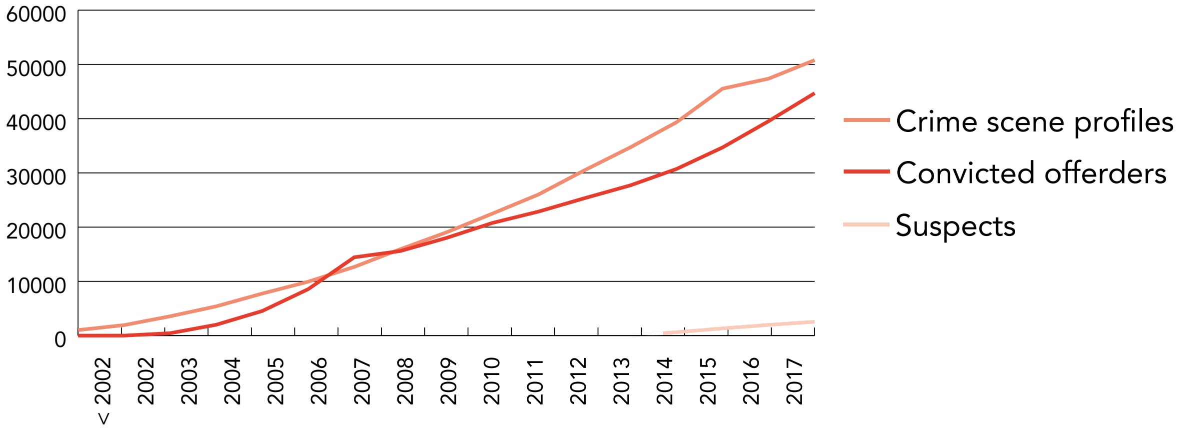


Figure 9: Number and type of DNA profiles in the NGDB per year

6.4. Conclusion

Before a forensic DNA profile is stored in the NGDB, it has already gone through several crucial phases. At each of these phases there is a chance of dropout. As a result, the NGDB, like any other DNA database, is only a selection of all crimes registered in the police database. The impact of this selection is also closely related to the crime type – for example, DNA profiles from contact crimes, where the offender makes physical contact with the victim or with an object at the crime scene, are more likely to be present as there is a much better chance that DNA profiles will be found.

7. The validity of DNA databases in studying the spatial behaviour of unknown offenders

The previous chapter explored why the NGDB only contains a minority of all the crimes stored in the police database. It can thus be regarded as a sample of the police-recorded crime data. But is it a representative sample? And, more specifically, can the NGDB be used to study unknown offenders and their behaviour? (The validity of police-recorded crime data is a discussion in itself, but falls outside the scope of this dissertation.) The central research question in this chapter is therefore formulated as follows:

Q 2: Are DNA databases less valid than police-recorded crime databases for the study of unsolved crimes?

The criminographic data available in the NGDB are limited. This also applies to the unsolved crimes registered in the police database. The place (i.e., judicial district) and time (i.e., year) of unsolved crimes are the only data registered in both databases. Therefore, this is the only information that can be used to evaluate the representativity of the DNA data compared to the police-recorded crime data. The present study, which is a further elaboration of De Moor et al. (2018a), uses only the geographical information. A comparison was made of the spatial concentration and distribution of unsolved crimes stored in the two separate databases.

7.1. Introduction

Recently, scholars have argued in favour of using forensic DNA databases to study offenders' spatial behaviour. For example, as mentioned in the introduction, Wiles & Costello (Costello & Wiles, 2001; Wiles & Costello, 2000) used DNA database hits in addition to geocoded police-recorded crime data and offender interviews in their seminal study on the mobility of high-volume offenders in and around Sheffield. They used DNA data to study the travel patterns of offenders, irrespective of whether they were known to the police or not. All three data sources revealed that most offender movements are relatively short. Lammers (Lammers, 2014a; Lammers & Bernasco, 2013) used Dutch DNA data to study the influence of the spatial dispersion of serial offenders' crime location on their probability of arrest. In order to do so, Lammers compared the spatial behaviour of known and unknown offenders stored in the DNA database. She concluded that offenders with a greater geographical offending range, measured as the number of police regions in which the offenders committed their crimes, were less likely to be arrested and therefore had a lower probability of featuring in police-recorded crime data (Lammers & Bernasco, 2013). However, the average distance between crime locations was unrelated to the probability of arrest (Lammers, 2014a).

Despite the successes of applying DNA data in criminological research and the potential to study unknown offenders, DNA data remain subject to selection biases. Discussing the potential problems of selectivity in DNA data, Tilley and Townsley (2009, p. 375) question whether accurate descriptions of populations can be based on small samples, but add that “of themselves, small samples are not much of a problem if the sample is representative”. Similar concerns were raised by Lammers (2013) and Wiles and Costello (2000). By comparing crime levels across the police and DNA data, Wiles and Costello (2000, pp. 25-26) concluded that DNA data are representative of the overall crime level in England and Wales. They evaluated the representativeness of the DNA database against all registered crimes, solved or not. No scholars, however, have specifically evaluated the representativeness of DNA data in unsolved crimes.

The study presented in this chapter was designed to address this lack of knowledge and evaluate the validity of DNA data for criminological research on unknown offenders and their crimes. The spatial distribution of unsolved crimes stored in the police-recorded crime data is compared with the spatial distribution of unsolved crimes stored in the DNA data. As such, the concentration and spatial similarity of the two sets of data is studied. Three important advancements over prior work are introduced. First, this is the first study to explicitly examine the validity of DNA data in the study of the behaviour of unknown offenders. Second, this study uses and compares two different data sources (i.e., DNA data and police-recorded crime data) to study the same phenomenon: unsolved registered crimes. In prior research, different databases have been used to complement each other: each data source contained information on unique crimes – police-recorded crime data have been used to study (the offending behaviour of) known offenders and DNA data or self-report data were used to study (the offending behaviour of) unknown offenders. Third, the present study also contributes to criminological research based on DNA data, which remains limited.

7.2. Data

7.2.1. Study area and spatial unit of analysis

The spatial units of analysis are Belgium’s 27 judicial districts, which cover on average 1,136 square kilometres (min. = 511.98 km², max. = 2,507.32 km², S.D. = 410.58 km²). According to the NUTS classification (nomenclature of territorial units for statistics), a Belgian judicial district is smaller than a NUTS 2 region but larger than a NUTS 3 region. Belgium counts 11 NUTS 2 regions (10 provinces and Brussels-Capital Region) and 44 NUTS 3 regions (arrondissements).

Judicial districts are meaningful units of analysis when comparing data from the police dataset and the DNA dataset. Judicial districts can set their own priorities in the types of crimes that the police have to respond to, and in the DNA retention policy, and these factors will influence the spatial pattern of registered crimes in each judicial district. Although using the judicial district level as the spatial unit of analysis can obscure the concentration and also the variability that would be found in smaller spatial units (Andresen & Malleson, 2011; Hibdon et al., 2016; Ratcliffe, 2010), this geographic scale is suitable for our purposes as it is not our intention to make statements about crime levels and dispersion at different geographical levels.

7.2.2. Police-recorded crime data and DNA data

As the focus of this research is not on the spatial stability/variability of crime over time (see for example Andresen et al., 2017) but on the spatial similarity between the Belgian General Police Database and the Belgian National Genetic Database, only data from one calendar year was analysed. The year 2014 was chosen for no substantive reason. The analyses concerning the other calendar years are included in **Appendix 2.** All other years yielded comparable results.

Crimes of aggravated burglary, violent theft, lethal violence and sexual offences committed in 2014 were selected from both databases. Crimes with no identified offenders were selected from the Belgian General Police Database (i.e., the 'police dataset'). Only crimes without reference profiles were selected from the NGDB so that no known offenders were involved (i.e., the 'DNA dataset'). This approach ensured that only unsolved crimes were selected from both databases and resulted in a police dataset of 181,483 crimes and a DNA dataset of 1,913 crimes, all committed by unknown offenders.

7.3. Methods

The suitability of DNA data for studying the offending behaviour of unknown offenders on different levels is assessed. In addition to some descriptive statistics, the crime concentration and the similarity between the spatial points patterns in the police and DNA datasets is studied.

7.3.1. Lorenz curve and Gini coefficient

The Lorenz curve and the Gini coefficient are used to analyse the crime concentration in the judicial districts in each dataset. The Lorenz curve (see **Figure 10**) plots the relationship between the cumulative percentage of the judicial districts, and the cumulative percentage of the crimes committed in those districts. The diagonal line with slope 1 represents perfect equality in the distribution of crime over the different judicial districts. The closer the Lorenz curve leans to the diagonal, the more the crimes are evenly distributed over the districts. In accordance with Bernasco and Steenbeek (2017, p. 4), the judicial districts are ordered from the highest crime level to the lowest crime level in order to facilitate the interpretation of the curve. Each of the points of the curve can be interpreted as 'Y per cent of all crimes occur in X per cent of the judicial districts'. Generally, the Lorenz curve is a smooth line. As there are only a limited number of judicial districts, it was decided to represent the curve by the 27 corresponding judicial districts. That allows the above-mentioned interpretation to be further simplified to the exact number of districts.

The Gini coefficient (G) is calculated as the ratio of the area between the Lorenz curve (area A in **Figure 10**) and the line of perfect equality, and the area above the line of perfect equality (area A and B in **Figure 10**): G = A / (A + B). The Gini coefficient ranges between 0 and 1. In this study, a value of 0 indicates that every judicial district has the same crime level (i.e., complete equality) and a value of 1 indicates that all crime is concentrated in only one district (i.e., complete inequality) (Bernasco & Steenbeek, 2017; Eck et al., 2017).

An important drawback of the Lorenz curve and the Gini coefficient is that both measures may overestimate the level of crime concentration if there are fewer crimes than places (see Bernasco & Steenbeek, 2017). A (high) number of places with no crimes artificially increases the crime concentration. For example, five crimes committed in five different districts results in a higher Gini coefficient (and thus a higher concentration) when there are ten possible districts (G = 0.5) compared to only five possible districts (G = 0), although in both scenarios the five crimes are maximally spread. The generalized Gini coefficient (G’) proposed by Bernasco and Steenbeek (2017) is used to address this problem of overestimation of the concentration. The analyses were conducted in the statistical programming language R using the ‘lorenzgini’ package (Steenbeek & Bernasco, 2018). G’ will be smaller than G if there are fewer crimes than places but G’ and G will have the same value if there are more crimes than places. In this study, the number of judicial districts outweighs the number of crimes only for lethal violence so the possible range of concentration for lethal violence is smaller than the range of the other three crime types. As such, there will only be a difference between the original Gini coefficient and the generalized Gini coefficient for lethal violence.

The Gini coefficient does not provide insight into the similarity between spatial point patterns in the police dataset and the DNA dataset—for example, whether the districts with the highest crime rates in the police dataset also have the highest crime rates in the DNA dataset. Therefore, a spatial point pattern test was also conducted.

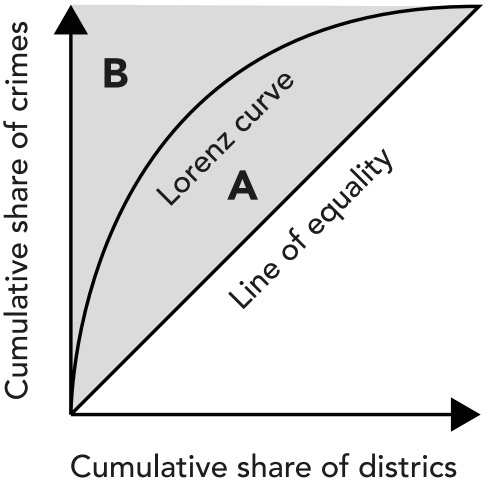


Figure 10: Graphical representation of the Lorenz curve and the Gini coefficient

7.3.2. Spatial point pattern test

This study uses the spatial point pattern test (SPPT) to examine the spatial similarity between the crimes registered in the DNA dataset and the police dataset over the 27 judicial districts. The SPPT is used to validate the representativeness of the unsolved crimes stored in the NGDB.

In the SPPT, a pairwise comparison is made of percentages[25] of crimes (i.e., 'points') in two datasets – a base dataset and a test dataset – within specified spatial units of analysis. In this dissertation, the police dataset is used as the base, and the DNA dataset as the test. Instead of comparing the exact percentages of crimes per judicial district, a Monte Carlo simulation is performed to create a confidence interval for each of the 27 judicial districts in the test dataset. Random samples (with replacement) of the test dataset are undertaken 200 times, selecting 85% of the entire test dataset (for details, see Andresen, 2016). When the percentage of crime in the base dataset falls within the confidence interval for the test dataset, then the specific judicial district has a similar proportion of crime in both datasets.

The SPPT generates two outputs: a local and a global similarity index. A local similarity index is computed for each judicial district separately. These local S-Indices indicate whether the point count in a specific judicial district is lower in the base dataset (local S-index = +1), higher (-1) or similar (0) compared to the test dataset. The output can be mapped in order to visualize the judicial district with significant difference. The SPPT also yields a global S-index indicating the overall similarity between the spatial

---

[25] As in much other research, the two datasets have different numbers of crimes or points: the police dataset includes far more crimes than the DNA dataset. Using the percentages instead of the exact number of points per judicial district makes it possible to compare datasets that are very different in size.

point patterns from the base dataset and the test dataset, ranging from 0 (no similarity) to 1 (perfect similarity). The mathematical form of the global S-index reads as follows:

$$S = \frac{\sum_{i=1}^{n} s_i}{n}$$

n represents the number of spatial units and s_i is equal to 1 if the spatial pattern of both the base and test dataset are considered similar for the spatial unit *i* and zero otherwise. It represents the percentage of judicial districts that have a similar spatial point pattern (local S-indices = 0). A global S-index of .80 (80%) or above is generally considered to be the threshold for spatial similarity (Andresen, 2009, 2016). In our research, this means that the local S-index must be 0 for at least 22 of the 27 judicial districts before the police and DNA dataset can be considered to be spatially similar at country level.

The SPPT is summarized in the following eight steps and **Figure 11**:

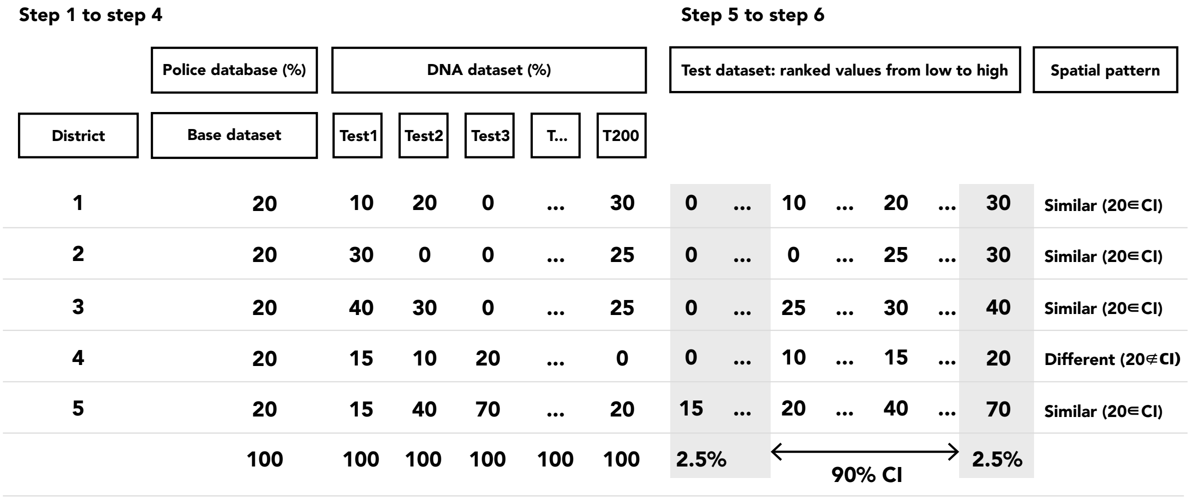
1) Identify a base dataset and count the number of 'points' inside each spatial unit.[26] Repeat this step for the test dataset.
2) Randomly sample (with replacement) 85%[27] of the points from the test dataset and count for each spatial unit the number of points.
3) Repeat step 2 200 times.[28]
4) For each of the 200 samples, calculate the percentage of points in each spatial unit. Rank for each of the spatial units these 200 values from lowest to highest.
5) Obtain a non-parametric 95% confidence interval for each spatial unit by removing the top and bottom 2.5% observations (i.e., the top and bottom five observations of the 200 observations in each spatial unit).
6) Compare within each spatial unit the percentage of points in the base dataset with the corresponding confidence interval generated from the test dataset. The spatial unit has a similar proportion of points in both datasets if the percentage of the base dataset falls within the confidence interval. If not, both datasets exhibit a significant different spatial pattern.
7) Map the output in order to visualize the spatial units with significant difference.
8) Calculate the global S-index to indicate the overall similarity between the spatial point patterns from the base dataset and the test dataset. (Andresen, 2009, p. 336; Andresen & Linning, 2012, p. 277; Andresen & Malleson, 2011, p. 64)

---

[26] The choice of which dataset is used as the base and which is used as the test is arbitrary (Andresen, 2009).

[27] Andresen (2009) relied on the research of Ratcliffe (2004a) to define the 85% boundary.

[28] Although research demonstrates that 50 or even 20 repeated samples would already provide good results, Andresen (2009) used 200 repeated random samples in his SPPT in order to be conservative and to provide convenient cut-off values for the confidence interval.



Step 1 to step 4

Step 5 to step 6

	Police database (%)	DNA dataset (%)					Test dataset: ranked values from low to high							Spatial pattern
District	Base dataset	Test1	Test2	Test3	T...	T200								
1	20	10	20	0	...	30	0	...	10	...	20	...	30	Similar (20∈CI)
2	20	30	0	0	...	25	0	...	0	...	25	...	30	Similar (20∈CI)
3	20	40	30	0	...	25	0	...	25	...	30	...	40	Similar (20∈CI)
4	20	15	10	20	...	0	0	...	10	...	15	...	20	Different (20∉CI)
5	20	15	40	70	...	20	15	...	20	...	40	...	70	Similar (20∈CI)
	100	100	100	100	100	100	2.5%		↔ 90% CI				2.5%	

Figure 11: Illustration of an SPPT

Judicial districts with no crimes for both the test and base dataset will also yield a local S-index of zero. This indicates a similar point pattern, although there is no point pattern at all. Therefore, if many spatial units contain no crimes, the global S-index is inflated. In that case, Andresen and Malleson (2011) suggest conducting a sensitivity analysis (see also Vandeviver & Steenbeek, 2017). This involves a SPPT with only non-zero spatial units: spatial units with at least one crime in either dataset. For all the different SPPTs conducted in this study, nearly all judicial districts have registered crimes in at least one of the datasets under consideration. Only one judicial district had zero registrations for lethal violence in both the police and the DNA dataset. Therefore, no such sensitivity analysis was needed in this study.

However, as will be demonstrate in the results section, many judicial districts in the DNA dataset (i.e., test dataset) have no or only a low number of crimes, which may severely influence the outcome of the SPPT (see Wheeler, Steenbeek, & Andresen, 2018). Judicial districts with no crime in the DNA dataset have a confidence interval from 0% to 0%. As a result, the SPPT will always indicate that the spatial pattern between the police data and DNA data is different for these districts, even if the proportion of crimes in the police dataset is only very small. Judicial districts with small crime proportions in the DNA dataset have larger confidence intervals, resulting in more similar spatial patterns (Wheeler et al., 2018). To meet these two limitations, Wheeler et al. (2018) suggest an SPPT based on the difference in proportions (hereafter 'adapted SPPT') in each spatial unit of analysis. In this study, the Chi-Square approach with Yates's correction with p-value adjustment for multiple comparisons present in the R package 'sppt' is used (Steenbeek, Vandeviver, Andresen, Malleson, & Wheeler, 2018).

7.4. Results

7.4.1. Descriptive statistics

Aggravated burglary is by far the most common crime type category in both our datasets: 85.90% (N = 155,894) in the police dataset and 81.02% (N = 1,550) in the DNA dataset. The second, much smaller, category is violent theft: 9.62% (N = 17,453) in the police dataset and 12.55% (N = 240) in the DNA dataset. The prevalence of sexual offences is 4.22% (N = 7,659) in the police dataset and 5.18% (N = 99) in the DNA dataset. Lethal violence is the smallest category in both the datasets: only 0.26% (N = 477) in the police dataset and 1.26% (N = 24) in the DNA dataset (see **Table 8**). Many judicial districts contain only few or even no crimes in the DNA dataset. Fifteen districts contain no lethal violence, eight districts contain no sexual offences, five contain no violent theft and two districts contain no aggravated burglary. For the police dataset, only one district contains no lethal violence crimes.

A chi-square goodness of fit test was calculated comparing the crime levels in the DNA dataset with those in the police dataset. Although the ranking of the four crime types in the DNA dataset is the same as in the police dataset, there are differences between the two datasets ($\chi^2 = 98.08$, df = 3, $p < 0.05$). The observed frequency of aggravated burglary in the DNA dataset (N = 1,550) is lower than would be expected (N = 1,643) based on the frequency in the police dataset ($z = -6.13$, $p < 0.05$). The reverse applies to the other three crime types: fewer crime counts of violent theft ($z = 4.35$, $p < 0.05$), lethal violence ($z = 8.47$, $p < 0.05$), sexual offences ($z = 2.08$, $p < 0.05$) were expected in the DNA dataset.

Table 8: Description of datasets

Crime type	Police dataset N	%	DNA dataset N	%
Violent theft	17,453	9.62	240	12.55
Aggravated burglary	155,894	85.90	1,550	81.02
Lethal violence	477	0.26	24	1.25
Sexual offences	7,659	4.22	99	5.18

7.4.2. Crime concentration

Figure 11 displays the Lorenz curve for both the police and the DNA datasets. The cumulative percentage of crimes, aggregated over the four crime types, is plotted against the cumulative percentage of judicial districts. When the two Lorenz curves do not cross, there is more inequality in the distribution of the crimes over the different districts in the upper curve. However, in this case, the Lorenz curves intersect at about 30% of the places, which makes the interpretation of the Lorenz curve ambiguous without the (generalized) Gini coefficient (Davies & Hoy, 1994).

Neither of the curves approximates the diagonal, which indicates that in both datasets crime is not evenly distributed over the 27 judicial districts. In contrast, crime is concentrated in a limited number of districts in both datasets. For the police dataset, about half of the judicial districts (N = 14) account for about 87% of all crimes, and about 50% of all crimes occur in fewer than four judicial districts. This corresponds to a (generalized) Gini coefficient for the police dataset of 0.56. The (generalized) Gini coefficient for the DNA dataset is comparable (0.54, p = 0.45). For the DNA dataset, about half of the judicial districts (N = 14) account for about 90% of all crimes, and about 50% of all crimes occur in fewer than five judicial districts.

The analysis illustrates that crime is concentrated in a limited number of districts, in both the police and the DNA dataset. Comparable results were obtained when studying the crime concentration of the four crime types separately. Violent theft and aggravated burglary have comparable high (generalized) Gini coefficients in the police data and the DNA data (**Table 9**). Lethal violence[29] ($p < 0.001$) and sexual offences ($p < 0.001$) are more concentrated in the DNA dataset than in the police dataset as the (generalized) Gini coefficients are significant higher for the DNA dataset than for the police dataset.

Table 9: Generalized Gini coefficients for different crime types

	Police dataset				**DNA dataset**		
	G'	**S.E. G'**	**Lower CI**	**Upper CI**	**G'**	**S.E. G'**	**p**
4 crime types	0.56	0.070801	0.4430009	0.6874887	0.54	0.05011042	0.447
Violent theft	0.70	0.07486759	0.5827504	0.8259792	0.62	0.05616732	0.287
Aggravated burglary	0.55	0.06931802	0.4371281	0.6799810	0.55	0.05262794	0.341
Lethal violence	0.54	0.06620676	0.4419252	0.6596354	0.67	0.1011423	< 0.001
Sexual offences	0.43	0.05244886	0.3444948	0.5447375	0.64	0.06190071	< 0.001

Figure 13 plots the districts according to their ranking in crime rates, aggregated over the four crime types, in the police and DNA dataset. All districts above the diagonal have a higher proportion in the DNA dataset than in the police dataset. The opposite is true for the districts below the line: these districts have a higher proportion in the police dataset than in the DNA dataset. The districts with the highest percentages of crimes are different for both datasets. Only two districts appear in the top five of both datasets. Antwerp (AN) has the highest crime rate in DNA dataset, and the second position in the police dataset. Brussels (BR) has the highest crime rate in police dataset, and the third place in the DNA dataset. Only two districts are located at the diagonal, having the same ranking in both datasets. Other districts seem to differ more in rank order between the two datasets. For example, the Hasselt district (HA) has a high ranking in the DNA dataset, but an average ranking in the police dataset. The Mons district (MO)

---

[29] As the number of this crime type outweighs the number of judicial districts, the original Gini coefficient differs from the generalized Gini coefficient. The Gini coefficient for lethal violence dropped from 0.70 to 0.67 to compensate for the overestimation of the level of crime concentration.

shows the opposite picture: a low ranking in the DNA dataset but a high ranking in the police dataset. Behind a difference in ranking, however, there can be a small difference in crime rate, so in itself the comparison of the rank order only gives an indication of the difference between the two databases with regard to the concentration (and spread) of crime over the 27 judicial districts. Further analysis needs to be done, and therefore an SPPT was performed.

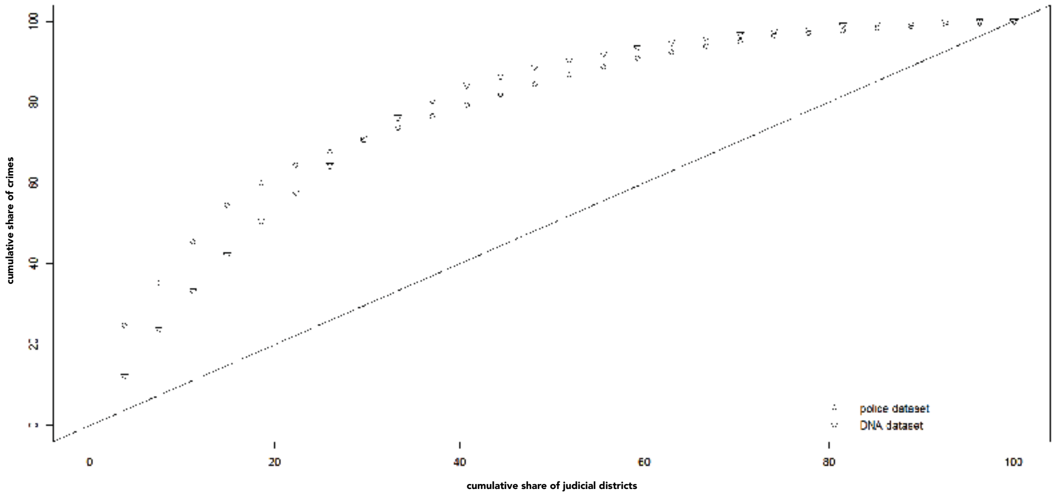


Figure 12: Lorenz curve for the police dataset and DNA dataset

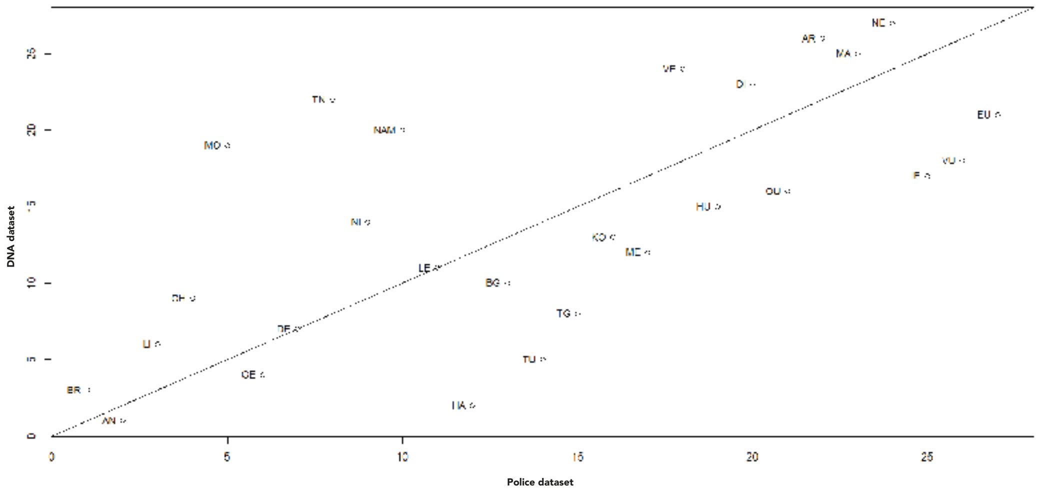


Figure 13: Rank order crime rates in 27 judicial districts in the police dataset plotted against rank order in the DNA dataset (high to low)

7.4.3. Spatial similarity

The adapted SPPT is used to compare the spatial distribution of the four crime types separately in the two datasets. When spatial patterns are similar in both datasets, the DNA dataset adequately reflects the spatial distribution of the unsolved crimes present in the police dataset.

Figure 13 maps the spatial similarity at judicial district level between the police dataset and the DNA dataset. Districts with a local S-index of +1 (i.e., black districts) have a higher crime proportion in the DNA dataset than in the police dataset. This means that in these districts there is an over-representation of the number of unsolved registered crimes in the DNA data. In districts with a negative local S-index (i.e., grey districts) the police dataset has a higher crime proportion than the DNA dataset. In these districts there is an under-representation of the number of unsolved registered crimes in the DNA data. Districts with similar percentages of crime in both datasets have a local S-index of 0 (i.e., white districts).

With a value of 0.22, aggravated burglary has the lowest global S-index resulting from the adapted SPPT for of all four crime types (**Table 10**). In contrast to the southern part of the county (Wallonia), most districts in the north (Flanders) have a higher percentage of unsolved aggravated burglaries in the DNA dataset than in the police dataset (**Figure 14**(2)). This clear distinction between Flanders and Wallonia says nothing about a difference in crime level or a difference in the clearance rate in both parts of the country, as only unsolved registered crimes were studied. For the other three crimes types, the global S-indices resulting from the adapted SPPT are all above the 0.80 threshold. Spatial similarity (white districts) is observed in 24 judicial districts for violent theft (**Figure 14**(1)), in all 27 judicial districts for lethal violence (**Figure 14**(3)) and in all but one judicial district for sexual offences (**Figure 14**(3)).

Table 10: S-index proportion difference test

	S-index
Violent theft	0.8518519
Aggravated burglary	0.2222222
Lethal violence	1
Sexual offences	0.962963

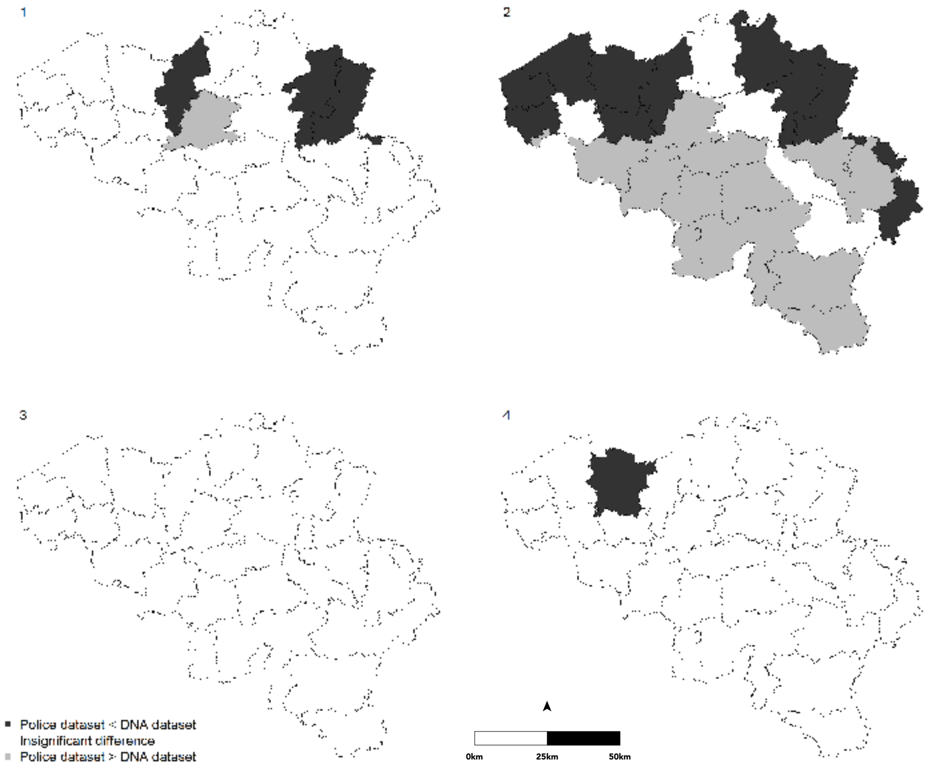


Figure 14: Local S-indices for violent theft (1), aggravated burglary (2), lethal violence (3) and sexual offences (4)

7.5. Conclusion and discussion

It is obvious that data on unknown offenders is difficult to collect. DNA databases do allow unsolved registered crimes and their unknown offenders to be studied based on the DNA traces they leave behind at crime scenes, and some authors have already successfully made use of DNA data to study the spatial behaviour of unknown offenders (see e.g., Lammers, 2014a; Wiles & Costello, 2000). As DNA databases only store a subset of all unsolved registered crimes, it is important to assess the accuracy of the DNA data in representing the spatial behaviour of unknown offenders. This will allow an evaluation of whether general statements based on DNA data can be made about unsolved crimes and unknown offenders, and whether there are differences between the spatial offending behaviour of known and unknown offenders.

In this study the spatial distribution of unsolved crimes committed in 2014 from two datasets were compared: the Belgian General Police Database and the Belgian National Genetic Database. Results from the Lorenz curves and the generalized Gini coefficients, aggregated over the four crime types, show that in both datasets crime is unevenly distributed over the 27 judicial districts and is concentrated in only a few districts. Both datasets show a similar degree of crime concentration: 50% of all crimes occur in fewer than five of the 27 judicial districts, although these are not the same five districts in both datasets. Comparable results were obtained when studying the crime concentration of the four crime types separately.

A spatially similar pattern between unsolved registered crimes in the police database and the DNA database is an important prerequisite for using DNA data to study the spatial behaviour of unknown offenders. A different spatial pattern in the DNA database, whether caused by an overestimation or under-estimation of crimes in certain districts, would lead to confounding results concerning the spatial distribution of registered crimes committed by unknown offenders. An adapted SPPT based on the direct differences in proportions (Wheeler et al., 2018) assessed the spatial similarity between the two datasets. Three important findings emerged from the adapted SPPT test. First, spatial similarity is different for each of the four crime types studied. Second, spatial similarity is (very) high for three of the four crime types: violent theft, lethal violence and sexual offences. Third, both an over-representation and an under-representation of the number of unsolved registered crimes in the DNA database was found for aggravated burglary.

It is important to take a closer look at where these distinct spatial patterns may originate. If the crime type alone had caused the difference between the two datasets, the same local S-index would be expected in every district. For example, the crime scenes of aggravated burglary may be more conducive to the collection of DNA samples. This could lead to an over-representation of this type of crime in the DNA database in all 27 judicial districts (i.e., 27 black districts in **Figure 14**). However, this is not the case. If the district alone had caused the difference between the two datasets, the same local S-index for every crime type per district would be expected. Only five of the 27 judicial districts have a local S-index of 0 (resulting from the adapted SPPT) for each of the four different crime types. None of the 27 judicial districts has a negative local S-index or a positive local S-index for all four different crime types. It seems more likely, however, that both the crime type and the DNA retention policy of a district may have an impact. Certain judicial districts may have a more active DNA retention policy, focusing on all crime types or only on specific ones. Districts can give a different priority to DNA as an investigative method or have a different budget for DNA analysis. For example, aggravated burglary shows a higher number of districts with an over-representation of crimes in the DNA dataset, mainly districts located in Flanders. Aggravated burglary crime scenes in Flanders may be systematically searched for DNA, resulting in a higher relative crime count in the northern part of Belgium. In which case, even a lower S-index may be valuable for spatial crime research.

By integrating information on law enforcement priorities or DNA retention policies, for example, insight is obtained into the (lack of) representativeness of the DNA data. The opposite is also true: an evaluation of the spatial similarity between the police and DNA data can provide insight into the DNA retention policy of the different districts. So even though no global spatial similarity was obtained for all crime types, this study has its value. First, conducting an SPPT offers an opportunity to evaluate the policy. Generally, Belgian DNA retention policies are not publicly available and it may be difficult to point out differences between the policies of different judicial districts. Indeed, the only information available on DNA retention policies comes from interviews with police officers, forensic experts and magistrates involved in collecting and analysing DNA traces. Moreover, the experience of the police in detecting DNA traces or limited budgets may also play a role in the nature and extent of the DNA profiles that are included in the database, regardless of the prescribed policy. Second, there are (many) districts in which the spatial pattern is similar to the police dataset. Although there may be more crime records available in the police data, the use of DNA data would be preferred since these data also contain information on unknown offenders. This is the reason why Lammers (2013) and Wiles & Costello (Costello & Wiles, 2001; Wiles & Costello, 2000) used DNA data in their spatial research. DNA databases allow us to study the spatial behaviour of serial (co-)offenders since a distinct unknown profile can be linked to different crime scenes and other (known and unknown) offenders. The fact that the offenders of unsolved crimes may be serial offenders or have co-offenders cannot be taken into consideration when relying solely on police-recorded crime data. This is a major drawback since research indicates that committing crime is not only a group behaviour, but also that a relatively small proportion of offenders is responsible for a large share of all crimes (Farrington et al., 2006; Felson, 2003; Lantz & Ruback, 2016; Reiss, 1988; Warr, 2002). However, some caution is needed when interpreting these high global S-indices as an indication for the validity of the DNA database as data source. After all, the high similarity between the spatial patterns of lethal violence and sexual offences is caused by the zero crime count in the DNA data (in combination with a low crime count in the police dataset). Therefore, there would be no unknown offenders to study in these districts.

This study provides a starting point for further research. As already mentioned, this research offers insights into the differences between the policies at the judicial district level, influencing the type and amount of profiles stored in the DNA database. Taking this a step further, these insights could help to equalize and improve policy decisions on the collection of forensic DNA across the different judicial districts. Second, in addition to crime type and DNA retention policy there may be some other, perhaps even more relevant, factors influencing the spatial similarity between police-recorded crime data and DNA data. For example, media coverage of crimes may increase offenders' awareness about the risk of leaving DNA traces, resulting in fewer DNA traces being found at crime scenes. Further research could identify those factors. Finally, the most pertinent research opportunity with DNA data lies in the study of unknown offenders and their behaviour. In this research, individual crimes were studied, but DNA databases are even better suited to study links between crimes committed by the same unknown offender and between different unknown offenders committing crimes together. This enables serial and co-offending behaviour of unknown offenders to be studied, which is difficult if not impossible with police-recorded crime data. In contrast, police-recorded crime data are more appropriate as a data source to study known offenders, as not all known offenders are stored in DNA databases (i.e., reference profiles). Taking both police-recorded crime data and DNA data together not only allows researchers to study and compare the criminal behaviour of known and unknown offenders but also allows them to make links between known and unknown offenders.

In summary, studying the spatial similarity between the two sets of data establishes how representative the DNA data is at a local level, keeping in mind the possible low or even absent crime counts for certain crime types in the DNA data, and taking into account different factors that may influence its representativeness. Once it is established how representative the DNA data is, it may then be used to contribute to an analysis of unsolved crimes and their offenders. The study of unsolved crimes and unknown offenders is an important issue in geospatial criminology, as it furthers the understanding of crime in general, and of the entire offending group. Therefore, used in context, DNA data could lead not only to new theoretical insights, but also to better detection and crime prevention strategies.

3

PART 3:

NETWORKS OF UNKNOWN OFFENDERS AND THEIR CRIMES

"One of the questions still to be answered by network researchers is whether, and to what extent, the patterns in co-offending found in official data translate to undetected crimes."
(Bouchard & Malm, 2016, p. 15)

In Part 3 of this dissertation, the focus is on the possibility of studying unknown offenders and their serial and co-offending behaviour using DNA data. The first chapter explains the importance of a network approach within criminological research. In the two subsequent chapters, this network approach is tested empirically on an integrated dataset of police-recorded crime data and DNA data.

8. Serial co-offending and the network approach

Committing crime is a group behaviour (De Moor et al., 2018b). At least half of all crimes involve more than one offender (Andresen & Felson, 2010; Felson, 2003; Lantz & Ruback, 2016; Warr, 2002) and about two-thirds of all offenders commit their crimes with others (Reiss, 1988). Generally, offenders do not have long-term co-offending partners (Reiss & Farrington, 1991) and most offenders commit joint offences as well as lone offences (Goldweber et al., 2011; Reiss, 1988; Reiss & Farrington, 1991; van Mastrigt, 2014). Burglary and robbery are the typical co-offending crime types (van Mastrigt & Farrington, 2009). Young offenders in particular are more likely to commit crimes in the company of others. Although serial offending and co-offending are generally studied separately, there is an import link between the two offending types: co-offenders often show higher crime rates than solo offenders (Andresen & Felson, 2010; Tontodonato, 1996). In other words: *"persistent deviance typically is not a solitary enterprise; rather it best flourishes when it receives group support"* (Matza, 1967, p. 63). Moreover, only a relatively small number of offenders is responsible for a substantial portion of all crimes (Everson, 2003; Farrington et al., 2006; Piquero, Farrington, & Blumstein, 2003; Staley, 2005; Thornberry & Krohn, 2000).

Offenders also learn how to commit crimes from others: criminal behaviour is learned behaviour. This is illustrated in a German study on the effect of popular TV crime series on criminal behaviour (i.e., the 'CSI effect'). The researchers found that convicted offenders considered friends and acquaintances to be the best source of information on how to commit a crime, whereas students and pupils preferred to obtain this information from the internet (Baranowski, Burkhardt, Czernik, & Hecht, 2017). The social aspect of criminal learning is also reflected in a number of criminological theories including the differential association theory by Sutherland (Sutherland & Cressey, 1960) and the self-control theory by Gottfredson and Hirschi (1990). Warr (2002, p. 4) perfectly summarizes the social nature of criminal behaviour – it "*is not merely an incidental feature of crime, but is instead a potential key to understanding its etiology and some of its most distinctive features*".

It is therefore not surprising that Papachristos (2011) vigorously defended the network perspective in his article entitled 'The coming of a networked criminology?' Criminologists use many network concepts that relate to interdependence between actors, such as peer influence (McGloin, 2009) or social bonding (Hirschi, 1969). Furthermore, social network analysis (SNA) is a methodological approach based on graph theory and is used to mathematically and visually analyse social networks. Papachristos stated that criminological theories could be evaluated and reworked using SNA (2011, pp. 102-103). The same is true for Bouchard and Malm (2016, p. 1), who stated that network methods can improve the study of criminal behaviour because they take into account the impact of the social environment on crime.

In this part of the dissertation, police-recorded crime data is supplemented with DNA data and then, based on a network approach, links are identified between both known and unknown offenders, and their crimes. This creates a one-to-one match between DNA data and police-recorded crime data on the same crime cases. To date, very few studies have used DNA data to study the co-offending behaviour of unknown offenders. Jeuniaux and colleagues (2016), for example, studied the links between known and unknown offenders within the same crime cases, but only used a DNA database. The problem with this approach is that not every offender registered in police-recorded crime data is stored in the DNA database, and many known offenders are therefore omitted from the analysis. However, if known offenders from police-recorded crime data are combined with unknown offenders from DNA data associated with the same crime case it is possible to detect previously unidentified co-offenders. These offenders can, in turn, be linked to additional crimes committed with (other) co-offenders. Unknown offenders in the DNA database could also demonstrate a previously unidentified link between other known offenders in the recorded crime data. Therefore, combining police-recorded crime data and DNA data could help to answer the pertinent question of Bouchard and Malm (2016, p. 15): "*whether, and to what extent, the patterns in co-offending found in official data translate to undetected crimes*".

9. Integrating police-recorded crime data and DNA data to study serial co-offending behaviour

The central research question in this chapter reads as follows:

Q 3: To what extent does the crime picture change when unsolved crimes are included in a network analysis?

Is the picture of serial co-offending behaviour qualitatively different when police-recorded crime data are enriched with DNA data, or do they just corroborate what is already known? To answer this question the spatiotemporal distribution of crimes committed by serial co-offenders stored in the NGDB is compared with the spatiotemporal distribution of crimes in an integrated dataset of both police-recorded crime data and DNA data. To this end, a network analysis was carried out with the crimes recorded in the police-recorded crime data and the DNA data as nodes. The analysis is explained below and is a further elaboration of De Moor et al. (2018b). The spatiotemporal data used to answer the second research question was also used for the third research question, because this is the only common information available in the two databases.

9.1. Introduction

Studies on the spatial and temporal dispersion of crimes are predominantly limited to known serial offenders registered in police-recorded crime data. Johnson (2013), for example, studied the space–time behaviour of serial burglars. Lundrigan, Czarnomski, and Wilson (2010) studied the spatial distribution of crimes committed by serial sexual offenders. van Sleeuwen, Ruiter, and Menting (2018) studied the temporal distribution of various crimes committed by serial offenders. All these, and many more, studies determined that the crimes committed by serial offenders are spatially and temporally clustered. If the behaviour of unknown offenders is also taken into account, a different picture could be revealed (see e.g., Lammers, 2013), although some studies have found little support for different spatiotemporal behaviour by known and unknown offenders (see van Sleeuwen et al., 2018).

Does this spatiotemporal clustering also apply to co-offenders? To date, the spatiotemporal behaviour of unknown serial co-offenders has not been studied. Information on where and when unsolved crimes are committed is registered in the police database, but they are seemingly independent crimes committed by different offenders. In contrast to the unsolved crimes stored in the DNA database, these crimes cannot be linked to each other due to the lack of a known offender. In this chapter a first step into this unexplored area is taken by combining spatiotemporal offending data from known and unknown serial co-offenders. This study takes a more methodological perspective. The prevalence and characteristics of serial co-offences obtained from the police-recorded crime data and the DNA data are used to evaluate differences between the networks obtained from both datasets. No statements are made about the offending behaviour of 'the' serial co-offender because the focus of this chapter is on the changes the networks obtained from police-recorded crime data undergo when integrated with data from unknown offenders in the DNA database. The research is therefore not limited to certain types of serial and co-offenders but preferred a broad conceptualization of serial co-offenders: offenders who committed at least two crimes, of which at least one crime committed in company of another offender.

9.2. Data

The study uses six years of recorded crime data (2010 through 2015) relating to the four selected crime types from both the NGDB and the Belgian General Police Database: aggravated burglary, violent theft, lethal violence and sexual offences.

In order to study the added value of the NGDB, a selection of the available data was made. First, to ensure that the offenders in the DNA dataset were different from those selected from the police-recorded crime data, only crimes committed by one or more *unknown* offenders were selected from the NGDB.[30] The reverse selection mechanism was applied to the Belgian General Police Database: only crimes committed by one or more *known* offenders were retained. Second, only crime cases that were present in both databases were selected. Not every crime in the Belgian General Police Database is registered in the NGDB – some crime scenes are simply not inspected for DNA traces, and in other crime scenes no suitable DNA traces may be found – but every crime registered in the NGDB *is* registered in the Belgian General Police Database, with or without known offenders. Third, only crimes involving serial offenders (i.e., offenders who committed at least two crimes in the six years under study) were selected from both databases. This resulted in a police dataset of 121 crimes committed by 71 serial co-offenders and a DNA dataset of 544 crimes committed by 232 serial co-offenders. As these figures indicate, the DNA dataset contains far more crimes and offenders than the police dataset, primarily due to the low clearance rates typical of recorded crime data. Finally, both datasets were integrated into a single dataset so that those crimes only committed by co-offenders (i.e., offenders who committed at least one of their crime in the company of another offender) could be selected. An extra selection criterion was adopted for this integrated dataset, because combining the two datasets creates an uncertainty about whether an unknown offender in the DNA dataset is the same person as a known offender from the police dataset. Therefore, networks that were composed of only two offenders of which one originates from the DNA dataset and the other from the police dataset were excluded. Thus, all remaining networks in the data definitely included crimes committed by serial co-offenders. The hypothetical example in **Figure**

---

[30] As mentioned in the introduction, offender profiles are stored in the NGDB using a code. The corresponding identity of the offender is managed by a third party, the National DNA Cell. As scientific researchers have no access to these data, it was not possible to check for possible matches between unknown DNA profiles and known offenders stored in the police dataset.

15 illustrates this. Suppose that unknown offender A from the DNA dataset and known offender B from the police-recorded dataset are actually the same person. If the offenders were incorrectly identified as two different offenders, a supposed network would be analysed that had actually been created by the researchers themselves.

The integrated dataset is more than the sum of the number of crimes and offenders from both datasets. Serial offenders who are not co-offenders in the police dataset or the DNA dataset can become co-offenders when both datasets are integrated. Looking at **Figure 15**, when only studying the police dataset data about offender B and his crimes would not be retained, as he did not commit any of his crimes with another known offender in the police dataset. The same is true for offender A from the DNA dataset. However, as **Figure 15** illustrates, when the police dataset and the DNA dataset are integrated, offenders A and B become co-offenders. Applying the selection process to the integrated dataset, data on 654 crimes committed by 372 serial co-offenders was obtained.

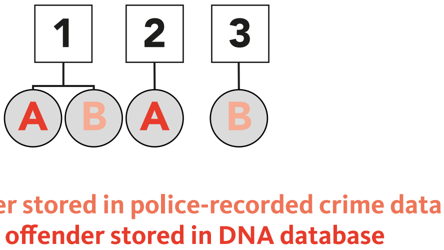


Figure 15: Hypothetical example of three crimes and two offenders

Notes: A and B become co-offenders when the police dataset is integrated with the DNA dataset. A prerequisite is that A and B are two different offenders, and not the same offender stored in both datasets.

9.3. Methods

9.3.1. Social network analysis

Graph-based techniques are used for studying pairwise relationships among the actors (Rossy & Morselli, 2018). The relationships between the actors are described in terms of *vertices* and *edges*. Vertices *V* (or nodes) are the individual actors within the network *G*, and edges *E* represent the links between the actors (Wasserman & Faust, 1994). **Figure 16** illustrates a hypothetical network G(V,E) of six offenders (A–F) responsible for six crimes in total (1–6). The bipartite graph represents a two-mode network, as there are two different nodal types: a set of actors (offenders) and a set of events (crimes). Non-directed lines connect actors on one side of the graph to the events on the other side of the graph.[31] Lines, or edges, between one offender and another or between one crime and another are not possible. Offenders

---

[31] Lines can be directed or non-directed in social networks. For example, in a one-mode network of people who live in a city, lines between siblings are non-directed (V is a sibling of W and vice versa) whereas lines representing parenthood are directed

are only linked to each other through a common crime, and crimes are only linked to each other when they have offenders in common (Knoke & Yang, 2008; Wasserman & Faust, 1994). One-mode network data can be derived from this two-mode network data, one for each set of nodes. In the one-mode network of offenders, for example, the nodes are offenders and the offenders are linked through their joint participation in different crimes (Marin & Wellman, 2011). **Figure 17** represents the one-mode network of offenders obtained from **Figure 16**.

In the one-mode network G there are two components or sub-networks: component 1 consisting of offenders A, B and C; and component 2 consisting of offenders D and E. Only offenders and their other co-offenders who commit a crime together are part of the same component. A component consists of at least two offenders connected to each other, directly or indirectly, but the offenders within a sub-network have no links with other offenders outside the sub-network (Wasserman & Faust, 1994). Node F is called an isolate as this offender has no adjacent nodes and is thus disconnected from the other nodes in the network.

The R packages 'igraph' (Csárdi & Nepusz, 2006) and 'Matrix' (Bates & Maechler, 2017) were used to conduct all network analyses.

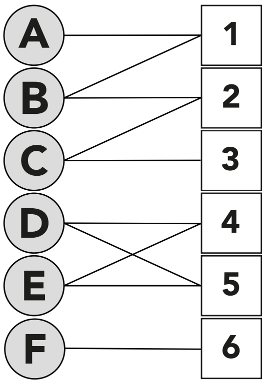


Figure 16: Bipartite graph of a hypothetical two-mode network of six offenders and six crimes

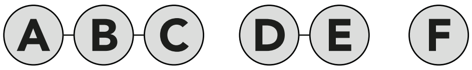


Figure 17: One-mode networks of offenders obtained from the bipartite graph in **Figure 16**

(X is a parent of Y but Y is not a parent of X). In two-mode networks, lines are always non-directed.

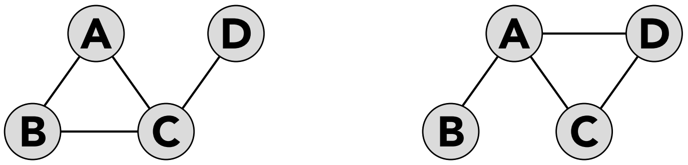
9.3.2. Jaccard similarity index

The Jaccard similarity index is a measure of similarity between two sets of data. The Jaccard index can be calculated as the number of data points the two datasets have in common, divided by the total number of different data points in the two datasets (Jaccard, 1912):

$$J(X,Y) = | X | Y | / | X U Y |$$

The minimum index is 0 (or 0%), which means that there is no overlap as both datasets have no data points in common. When the data points in both datasets are the same, the index reaches a value of 1 (or 100%), which is a perfect overlap of the two datasets (Simpson et al., 2013). A simple example can illustrate the index. Dataset X has five data points: X = {a, b, c, d, e} and dataset Y has eight data points: Y = {a, b, e, f, g, h, i, j}. As both datasets have three data points in common and there are ten different data points in total, the Jaccard similarity index is equal to 3/10 or 30%.

In this study, the Jaccard similarity index is used to determine the similarity between networks (**Figure 18**). The similarity index is then based on the number of edges two networks have in common, the number of edges in each network and the total number of edges (Fuxman Bass et al., 2013).



Edge	Network 1	Network 2
A - B	1	1
A - C	1	1
A - D	0	1
B - C	1	0
B - D	0	0
C - D	1	1

→ J(1,2) = 3/6 = 50%

Figure 18: Illustration of the Jaccard similarity index (Fuxman Bass et al., 2013, p. 1174)

9.3.3. Normalized diversity index

Diversity indexes were computed to assess the spatial and temporal variability within the obtained crime networks. In our study, the **temporal diversity index** reflects the probability that any two offences drawn randomly from a network are committed in different years. The **spatial diversity index** reflects the probability that any two offences drawn randomly from a network are committed in different judicial districts (Mazerolle et al., 2000, pp. 1153-1154).

The diversity index d_i is computed as follows:

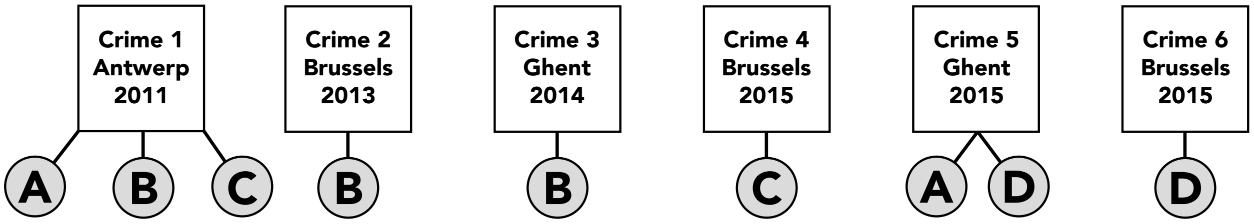
$$d_i = 1 - \sum_{m=1}^{M} p_m^2$$

where p_m is the proportion of crimes registered in a specific year or district m (Mazerolle et al., 2000, pp. 1153-1154). M represents the total number of categories or the number of possible years and districts, which is six for the temporal diversity index as there are crime data on six different years, and 27 for the spatial diversity index as all crimes are registered in one of the 27 judicial districts.

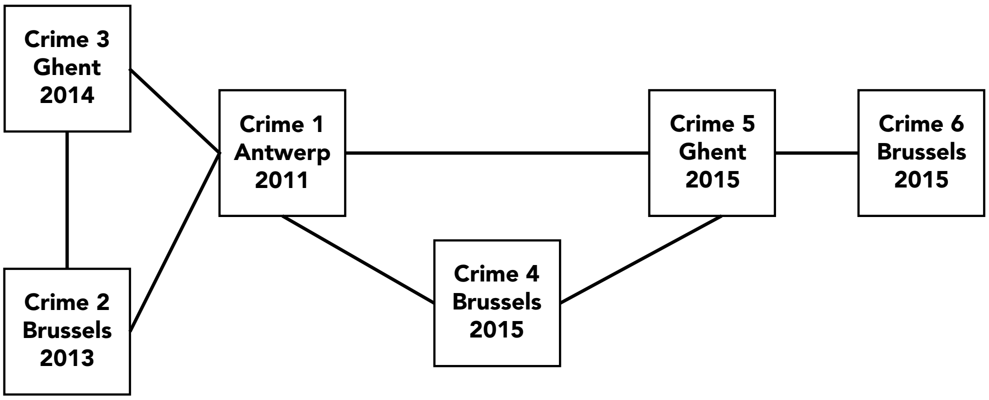
The minimum of the diversity index is zero, and this indicates that all crimes were committed in the same year/district. The maximum depends on the number of categories M or the number of possible years (six) and districts (27): $d_{max} = (M-1)/M$. A maximum spread of the crimes in time/place is obtained when every category entails an equal number of offences. This is illustrated by the network of six crimes committed in four years and in three districts presented in **Figure 19**, provides the following indexes:

$$d_{year} = 1 - \{ (0/6)^2 + (1/6)^2 + (0/6)^2 + (1/6)^2 + (1/6)^2 + (3/6)^2 \} = 0.6667$$

$$d_{district} = 1 - \{ (3/6)^2 + (1/6)^2 + (2/6)^2 + (0/6)^2 \} = 0.6111$$



Six crimes (1 - 6) committed by four different offenders (A - D)



One-mode network of crimes

Figure 19: Fictional example of six crimes committed in four years and in three districts

The index as calculated above depends on the number of categories *M*. A network A with three crimes spread over three districts will have a lower index ($d_{district}$ = 0.6667) than a network B with nine crimes spread over nine districts ($d_{district}$ = 0.8889), as only three of the 27 districts (categories) are involved in the network. However, a network with three crimes can only be spread over a maximum of three different districts. Network A and network B would have the same diversity index when the index is based on the number of crimes *n* in the network.

To be able to interpret the diversity index independently of the number of categories, a **normalized diversity index** with $d_{min} = 0$ and $d_{max} = 1$ was calculated by the formula: $d_{Norm} = d_i / (n-1)/n$. A value of zero indicates that there is no diversity in the network because all crimes are committed within the same district/year. A value of one indicates that all crimes in these networks are evenly spread over the different districts/years present in the network. For the network in **Figure 19**, this results in $d_{Norm, year} = 0.80004$ and $d_{Norm, district} = 0.73332$. Both the d_{Norm} for year and district are high, as the network is spread over four of a maximum six different years and over three of a maximum six different districts.

9.4. Results

9.4.1. Descriptive statistics

The police dataset consists of 121 crimes committed by 71 known serial co-offenders. There are 168 edges between pairs of crimes in this dataset. With only 30 networks, the police dataset has the lowest number of networks. The average network size is 4.03 crimes (min = 2, max = 10, S.D. = 2.14). The DNA dataset consists of 232 unknown serial co-offenders responsible for 544 offences. The offences are spread over 93 networks with a mean size of 5.85 crimes (S.D. = 6.11). All networks consist of between 2 and 24 crimes. Only one network contains 51 crimes. There are 1,496 edges visible in the DNA dataset, which is about nine times more than in the police dataset (**Table 11**).

By integrating the police dataset and the DNA dataset, a dataset of 654 crimes committed by 372 offenders was obtained: 102 known offenders and 270 unknown offenders. There are 1,699 edges between two crimes, spread over 108 networks with an average size of 6.06 crimes (S.D. = 6.38). As in the DNA dataset, the network size is between 2 and 24 crimes. Only one network contains 55 crimes. Combining the two datasets allows new links between crimes from each dataset to be identified. 1.85% (N = 2) of the networks are the result of the expansion of an already existing DNA network with additional crimes from the police dataset and the DNA dataset, as new links between crimes from both databases are made. 12.04% (N = 13) of the networks are the result of a combination of networks already present in the police dataset and the DNA dataset. 5.56% (N = 6) of the networks are new and could only be identified by combining both databases. In 2.78% (N = 3) of the networks the DNA dataset allows for an extension of the networks of crimes that were already present in the police dataset. Only 9.26% of the networks (N = 10) have the same crime composition as observed in the police dataset alone. Combining the police dataset and the DNA dataset has no consequences on the appearance of these networks in terms of crimes, compared with the police dataset alone. In some of these ten networks, the DNA dataset adds some extra offenders, but no extra crimes. The remaining 79.92% of the networks (N = 74) are composed of the same crimes as in the DNA dataset only. Fifty-seven of them are exactly the same as in the DNA dataset. No information from the police dataset is integrated in these networks. One network of three crimes can be found in the two separate datasets, though with different offenders. Since the crime networks, and not the offender networks, are studied, the DNA dataset has no added value in this network. In another 16 networks previously visible in the DNA dataset, the police dataset adds new offenders, but no new crimes.[32] To conclude, 89.81% of the crime networks (N = 97) from the integrated dataset clearly benefit from the integration of the two separate datasets compared with the police dataset alone.

---

[32] The two-mode network of offenders and crimes is used for the integrated dataset to make this evaluation.

Table 11: Description of the three datasets under study

	Police dataset	DNA dataset	Integrated dataset
Number of nodes (crimes)	121	544	654
Number of edges	168	1,496	1,699
Number of networks	30	93	108
Mean network size in number of crimes	4.03	5.85	6.06

9.4.2. *Jaccard similarity index*

The overlap between the different datasets is quantified in terms of shared edges. Both the number of shared edges and the Jaccard index (i.e., the proportion of shared edges relative to the total number of edges) are given in **Table 12**. Although all edges from the police dataset are present in the integrated dataset, the Jaccard index is very low, with a value of 0.10, as both datasets only share 168 of the 1,699 edges in total. Nearly all edges from the DNA dataset (1,490 of 1,496 edges)[33] are present in the integrated dataset of 1,699 edges. This results in a much higher Jaccard index of 0.87. In other words, the DNA dataset and the integrated dataset have 87% of all their edges in common, indicating that both datasets are very similar.

These two indexes provide an indication of how much the police dataset and the DNA dataset separately contribute to the integrated dataset. However, they give no indication of how big the overlap between the two separate datasets is. The same edge can be present in both the police dataset and the DNA dataset. Also, as illustrated above, a new link can be made between a crime from the police dataset and a crime from the DNA dataset. Therefore, the integrated dataset is not just the sum of the two separate datasets but includes, for example, an extra 35 edges. If the police dataset were to completely overlap with the DNA dataset, the resulting Jaccard index would be only 0.11. As both datasets only share 85 edges, the Jaccard index is just 0.05, although about half of all edges present in the police dataset overlap with edges present in the DNA dataset. This low Jaccard index indicates that the separate datasets both make a contribution to the integrated dataset in terms of the number of unique edges.

---

[33] One would expect this number to be 1,496. However, by excluding the networks composed of only two offenders of which one originates from the DNA dataset and the other from the police dataset, some crimes and related edges were removed from the integrated dataset (see 9.2).

Table 12: Jaccard similarities between edges present in the three different datasets.

	Police dataset	DNA dataset	Integrated dataset
Police dataset	1.0 (168)	0.05 (85)	0.10 (168)
DNA dataset		1.0 (1,496)	0.87 (1,490)
Integrated dataset			1.0 (1,699)

Notes: Number of edges between brackets.

9.4.3. Spatiotemporal distribution

Studying the spatiotemporal distribution of the networks in the datasets allows us to evaluate whether the three networks differ not only in size, but also in their characteristics. In ten networks present in the police dataset, all crimes were committed in only one year. Obviously, this has an impact on the average temporal distribution measures of the 30 networks in total. Focusing instead on the 20 networks active in more than one calendar year, the mean temporal diversity index increases from 0.44 to 0.65, the mean timespan increases from 2.23 to 2.35 years and the average network size is now 4.75 crimes (S.D = 2.20). A similar phenomenon was found relating to the spatial distribution. Over half of all (N = 17, 56.67%) networks in the police-recorded crime data were only active within one district, as indicated by the low overall average spatial diversity index of 0.37. When only the 13 networks with different crime locations are taken into account, the average diversity index increases to 0.85, indicating a high chance that two random crimes from one of these 13 networks are committed in different districts. These networks are on average active in 3.38 different districts and have an average network size of 4.54 crimes (S.D. = 2.79).

Table 13: Temporal and spatial spread of the three datasets under study

	Police dataset	**DNA dataset**	**Integrated dataset**
Mean (median – S.D.) timespan in number of years	2.23 (2 - 1.22)	2.61 (2 - 1.37)	2.64 (2 - 1.46)
Mean (median – S.D.) unique number of years	1.9 (2 - 0.80)	2.22 (2 – 0.99)	2.26 (2 – 1.10)
Number of networks only active in one year	10 (33.33%)	26 (27.96%)	30 (27.78%)
Mean (median – S.D.) temporal d_{Norm}	0.44 (0.50 – 0.35)	0.48 (0.60 – 0.34)	0.49 (0.63 – 0.34)
Mean (median – S.D.) unique number of districts	2.03 (1 – 1.77)	2.65 (2 – 2.04)	2.6 (2 – 2.08)
Mean (median – S.D.) spatial d_{Norm}	0.37 (0 – 0.45)	0.50 (0.6 – 0.39)	0.49 (0.5 – 0.39)
Number of networks only active in one district	17 (56.67%)	29 (31.18%)	34 (31.48%)

The identified networks in the DNA dataset have a larger spatial and temporal spread than in the police dataset (**Table 13**). The measures for spatial and temporal distribution in the integrated dataset are similar to those of the DNA dataset separately. This is not surprising since 74 of the 108 networks (79.92%) are composed of the same crimes as in the DNA dataset only. In the integrated dataset, the average diversity index for time is 0.49, compared with 0.44 in the police dataset. This already represents a 11.36% larger temporal spread, although the difference is not significant ($p = 0.41$).[34] The timespan is extended from 2.23 to 2.64 years and the crimes are spread over more years (from 1.9 to 2.26 years on average) ($p = 0.08$). The proportion of networks only active in one year decreases from 33.33% in the police dataset to 27.78% in the integrated dataset. The decrease is even greater for the number of networks only active in one district: from 56.67% to 31.48%. The networks from the integrated dataset are spread across more districts than the police dataset (2.6 versus 2.03) and also the mean spatial diversity index increases significantly to 0.49 ($p<0.01$).[35] The spatial dispersion of the crimes at network level is visually displayed in **Figure 20** and **Figure 21**. For example, network X in the plot of the networks in the integrated dataset is the biggest, with 55 crimes spread over 14 different districts. Only one crime was committed in two of these districts, and between two and seven crimes were committed in each of the other districts. This results in a high normalized spatial diversity index of 0.93. Network Y, in contrast, is composed of four crimes, all committed in the same district. The normalized spatial diversity index of this network is zero.

---

[34] The one-sample Wilcoxon signed rank test was significant at 0.05 level (V = 119).

[35] The one-sample Wilcoxon signed rank test was significant at 0.05 level (V = 119.5).

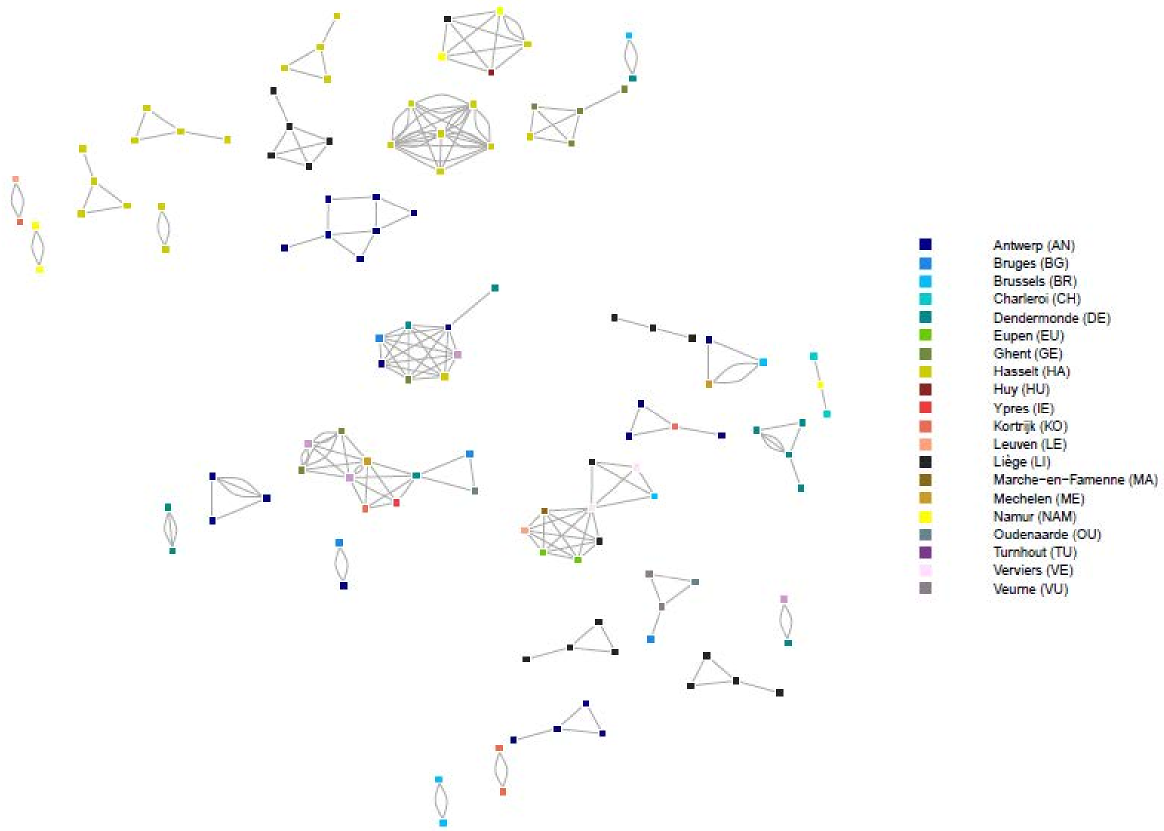


Figure 20: Spatial dispersion of crimes at network level for the police dataset

Notes: The districts Arlon (AR), Dinant (DI), Mons (MO), Neufchâteau (NE), Nivelles (NI), Tongeren (TG), Tournai (TN) are absent in the police recored crime data.

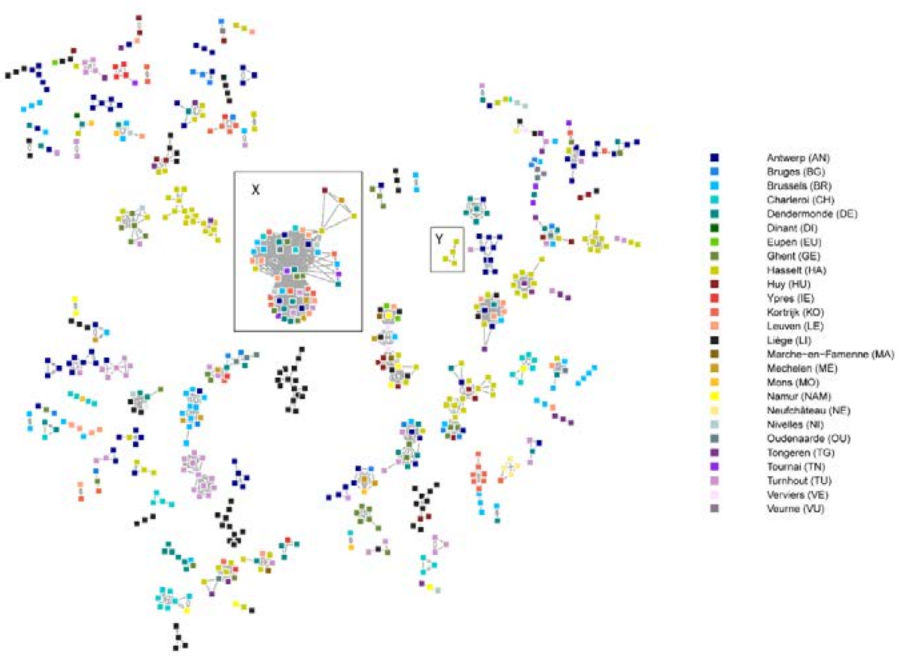


Figure 21: Spatial dispersion of crimes at network level for the integrated dataset

Notes: The district Arlon (AR) is absent in the integrated dataset.

9.5. Conclusion and discussion

Serial co-offenders are particular difficult to study because police-recorded crime data underestimate the true magnitude of crime, and also because arrested offenders are not willing to inform on potential accomplices or report additional offences (Alarid et al., 2009). The use of DNA databases makes it possible to study unknown offenders of detected crimes, and multiple offences by the same unknown offender can be linked. In this chapter the NGDB was used to study serial co-offences. The central question is whether and how the picture of offending networks is different in a dataset that integrates police-recorded crime data and DNA data on the same cases, compared to the police-recorded crime data only.

The DNA dataset was found to be an important quantitative enrichment for the police dataset. A low Jaccard index illustrates that the two datasets only have a small overlap, indicating that both contain different information. Indeed, combining both datasets on the same crime cases made it possible to study five times more crimes than would have been possible with just the police dataset, and the offences and offending behaviour of unknown offenders could be studied. In addition, links were made between crimes committed by known offenders from the police dataset and crimes committed by unknown offenders from the DNA dataset. As a result, four times more networks were identified, which involve on average 1.5 times more crimes than those in the police dataset. A high number of these networks only consisted of crimes committed by unknown offenders from the DNA dataset. Moreover, less than 10% of the crime networks in the integrated dataset have exactly the same crime composition as those in the police dataset alone.

This research confirms the findings of Lammers (2013) that police-recorded crime data underestimates the spatial dispersion of serial offenders. Both the spatial *and* temporal distribution of the networks of crimes is lower in the police dataset than in the integrated dataset. The networks from the integrated dataset have a longer time span and the crimes are committed in more distinct years and districts: on average, the offending networks have about 12% larger temporal spread and 15% larger spatial spread than the networks identified in the police dataset. However, some caution is needed in interpreting the results, especially as small sample sizes are involved. For example, the larger temporal spread in the integrated dataset was not significant. Integrating police-recorded crime data with DNA data did not have a substantial impact on the temporal spread of the networks found. Nevertheless, these findings provide additional empirical support to the previously mentioned statement that police-recorded crime data only represent a proportion of the total offender population. Disregarding unknown offenders may mean that a substantial number of crimes are ignored, and this may result in a considerable loss of spatiotemporal variation.

The proposition to integrate police-recorded crime data and DNA data is important for both theoretical and operational research. As has already been mentioned, DNA databases have an added value in the study of serial co-offenders, especially when combined with police-recorded crime data. In particular, the results of this research indicate that known and unknown offenders may have different offending patterns. New theoretical insights into offending behaviour will follow from the use of a SNA on an integrated dataset. Police forces can also benefit from the new insights obtained from an integrated dataset, both to prevent and solve crimes. First, a social network analysis of the integrated dataset can also help to discover networks and, ultimately, to prevent crime. Our analysis illustrates that the integration of the police-recorded crime data and the DNA data not only expands the networks visible in police-recorded crime data, but also reveals crime networks that were not visible in the police-recorded crime data. As a result, offenders who appeared to be of marginal importance in the police database may have a central role in the integrated database, or vice versa. This new information may guide the police in their investigation. Arresting those key offenders may result in a disruption of the network (Berlusconi, 2017). This may also prevent other actors from committing new crimes as they may have lost their 'partner in crime'. Second, an integrated dataset increases the opportunities for crime 'scenario building' (van der Hulst, 2009) – reconstructing what might have happened and identifying who was involved. For example, identifying unknown offenders becomes more feasible when it is discovered that they have committed a crime with a known offender. Unknown offenders present in DNA databases could also resolve a missing link between other known offenders in police-recorded crime data. Offending groups (i.e., 'Who offends with whom?') become visible, as do the broader accomplice networks of offenders indirectly connected to each other (i.e., 'Who (probably) knows whom?') (Chattoe & Hamill, 2005). Studying a pattern of crimes committed by an unknown offender may also create new leads. Furthermore, the fact that offenders are committing their crimes in more districts and time periods than previously thought means that the police should look beyond the crime case that is initially brought to their attention, to a broader time scope, but also a broader geographic area. Thus cooperation between police forces from different judicial districts will be needed before the benefits of an integrated dataset can be fully exploited.

Two limitations mentioned in the introduction to this dissertation specifically apply to this study. First, the identified networks may have been even larger and more dispersed if other crime types represented in the NGDB, such as drug use or organized crime, had been integrated into the analysis. This seems likely, as most offenders are generalists and do not limit themselves to committing only one type of crime (Leary & Pease, 2003; Nieuwbeerta et al., 2011; Piquero et al., 2003). The second potential limitation concerns the aggregated spatial level used in the study. The information that was made available was restricted to the district level due to privacy concerns, and it was therefore not possible to detect spatial variation between two crimes committed in different cities within the same district. The geographic dispersion could be much higher than was found in this study. The same is true for an aggregated temporal level. In this study, only variations between different years, but not within one year, could be made visible. However, the problem of aggregated data should not be exaggerated. The added value of an integrated dataset for criminological research was demonstrated simply by observing a difference between the temporal and spatial dispersion at the aggregate level between the police dataset and the integrated dataset.

This study is only a first exploration of the potential of a dataset that integrates police-recorded crime data and DNA data by means of one-mode crime networks. Although this perspective was clearly valuable, the necessary next step seems to be to study the one-mode network of offenders, especially in the light of research on hidden populations, of which criminal networks are an example. This missing data problem will be addressed in the next chapter by using networks of offenders that include both known and unknown offenders.

10. Assessing the missing data problem in criminal network analysis using DNA data

The research question to be answered in this chapter reads as follows:

Q 4: To what extent does the offender profile change when unknown offenders are included in a network analysis?

The chapter explores whether and how the picture of offender networks in police-recorded crime data differs from the picture of offender networks in a dataset that integrates this data with DNA data. The chapter takes a first step into the unexplored area of DNA datasets to search for an answer to the missing data problem in criminal network analysis by comparing relevant network statistics.

10.1. Introduction

It is not easy to map criminal networks. They have fuzzy boundaries and dynamic relationships (Sparrow, 1991). The partnerships between different offenders can be temporary (only for one crime, for example) or more permanent in nature (Weerman & Kleemans, 2002), although offenders do not usually commit multiple crimes with the same co-offender, except in more specialized groups (Reiss & Farrington, 1991). As a result, co-offenders often belong to multiple offending groups at the same time (Warr, 1996). But perhaps the main obstacle to the study of criminal networks is the incompleteness of the available network data. Unlike social networks such as friendships or working relationships, criminal ties are less visible, as offenders try to conceal their crimes and ties with criminal friends. Consequently, criminal networks are incomplete and both nodes and edges are missing (Sparrow, 1991; Xu & Chen, 2005).

Depending on the data collection method, missing data in networks can have multiple causes: the boundary specification problem (BSP), respondent inaccuracy and non-response in network surveys or interviews and the study design. For example, the study design can create a fixed choice effect, where bias is caused by limiting respondents to naming, say, three friends when in reality they have at least ten friends (Kossinets, 2006). The BSP is the most important factor in this dissertation. The BSP refers to the impact that the non-inclusion of actors or nodes and the different affiliations between those actors can have on the network (Laumann, Marsden, & Prensky, 1987). The BSP is of particular interest in the study of criminal networks, as only detected offenders and their crimes can be integrated in the network. In other words: the external boundaries of the network lie where the police and court files end (Berlusconi, 2013, p. 63; Campana & Varese, 2011, p. 20). However, these boundaries can be very restrictive, as in many Western countries the clearance rate of registered crimes is very low (De Wree et al., 2006; Lammers &

Bernasco, 2013). Unknown offenders and their offences, and the unknown crimes of known offenders, remain out of reach to those studying offending behaviour, and as a result criminal network data is incomplete as part of the existing nodes and/or links are not visible (Coles, 2001).

Criminological researchers have used several sources to extract network data: surveillance data from communications using telephone, e-mail or personal contact (e.g., Campana & Varese, 2011), interview data with offenders (e.g., Vlaemynck, 2016), experiential knowledge of police officers and other criminal justice agencies (e.g., McGloin, 2005), police crime reports (e.g., Papachristos & Wildeman, 2014), transcripts of court proceedings (e.g., Reid, Tayebi, & Frank, 2013) and open source media reports (e.g., Burcher & Whelan, 2015; Morselli, Giguère, & Petit, 2007). In recent years, a number of researchers have evaluated (the completeness of) these different data sources for use in network analysis. D. A. Bright, Hughes, and Chalmers (2012) recommended the use of judges' sentencing comments. The researchers admitted that these data might not provide as valuable information as that gained from other sources, but highlighted that, in contrast to offender databases or wire taps for example, judges' sentencing comments are free of charge, publicly available and concise.

Other researchers have focused on methodological issues and the impact of incomplete data on criminal networks. Malm and Bichler (2011) concluded that the use of multiple data sources can provide a more comprehensive picture of drug market networks. Berlusconi (2013) used wire taps, arrest warrants and judgment data on groups operating in an Italian province. She noted that the number of nodes and ties present in the data decreases from wire taps to arrest warrants to judgment data, although the statistical measures that describe the position of an actor within a network (degree and betweenness centrality) remain quite robust.

A more systematic way of evaluating the impact of incomplete network data is by simulating network errors in an observed network: nodes and/or edges are extracted from or added to a known network. Statistics on this reduced or completed network are then compared with the statistics from the true or real-world network (Borgatti, Carley, & Krackhardt, 2006; Kossinets, 2006; J. A. Smith & Moody, 2013; J. A. Smith, Moody, & Morgan, 2017; Xu & Chen, 2008). This allows the different error types that are possible in network data to be assessed. For example, the impact of false negative nodes and false negative edges can be studied by deleting nodes or edges in criminal offending networks. A false negative node refers to the absence of a person in the network who should be present as he is an offender. A false negative edge means that the relation between two offenders is not observed in the network. The two offenders are not registered as co-offenders even though they actually are. The impact of false positive nodes and false positive edges can be assessed by adding nodes or edges in criminal offending networks. A false positive node refers to a person registered as an offender who is not the offender of the crime and should therefore not be part of the network. False positive edges appear when relationships between offenders are incorrectly present in a network: two offenders are registered as co-offenders, but have not committed any crimes together (Frantz, Cataldo, & Carley, 2009; Wang, Shi, McFarland, & Leskovec, 2012). A less commonly studied measurement error is false aggregation and disaggregation. Two nodes are falsely aggregated when they are wrongly regarded as one node. The opposite applies to false disaggregation: one node is wrongly regarded as two separate nodes in the network. The impact of these errors is assessed by aggregating or disaggregating nodes. In the former, edges of node A are connected to node B and node A is removed afterwards. In the latter, node A is split into two nodes A and B. Some of the edges of node A are randomly removed and added to the new node B (Wang et al., 2012).

Most researchers using simulated network errors apply random errors to real-world (i.e., observed) networks (e.g., Huisman, 2009; J. A. Smith & Moody, 2013; Wang et al., 2012) or to simulated, random networks (e.g., Borgatti et al., 2006; Frantz et al., 2009) to assess the impact on centrality measures (see 10.3). In a recent study, J. A. Smith et al. (2017) applied non-random errors to 12 real-world networks by removing nodes proportional to their centrality. Most of these studies simulating network errors conclude that the effect of missing data depends on a number of factors (J. A. Smith et al., 2017). First, the lower the sample coverage, the more the network estimates are corrupted (Galaskiewicz, 1991). The nature of the missing data is a second factor. For example, bias is worse when more central nodes are missing. In other words, non-random missing data cause a higher bias. Third, missing data do not have the same effect on every network statistic (Costenbader & Valente, 2003). Centrality measures seem to be more robust against missing network data than other network statistics (J. A. Smith et al., 2017). Finally, the characteristics of the network will also influence the effect of missing data. Smith & Moody (2013), for example, found in their research on random missing nodes in different empirical networks that larger, more centralized networks are generally more robust to missing data. Borgatti et al. (2006) concluded that, except for edge deletion, centrality measures of dense networks are the most robust against random errors in network data. Frantz et al. (2009) also found that network errors may have a different impact on centrality measures, depending on the network topology (uniform random, small-world, core-periphery, scale-free or cellular networks). **Table 14** gives an overview of the main characteristics and results of the studies discussed above.

Current research on imperfect or missing data in networks has an important limitation. Generally, random errors are applied to real-world or generated networks. However, as in many other network contexts, missing data are non-random in criminal networks (Sparrow, 1991). Some offenders may be more likely than others to be absent in police-recorded crime data. The non-random removal of central nodes, as performed in the study by J. A. Smith et al. (2017) also doesn't seem an adequate answer, as the missing of nodes and edges may be related to features other than the position one takes in a network. Although not in a network context, Lammers et al. (2012), for example, found that unknown (i.e., not arrested) offenders may differ from arrested offenders as the latter have a longer criminal career (i.e., commit multiple crimes) and have a more versatile offending pattern (i.e., are not limited to only one crime type). Moreover, in light of the low clearance rates, the most central nodes in police recorded crime data may not be the main offenders in the network. Indeed, Sparrow (1991, p. 256) stated that “the determination of centrality will depend upon *who you know most about,* rather than who *is central or pivotal in any structural sense*” (italics in reference).

The central question in this chapter is whether the offender profile changes when unknown offenders are included in a network analysis. Do known and unknown offenders differ in network measures? In other words, are unknown offenders random missing nodes or not? To assess these questions, police-recorded crime data are used to construct the ‘reduced network’ and the integrated dataset of police-recorded crime data and forensic DNA data is used to construct the ‘real-world network’, combining DNA data on unknown offenders and police-recorded crime data on known offenders. By comparing network measures of both networks, the effect of missing data (i.e., unknown offenders) in police-recorded crime data can be evaluated.

Table 14: Overview of main characteristics and results of studies on missing data in networks illustrated in the text

Study	Type of network	Network measure	Type of error		Error %	Replications	General conclusions
Borgatti et al. (2006)	Random networks	Centrality: - degree - betweenness - closeness - eigenvector	Random	Node removal/addition Edge removal/addition	1 5 10 25 50	10,000	- Accuracy of centrality measures declines smoothly and predictably with the amount of error - Different types of error had relatively similar effects on centrality robustness
J. A. Smith and Moody (2013)	Empirical networks (directed and non-directed)	Centrality: - degree - closeness - betweenness - Bonacich power score Centralization Topology Homophily	Random	Node removal	1 2 5 10 15 25 30 40 50 60 70	1,000	- Measurement bias generally increases with more missing data - Exact rate and nature of increase varies systematically across network measures - Bias dependent on the features of the network
J. A. Smith et al. (2017)	Empirical networks (directed and non-directed)	Centrality: - degree - closeness - betweenness - Bonacich power score Centralization Topology Homophily	Non-random	Node removal	1 2 5 10 15 25 30 40 50 60 70	1,000	- Bias is worse when more central nodes are missing - Bias dependent on the features of the network
Wang et al. (2012)	Empirical networks	Centrality: - degree - eigenvector Clustering coefficient Network constraint	Random	Node removal/addition Edge removal/addition Aggregation/disaggregation of nodes	From 5 to additionally removing up to 95	10	- Networks with low average clustering and less positively skewed degree distributions are most resistant to measurement error - Bias dependent on the features of the network

Kossinets (2006)	Random networks Empirical networks	Degree Clustering Assortativity Fractional size largest component Average path length	Random and non-random	Node removal Edge removal			- Boundary specification (non-inclusion of nodes or edges) can dramatically alter estimates of network-level statistics
Frantz et al. (2009)	Random networks	Centrality: - degree - between-ness - closeness - eigenvector Local clustering	Random	Node removal Edge removal	1 5 10 25 50	10-250	- The topological form of true network (uniform random, small-world, core-periphery, scale-free or cellular) has a measurable effect on robustness - Results are consistent with Borgatti et al. (2006) for the uniform random topology
Costenbader and Valente (2003)	Empirical networks	11 centrality measures	Random	Node removal	From 20 to 80, in steps of 10	25	- Some measures are more stable than others - Bias dependent on the features of the network
Huisman (2009)	Empirical networks (directed and non-directed)	Degree Reciprocity Clustering Assortative mixing (on degree) Distance	Random and non-random	Node removal Edge removal	From 10 to 90, in steps of 10	100	- Missing data can have large negative effects on structural properties of the network
Galaskiewicz (1991)	Empirical networks	Point centrality	Random	Node removal	25 50 75	10	- Bias increased considerably as sampling percentage decreased

10.2. Data

The study makes use of both the NGDB and the Belgian General Police Database. The dataset contains six years of recorded crime data (2010 through 2015) relating to the four selected crime types (i.e., aggravated burglary, violent theft, lethal violence and sexual offences). All known offenders involved in crimes that matched these criteria were selected from the Belgian General Police Database, resulting in a police dataset of 73,837 known offenders.

The police dataset was enriched with offender data from the DNA dataset to construct a dataset with both known and unknown offenders. Some precaution is needed when combining data from the Belgian General Police Database and the NGDB. The same person may be registered in both as a known offender. Profiles of known offenders (i.e., suspects or convicted offenders) are stored in the NGDB using a DNA code number (see 6.3). However, possible matches between known DNA profiles and known offenders stored in the police dataset could not be checked because access to the corresponding identity of the offender is prohibited for scientific research. This could lead to what Wang et al. (2012) define as false disaggregation in network data. To avoid this type of false disaggregation, only unknown offenders were selected from the NGDB.

Table 15 illustrates the process used to select unknown offenders from the NGDB. The same DNA profile can be linked to different crimes, whether it is a reference profile and/or a forensic profile. For example, the DNA profile of offender O5 is found at a crime scene, and a reference profile is obtained from the same offender O5 in the context of crime C4. Offender O5 is therefore a known offender and would not be included in the analysis. In another example, although two offenders are related to crime C6, only O6 would be retained in the analysis as the other offender, O7, is a known offender. This selection procedure resulted in a DNA dataset of 16,092 different unknown offenders.

Table 15: Data selection of unknown offenders from the NGDB

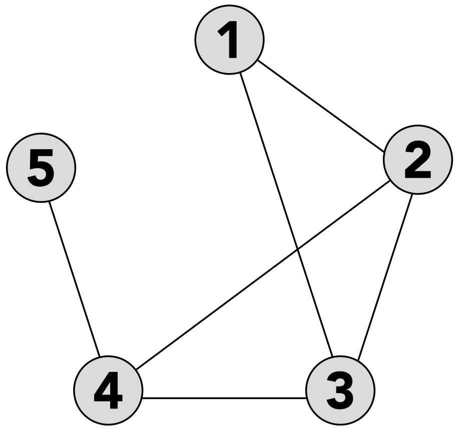
Crime	Offender	
C1	Forensic profile	O1
C1	Reference profile	O2
C2	Forensic profile	O3
C3	Forensic profile	O4
C4	Reference profile	O5
C5	Forensic sample	O5
C6	Forensic profile	O6
C6	Reference profile	O7
C7	Forensic profile	O6

10.3. Methods

10.3.1. Network analysis

As in the previous chapter, a social network analysis was performed making use of the R packages 'igraph' (Csárdi & Nepusz, 2006) and 'Matrix' (Bates & Maechler, 2017) to identify different components in the two datasets. An important variance with the previous research question is that in this chapter the analysis was carried out at the level of the offender and various network measures were used, which are explained below.

The *size* of a network is equal to the number of nodes or links (Rossy & Morselli, 2018). The geodesics is the shortest path between a pair of offenders. The *geodesic distance* is equal to the length of the shortest path (Wasserman & Faust, 1994). The *average path length* is the average length of the shortest paths for all possible pairs of nodes. *Density* describes the network cohesion. It is the proportion of the actual present edges to all the possible edges. A network containing all possible edges is a *clique* (Rossy & Morselli, 2018). Size, geodesics and density are illustrated by a small network in **Figure 22**.



Size	Geodesics	Density
5	d(1,2) = 1	6/10 = 0.6
	d(1,3) = 1	
	d(1,4) = 2	
	d(1,5) = 3	
	d(2,3) = 1	
	d(2,4) = 1	
	d(2,5) = 2	
	d(3,4) = 1	
	d(3,5) = 2	
	d(4,5) = 1	

Figure 22: Network illustrating the concepts size, geodesics and density (Wasserman & Faust, 1994, p. 111)

This study was limited to two centrality measures (i.e., degree and betweenness) for substantive reasons. Not all centrality measures are meaningful for disconnected networks composed of several distinct components like the network data in this study (e.g., closeness, see Haythornthwaite, 1996; Prell, 2013; Wasserman & Faust, 1994).[36] The *degree* refers to the number of direct links an offender has with other

---

[36] In addition, an analysis on a network composed of 73,837 known offenders and 16,092 unknown offenders demands significant computational power. To illustrate, an undirected network with n nodes can contain a maximum of n(n-1)/2 distinct edges (Scott, 2013). In theory, in the integrated network with 89,929 known and unknown offenders, in total 4,043,567,556 distinct edges are possible between two offenders

offenders by committing a crime together (Freeman, 1979; Wasserman & Faust, 1994). The minimum degree is 0 (or 0%), which means that the offender committed all his crimes without any co-offender. An offender with degree 0 is called an isolate. The maximum degree is equal to the number of nodes in the network minus 1 (or 100%). Offenders with a maximum degree committed at least one crime with every single other offender in the network. The degree only takes the local positon of the actor into account, as it is not concerned by how the other offenders are connected in the network (Morselli, 2009, p. 39; Wasserman & Faust, 1994). *Betweenness* centrality is the proportion of times an offender is located along the geodesics between any two other offenders. In other words: to what extent is an offender the direct link between two other offenders? Unlike degree, the quantity of contacts is not important, but the quality of the connections is. An offender with a relatively low degree may play an important 'intermediary' role and so be very central to the network (Scott, 2013, p. 87). As such, a network can easily be disrupted when an offender with a high betweenness centrality is arrested and thus removed from the network. Prell (2013, p. 107) describes the differences between these two centrality measures as degree centrality, emphasizing activity, and betweenness centrality, emphasizing potential control over information flow. An offender with many co-offenders will be central according to the degree centrality measures. However, offenders with fewer contacts may become more central when the betweenness centrality is measured. These two centrality measures are illustrated by a small network in **Figure 23** and **Table 16**. **Figure 24** presents in one image a visual clarification of betweenness centrality.

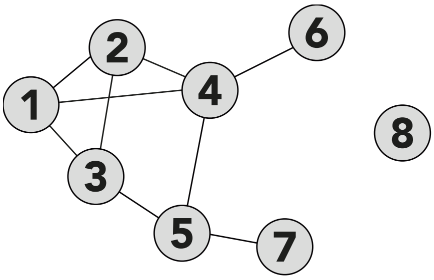


Figure 23: Network G(V,E)

Table 16: Centrality measures of network G in **Figure 23**

Node	Degree	Betweenness
1	3	2
2	3	5.67
3	3	0.67
4	3	0.67
5	4	7
6	1	0
7	1	0
8	0	0

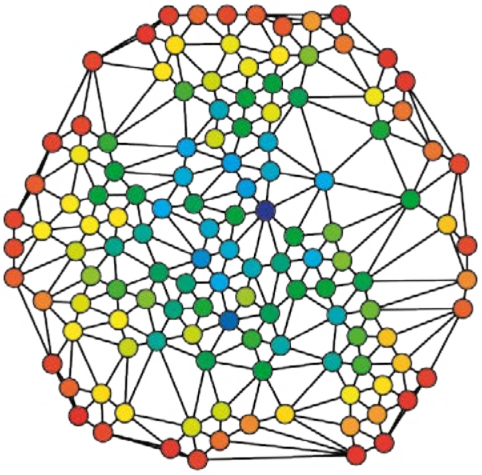


Figure 24: Betweenness centrality of each node coloured from least (red) to greatest (blue) (Rocchini, 2007)

10.3.2. Monte Carlo simulation

Previous research on the effect of missing data in networks, or network errors in general, usually performed Monte Carlo simulations to evaluate bias. This procedure entails six steps: (1) Identify or simulate a real-world network G(V, E). This network is assumed to be complete. (2) Calculate the network measures of interest for this real-world network. (3) Apply (random or non-random) network errors to the real-world network by adding or deleting a certain fraction of nodes and/or edges. The result is the distorted or reduced network G'(V', E'). (4) Calculate the network measures of G'(V', E'). (5) Repeat step 3 and 4 to obtain distributions and confidence intervals of the network measures. (6) Compare network measures of G(V, E) with those of G'(V', E') to assess the impact of the different error levels (Kossinets, 2006; J. A. Smith & Moody, 2013; J. A. Smith et al., 2017; Wang et al., 2012).

In this study, three (instead of two) network types were identified: a real-world network (G(V,E)), a real-reduced network (G'(V',E')) and different simulated-reduced networks.[37] The real-world network is composed of known offenders from the police database and unknown offenders from the DNA database. Based on the police database, the real-reduced network is created. It is a real network as it can be observed, but it is reduced because a lot of unknown offenders are missing. Finally, the simulated-reduced networks are obtained by randomly removing a certain percentage of nodes from the real-world network. Eleven different levels of missingness were applied to the real-world network by randomly

---

[37] The VSC (Flemish Supercomputer Center) provided the computational resources (Stevin Supercomputer Infrastructure) and services to compute the simulated-reduced networks and the corresponding network measures. The VSC is funded by Ghent University, the Hercules Foundation and the Flemish Government – department EWI.

removing 2, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70% of the nodes of the real-world network. Each error level was repeated 999 times.

Generally, the impact is assessed by calculating the correlation between the network measure in G(V,E) and G'(V',E') at the network level (e.g., Costenbader & Valente, 2003) or at the individual node level. In the latter case, only nodes present in both the real-world network and the reduced network can be taken into account (e.g., Wang et al., 2012). In this study, the networks with the random generated errors (i.e., simulated-reduced networks) are compared with the real-reduced network G'(V',E'). This allowed the impact of random missing nodes versus non-random missing nodes to be assessed by measuring the degree and betweenness for each of the different error levels. Furthermore, the network measures were evaluated at the network level in this study. Correlations at the individual node level would only take the known offenders into account, as only these nodes are present in both the reduced and the real world network. As the goal of this study is to assess whether and how the network measures change when integrating unknown offenders, the analysis was done at the global (i.e., network) level. The two centrality measures are averaged across the 999 replications for each error level, generating one value for each of the two measures for each of the eleven error levels.

10.4. Results

10.4.1. Descriptives

The characteristics of the real-world network and real-reduced network are summarized in **Table 17**. The real-reduced network is composed of known offenders derived from the police-recorded crime data. The real-world network contains 16,092 more nodes (21.79%) than the real-reduced network. All these additional nodes are unknown offenders from the DNA data. An extra 21,329 edges (43.83%) between offenders are created by integrating unknown offenders in the network. These edges can be between an unknown offender and a known offender who have committed a crime together, but also between unknown co-offenders. Almost half of all the components have a size of only one node (i.e., isolates), both in the real-world network and the real-reduced network (43.36% versus 45.41%). These offenders did not commit any crime with another offender.

The previous chapter showed that existing components can be supplemented with additional crimes, or several components could be merged, when DNA data is integrated with police data. The same applies to offenders. This means that the composition of the 44,743 components in the real-reduced network may have changed after the integration of DNA data. Moreover, the real-world network contains 8,305 more components (18.56%) than the real-reduced network. These components are composed of only unknown offenders.

The *average path length* is around 17 for both network types. This is quite large, given the average degree is below 2. However, this mean value can be distorted as both the real-world network and the real-reduced network contain one huge component, respectively 5,838 and 5,282 nodes. The second largest component contains only 144 and 136 offenders respectively. The density or network cohesion, by contrast, is low, caused by the high number of components in both the real-world and real-reduced network.

Table 17: Characteristics of the real-world network and the real-reduced network

	Real-world network G(V,E)	**Real-reduced network G'(V',E')**
	Police data and DNA data	**Police data**
Nodes	89,929	73,837
Edges[38]	69,995	48,666
Average degree	1.56	1.32
Number of components	53,048	44,743
Size of largest component	5,838	5,282
Size of smallest component	1	1
Number of isolates	38,989 (43.36%)	33,531 (45.41%)
Density	1.731021e-05	1.785309e-05
Average path length	17.44058	17.50912

10.4.2. The effect of missing data on degree and betweenness

Table 18 gives an overview of the impact of missing nodes on the degree in network analysis. Results on the degree for the real-world network, the different simulated reduced networks and the real-reduced network are summarized in the table. Obviously, the real-world network has an error level of 0%. This network contains almost 90,000 nodes or unknown and unknown offenders and almost 70,000 links between these offenders. On average, every offender in the real-world network has 1.56 co-offenders. At the bottom of the table, the values for the real-reduced network, which only includes known offenders from the police-recorded crime data, are presented. The average degree (1.32) is lower than in the real-world network (1.56), which includes both known and unknown offenders. In other words, integrating unknown offenders also revealed more co-offending relationships in the real-world network.

When comparing the number of nodes, the real-reduced network has an error level of about 18% compared to the number of nodes in the real-world network. **Table 18** also presents the results of the eleven different error levels applied on the real-world network, ranging from 2% to 70% of the nodes being randomly

---

[38] The number of edges corresponds to the number of edges present in the simplified networks. Simplified networks do not contain multiple edges between two nodes. In a simplified network, only one edge between two offenders is possible, even if these offenders may have committed multiple crimes together.

removed. Removing nodes clearly has an impact on the average degree: the degree decreases as the error percentage increases. For example, an error percentage of 40 or higher corresponds with a mean degree below one, illustrating that this reduced network contains many isolates (i.e., offenders who did not commit any crime with another offender). Logically, the number of edges also decreases with an increasing error level.[39]

A simulated error level of 15% results in the same value for degree as the 18% error level in the real-reduced network.[40] Randomly removing 15% of the known and unknown offenders from the real-world network results in the same degree as non-randomly removing only the unknown offenders from the real-world network. This can also be deduced from **Figure 25**, representing the 95% confidence intervals for degree of the 11 simulated reduced networks. The vertical dotted line on the left represents the mean degree of the real-reduced network (1.32) and the vertical dotted line on the right represents the mean degree of the real-world network (1.56). The mean degree of the real-reduced network falls within the 95% confidence interval of the 15% error level network.

Taking into account that known and unknown offenders have a similar impact on the mean degree centrality and that the average degree increased in the real-world network, this means that integrating unknown offenders also changed (i.e., increased) the degree of the known offenders. In other words, co-offending relationships between known and unknown offenders become visible when DNA data and police data are integrated.

---

[39] As different Monte Carlo simulations were performed for degree and betweenness centrality, the number of edges for the different error levels in **Table 18** is different from the number of edges for the different error levels in **Table 19**.

[40] There is a significant difference between the degree of the real-reduced network and the other ten simulated error levels.

Table 18: Degree of real-world network and reduced networks

Network type	Error percentage	Mean degree	Nodes	Mean edge count
Real-world network G(V,E)	0	1.56	89929	69995
Simulated-reduced networks	2	1.53	88130	67223
	5	1.48	85433	63181
	10	1.40	80936	56697
	15	1.32	76440	50584
	20	1.25	71943	44793
	25	1.17	67447	39369
	30	1.09	62950	34304
	40	0.93	53957	25206
	50	0.78	44965	17492
	60	0.62	35972	11200
	70	0.47	26979	6305
Real-reduced network G'(V',E')	17.89	1.32	73837	48666

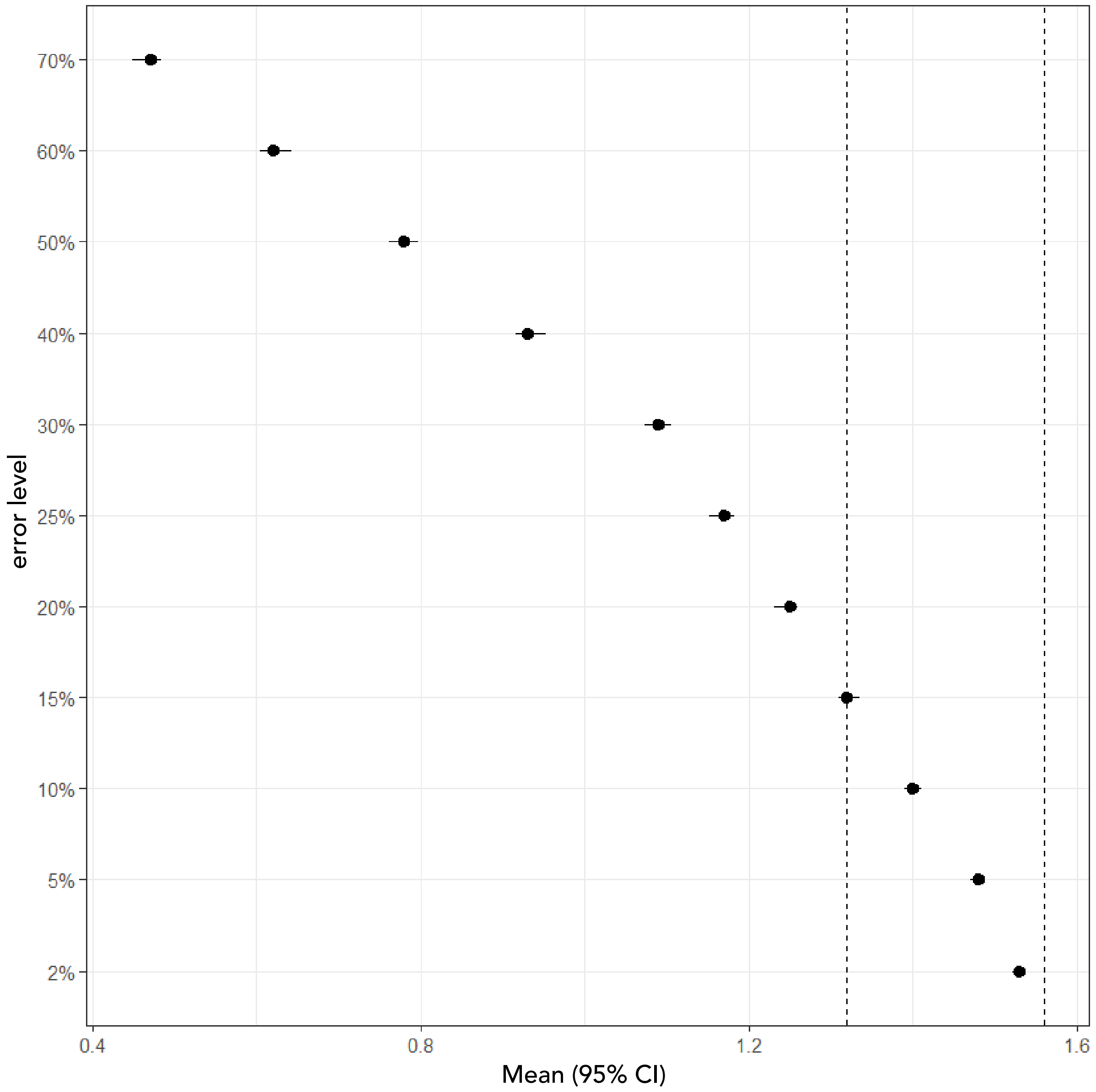


Figure 25: Forest plot 95% confidence interval for degree of simulated-reduced networks

The mean betweenness scores across the real-world network, the different simulated-reduced networks and the real-reduced network are presented in **Table 19**. It is remarkable that the betweenness centrality of the real-reduced network with a non-random error level of about 18% is about the same as the betweenness in the real-world network (3723.58 and 3715.44 respectively). The integration of unknown offenders seems not to have affected the betweenness in the real-world network.

Randomly removing a number of nodes, both known and unknown offenders, has a clear impact on the average betweenness centrality of the offenders. **Figure 26** illustrates that the mean betweenness decreases as the error level increases. Furthermore, the dotted line representing the mean betweenness of the real-world network (3715.44) is not within the 95% confidence interval of any simulated-reduced network.[41] Therefore, the results for betweenness centrality are totally different to those for degree centrality. Removing about 15% of the offenders randomly would result in a much smaller average betweenness (1328.99) than removing the same percentage non-randomly (i.e., 3723.58 in the real-reduced network). Even randomly removing only 2% of the known and unknown offenders from the real-world network results in a lower betweenness centrality than the 18% error level in the real-reduced network with only known offenders ($p = 0.01$). Known and unknown offenders have a different impact on betweenness centrality.

Table 19: Betweenness of real-world network and reduced networks

Network type	**Error percentage**	**Mean betweenness**	**Nodes**	**Mean edge count**
Real-world network G(V,E)	0	3715.44	89929	69995
Simulated-reduced networks	2	3368.27	88130	67238
	5	2845.52	85433	63194
	10	2029.64	80936	56702
	15	1328.99	76440	50606
	20	815.97	71943	44811
	25	464.79	67447	39403
	30	241.46	62950	34336
	40	51.83	53957	25208
	50	10.30	44965	17462
	60	2.00	35972	11219
	70	0.42	26979	6317
Real-reduced network G'(V',E')	17.89	3723.58	73837	48666

---

[41] As the mean betweenness of both the real-reduced network (3723.58) and the real-world network (3715.44) are similar, the dotted lines representing these values are very close to each other in **Figure 26**.

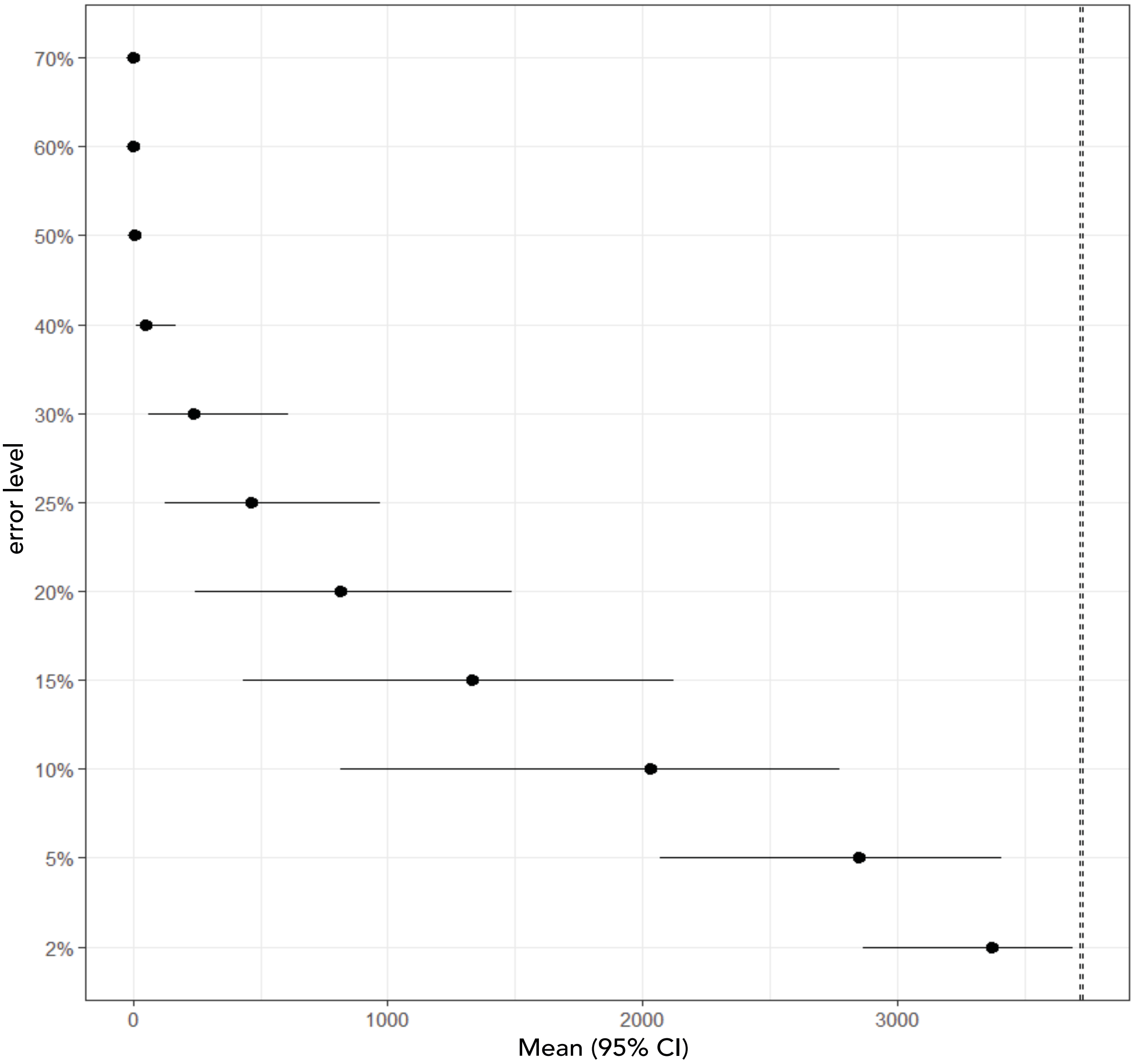


Figure 26: Forest plot 95% confidence interval for betweenness of simulated-reduced networks

10.5. Conclusion and discussion

Criminal networks are a textbook example of hidden networks, as many registered crimes are unsolved and the offenders remain unknown. In order to assess the validity of research on criminal networks it is therefore important to assess the robustness of basic network measures under the condition of missing data (Borgatti et al., 2006). As far as is known, this is the first study to integrate forensic DNA data on unknown offenders with police data on known offenders in order to study the missing data problem in criminal networks. The DNA data provides a unique opportunity to integrate missing data into police networks and is an important advancement over prior research.

This study confirms the findings from previous research. First, the higher the error level in the simulated networks, the more the network estimates are affected. This applies both to the degree and to the betweenness centrality. Second, the impact of the error level is not equal for both centrality measures. Betweenness centrality seems to be more affected by an increasing error level than does degree centrality. Third, the nature of the missing data is an important factor to consider. Although there does not seem to be much of a difference for degree centrality, there is a difference in randomly and non-randomly removing offenders from the network for betweenness centrality.

The central question in this chapter is whether and how the picture of offender networks is different in a dataset that integrates police data and DNA data, compared to the police data only. Offender networks with both known and unknown offenders may not only be bigger but also have a different structure to networks with only known offenders. The results of this study show that integrating unknown offenders has an impact on the degree, but not on the betweenness centrality. The degree is higher in the real-world network, which means that many offenders stored in the DNA database could be linked to the known offenders in the police data or to other unknown offenders. As such, the degree of the known offenders also increased by integrating the data. Removing only unknown offenders from the real-world network (i.e., real-reduced network) had no impact on betweenness. On the contrary, when known offenders are also removed from the network (i.e., simulated-reduced networks), betweenness decreases. In other words, known offenders may be more central nodes than unknown offenders in relation to betweenness.

These research findings have implications for both theory and practice. Including the unknown offenders stored in the NGDB in the database resulted in not only about 22% more offenders (i.e., nodes) but also about 44% more co-offending relations (i.e., edges), in comparison with a database solely based on police-recorded crime data. Therefore, the generally accepted assertion in criminology that at least half of all crime involves more than one offender (Andresen & Felson, 2010; Felson, 2003; Lantz & Ruback, 2016; Warr, 2002) and that about two-thirds of all offenders commit their crimes with others (Reiss, 1988) is probably an understatement. Furthermore, this research provides a unique view of the position these offenders may take in the whole offending network. Unknown offenders may be more peripheral nodes in the network. Illustrated in **Figure 24**, a blue or green node may represent a known offender and a red or orange node may represent an unknown offender. The question therefore arises as to whether the unknown offenders remain unidentified by the police because of their peripheral position, or is it, as Sparrow (1991, p. 256) states, just because they stay unidentified by the police that they have a more peripheral position in the network, although they may be more central in reality? This is an important nuance, as in the first case the integrated dataset would give an accurate picture of the centrality of the unknown offenders, whereas in the latter case the picture would be distorted. It is important to be aware of this uncertainty, because it could mean that arresting unknown offenders has a bigger impact on crime prevention than would be assumed based on their peripheral position in the network.

Some potential limitations of this study need to be acknowledged. First, it is important to bear in mind that the real-world network is unlikely to include all unknown offenders from police-recorded crime data. It is, however, an approach that cannot be achieved with any other data. Second, both the real-world and real-reduced networks contain many isolates and small components, which limits the network research possibilities but also has an impact on the mean degree and betweenness at network level. Third, as stated in the introduction to this dissertation, it is not possible to be certain that all profiles stored in the database belong to offenders and not to victims, for example. The impact of this possible error differs according to the research point of view: operational or criminological research. For example, in operational research, an unknown victim connecting two known offenders can provide new investigative leads. For criminological research, this only distorts the results. Finally, false disaggregation could also apply to a known offender from the police dataset and an unknown offender from the DNA dataset. However, according to the Belgian DNA law of 2011, offenders of a crime or attempted crime mentioned in a restrictive list of crimes have to provide their DNA profile upon conviction (Art. 14 DNA law 2011, Belgisch Staatsblad, 2011a). As the four crime types selected for this study are part of this list, an unknown DNA profile will become 'known' when there is a match with the DNA profile of a convicted offender. Consequently, except for administrative delays in taking reference samples of convicted offenders or not being able to find the convicted offender to take a reference sample, for example, this problem does not apply to the current study.

Future research might progress the work developed here. In this study only two centrality measures were taken into account. Future research could explore the possibilities to measure the impact of missing data on other network measures. The present study could also be replicated using the traditional approach in missing data research by studying the correlations of the centrality measures at individual node level in networks with different error levels. Finally, to address the limitation of the high degree of isolates mentioned above, future research could focus on only the biggest component(s) present in the real-world network. All these suggestions for future research could foster the theoretical insights on known and unknown offenders.

In summary, this study emphasized the importance of integrating unknown offenders in criminological research and gave further insights into the difference between known and unknown offenders. As in the previous chapter, this study illustrated that the difference between unknown and known offenders is not straightforward. The implications of this are further discussed in the conclusion of this dissertation (Chapter 13).

PART 4:

CONCLUSION

The most important research findings are summarized in this final part of the dissertation. In addition, the specific advantages and disadvantages of the research are described, as well as the implications of the findings for both theory and practice. Finally, an ethical reflection is provided.

11. Context

Do known and unknown offenders exhibit similar characteristics and offending behaviour (De Moor et al., 2018a)? This question is pertinent for both criminological theory and the development of crime policies. It is frequently asked in criminology, but to date research on unknown offenders has been limited because of the difficulty of studying them. This dissertation examined how DNA databases can contribute to the study of unknown offenders and their behaviour. The implications of the dissertation are threefold: methodological, as the validity of the DNA database is assessed and discussed; theoretical, as it provides more information about the behaviour of unknown offenders; and practical, as guidelines for crime prevention and control can be derived from the methodological and theoretical insights.

Literature on forensic science and forensic intelligence was studied to obtain a broader context on the use and content of DNA databases. As illustrated in Chapter 2, DNA databases have been the subject of many different studies on a range of issues, from the legal and ethical concerns surrounding their use and expansion, to the opportunities they provide for operational policing. Their use in criminological research to learn more about criminal behaviour and the broader criminal context is still limited (De Moor et al., 2017; De Moor et al., 2018b). However, DNA databases have, as a source of criminological research, three major advantages over the commonly used police databases. First, unknown offenders are stored in DNA databases as DNA profiles that are collected at crime scenes. Second, different crimes can be linked to the same unknown offender when their DNA profile is found at different crime scenes. Third, several unknown offenders can be linked to each other if their DNA profiles are found at the same crime scene. This makes it possible to study the behaviour of serial offenders and of co-offenders without these individuals being known to the police.

In other words, DNA databases provide – at least in theory – a unique data source for studying unknown offenders and their behaviour. Police data do not offer this opportunity. However, the police do not examine every crime scene for DNA traces, and not every crime scene provides suitable DNA profiles. Crimes such as money laundering or phishing, for example, are generally not represented in the DNA database. DNA databases therefore only comprise a proportion of all crimes stored in the police database.

The purpose of this dissertation was to evaluate the use of DNA databases for scientific research, taking into consideration their unique advantages and disadvantages. The central research question underpinning this dissertation was:

How can DNA databases contribute to the study of unknown offenders and their criminal behaviour?

This research question was further operationalized in four sub-questions. The first two deal with the selectivity of DNA databases and the last two focus on the advantages of the DNA database:

Q 1: Which actors and factors have an impact on the selectivity of DNA databases?

Q 2: Are DNA databases less valid than police-recorded crime databases for the study of unsolved crimes?

Q 3: To what extent does the crime picture change when unsolved crimes are included in a network analysis?

Q 4: To what extent does the offender profile change when unknown offenders are included in a network analysis?

12. Summary of research findings

This chapter summarizes the main findings of this dissertation. Each of the four research questions is discussed. First, the selectivity of DNA databases is reviewed as a (possible) disadvantage for criminological research (Q1 and Q2). Next, the advantage of the DNA database, being the ability to study unknown offenders and their behaviour, is considered (Q3 and Q4).

12.1. Selectivity

Chapters 6 and 7 of this dissertation considered whether the selectivity of DNA databases is a limitation to their use as a source for scientific research. First, an answer was sought to the question, **Which actors and factors have an impact on the selectivity of DNA databases (Q1)**, and in particular on the forensic profiles stored in the criminalistics database of the NGDB? These are the profiles that will enable us to study unknown offenders and their behaviour. On the basis of a literature study of both scientific sources and legislation, and observations on relevant actors involved in the process that a DNA trace profile passes through before it reaches the database, various selection mechanisms were defined:

- An offender may or may not leave DNA traces at a crime scene (often depending on the type of crime).
- The police may or may not be aware of the crime.
- The crime scene may or may not be investigated for DNA traces, for example by the LTWP.
- A magistrate may or may not instruct a DNA laboratory to examine a piece of evidence for DNA traces.
- A DNA laboratory may or may not find DNA profiles on a piece of evidence.
- The DNA profile drawn up by the DNA laboratory may or may not be of sufficient quality to be stored by DIS in the NGDB.

From the list above two cases can be concluded. First, a number of these selection mechanisms are general in nature and therefore will be valid in every context and in every country. For example, some types of crime do not have a clear crime scene where DNA traces can be found (e.g., phishing). However, other mechanisms are country or region specific. Each judicial district in Belgium can set its own policy priorities for investigation and DNA retention – for example, judicial districts can opt to have every crime scene examined for DNA traces. Second, the overview shows that both conscious and unconscious processes have an impact on the content of the DNA database. The most difficult selection mechanism to control is when an offender leaves no DNA at the crime scene, or when there is no clear crime scene. This and other selection mechanisms mean that the DNA database contains far fewer crimes than the police database. The datasets made available for this dissertation have illustrated this. The police dataset used in this study contains far more crimes (N = 1,186,061) than the DNA dataset (N = 25,492), although the same inclusion criteria were used for both. The research used data on the four most-recorded forms of crime in the NGDB – aggravated burglary, violent theft, sexual offences and lethal violence – committed between 2010 and 2015.[42]

What does this mean for the validity of DNA databases as a source for criminological research? Tilley and Townsley (2009, p. 375) suggested that "*of themselves, small samples are not much of a problem if the sample is representative*". In this study, all crimes stored in the DNA database are also stored in the police database (but not vice versa), and so the representativeness of the DNA database was evaluated by making a comparison with the police database. The following research question is central: **Are DNA databases less valid than police-recorded crime databases for the study of unsolved crime (Q2)?** Police-recorded crime data are frequently used in scientific research, even though this source also has a high degree of selectivity (e.g., the dark number of crime). However, this dissertation has sought to compare the validity of two data sources, and does not address the validity of police data per se; hence the term 'less' was included in the research question.

In comparison with police data, DNA data have (to date) only been used to a limited extent in criminological research. Nevertheless, some ground-breaking studies have made use of DNA data, for example Lammers' (2013) study on the probability of arrest and Wiles and Costello's (Costello & Wiles, 2001; Wiles & Costello, 2000) research into offender mobility. Wiles and Costello used DNA data to supplement police data in order also to study unknown offenders and their behaviour. Like other researchers making use of DNA databases, they were particularly interested in the benefits the database offers and they often did not pay sufficient attention to the validity of the data source. This dissertation is therefore unique with respect to previous research, as the validity of the DNA database was explicitly studied.

To answer the second research question, and thus to evaluate the representativity of the DNA data in relation to the police data, the concentration and spread of crimes across the different judicial districts between the two databases were compared. Only unsolved crimes were studied, because the specific added value of the DNA database lies in being able to study unknown offenders and their behaviour. The analyses show that in both the police dataset and the DNA dataset crime is concentrated in a limited number of districts. In the DNA dataset about 50% of all crimes occur in fewer than five judicial districts; in the police dataset about 50% of all crimes occur in fewer than four judicial districts. However, the rank order of crime percentages over the different judicial districts differs between both datasets. The districts

---

[42] It is important to mention here that the datasets that were supplied contain both solved and unsolved crimes. Specifically regarding the DNA dataset, this means that it includes crimes that are linked to both forensic profiles and/or reference profiles. To answer research questions 2, 3 and 4, a specific selection was made from these datasets.

with the highest number of registered crimes in the police-recorded crime data do not necessarily have the highest number of registered crimes in the DNA data. For example, only two judicial districts appear in the top five of both datasets.

An SPPT allowed a further validation of the representativeness of the unsolved crimes stored in the NGDB. Three of the four crime types studied (i.e., violent theft, lethal violence and sexual offences) show a high spatial similarity between the two databases. In only a limited number of districts there is an over-representation or under-representation of the number of unsolved crimes in the DNA database. Although the spatial pattern may be the same between the police-recorded crime data and the DNA data for violent theft, lethal violence and sexual offences, many judicial districts have a crime count of zero for these crime types in the DNA dataset. In other words, no crimes or (unknown) offenders are registered in these districts on which further analyses can be carried out. Aggravated burglary, on the contrary, exhibits a low spatial similarity. However, there is a strikingly higher proportion of crimes of this type registered in the DNA database than in the police database.

It is difficult to come to an overall conclusion regarding the spatial similarity between the two databases. The results depend on the type of crime and the judicial district level. This highlights the conscious and unconscious selection that came to light when considering the first research question. In addition to providing insight into representativeness, the test also made the different policies between districts tangible, which was not the original aim of the study. For example, the over-representation of aggravated burglary in the DNA database in the (most) Flemish districts corresponds with these districts' more active policy in investigating an aggravated burglary crime scene.

What do these results mean for the representativeness of DNA databases and their use in criminological research? Their use is clearly limited by the low number of crimes registered in the database, although the most frequently registered crimes in the DNA database were studied in this study. A study that only uses a DNA database will provide very little information about some crime types (e.g., lethal violence). It is important to keep this drawback in mind; but simply rejecting DNA databases as a possible data source is a bridge too far. The next section shows that the added value of DNA databases lies in combining them with police databases.

12.2. Studying unknown offenders and their crimes

Several authors point out that the use of network methods is a logical choice, given the influence of relationships on criminal behaviour. Network analysis could also evaluate current theories and insights or even supplement them (Bouchard & Malm, 2016; Papachristos, 2011). However, network analysis has only been used to a limited extent in criminological research, and its use is often limited to known offenders. This is because the ability to link offenders and their crimes is a prerequisite for network analysis, and police-recorded crime data does not contain information on unknown offenders. In this dissertation, police-recorded crime data was supplemented with DNA data to study (the behaviour of) serial co-offenders via network analysis.

The third research question focused on the behaviour of serial co-offenders: **To what extent does the crime picture change when unsolved crimes are included in a network analysis (Q3)?** Therefore, the prevalence and characteristics of serial co-offences obtained from the police-recorded crime data and the DNA data (i.e., known and unknown offenders) were compared with those obtained from the police-recorded crime data (i.e., known offenders) only. More precisely, the differences between the size and spatiotemporal distribution of crime networks from both datasets were evaluated. Three important findings emerged from this research. First, both the number and size of the networks is larger in the integrated dataset than in the police dataset alone. The larger networks in the integrated dataset mean that more crimes were committed by serial co-offenders than one can determine only on the basis of the police-recorded crime data. Combining the two datasets allows new links to be identified between crimes from each dataset: serial offenders who are not co-offenders in the police dataset or the DNA dataset can become co-offenders when both datasets are integrated. Second, both the spatial and temporal distributions of the networks in the integrated dataset are greater. On average, the crime networks in the integrated dataset had about 12% larger temporal spread and 15% larger spatial spread than the crime networks identified in the police dataset. Third, only a small minority of the crime networks in the integrated dataset appear to have the same composition as in the police dataset alone or in the DNA dataset alone. In other words, the merging of the police-recorded crime data and the DNA data did not change the composition of these crime networks, although in most cases extra offenders were added to the network. This last observation also implies that although the network analysis was carried out only at the crime level, it may also be important to take the offender level into account.

In summary, disregarding unknown offenders means that a substantial number of crimes will be ignored, with a considerable loss of spatiotemporal variation in criminal behaviour as a result. The crime picture that can be derived from the behaviour of known offenders differs from the crime picture of unknown offenders. Knowing that the DNA database is selective, it is even more striking that the crime picture changes significantly when integrating DNA data with police-recorded crime data. Moreover, this analysis only concerns four offence types. There are no immediate reasons why the crime picture of some other offence types would not change in the same way if unknown offenders were taken into account.

The final research question focused on the impact on the offender network of integrating DNA data with police recorded crime data. **To what extent does the offender profile change when unknown offenders are included in a network analysis (Q4)?** This research question was approached from the perspective of missing data in network analysis. Generally, random or non-random network errors are applied to a real-world network G(V, E) by deleting a certain fraction of nodes and/or edges. This creates a reduced network G'(V', E'). By comparing network measures (often centrality measures) of G(V, E) with those of G'(V', E'), the impact of missing data can be assessed (Kossinets, 2006; J. A. Smith & Moody, 2013; J. A. Smith et al., 2017; Wang et al., 2012). In this dissertation, the real-world network consist of known offenders from the police-recorded crime data and unknown offenders from the NGDB. By randomly deleting a certain percentage of the nodes (i.e., known and unknown offenders) from this real-world network, eleven reduced networks were obtained. However, the network that only uses police-recorded crime data is also a reduced network compared with the real-world network. Unknown offenders were non-randomly deleted from the real-world network, resulting in a real-reduced network with an error level of about 18% missing nodes. This enabled the impact of random missing nodes versus non-random missing nodes to be assessed by measuring the degree and betweenness centrality for each of the different error levels. The centrality measures are evaluated at the global (i.e., network) level, as the goal of this study is to assess whether and how the picture of offender networks is different in a network that integrates police-recorded crime data and DNA data, compared to one that is based solely on police-recorded crime data.

Two important findings emerged from this missing data research. First, the real-world network with both known and unknown offenders contains more co-offending relationships compared to the network that only contains known offenders. New links (i.e., edges) become visible between known and unknown offenders, and also between unknown offenders. Therefore, there may be more co-offenders than can be deduced from police-recorded crime data alone. Second, integrating unknown offenders from the DNA database with the police-recorded crime data on known offenders has an impact on the degree but not on the betweenness centrality of the resulting network. On the contrary, compared with the error level in the real-reduced network, randomly removing the same percentage of known and unknown offenders from the real-world network results in much lower levels of betweenness. In other words, known offenders may be more central nodes than unknown offenders, as far as betweenness is concerned.

In summary, offender networks made up of both known and unknown offenders may not only be bigger but may also have a different structure from networks that only include known offenders. As was the case in the network study on crime level, this network study on offender level emphasized the importance of integrating unknown offenders and their crimes in criminological research.

13. DNA as the silver bullet?

The DNA database was used as a data source to answer the research questions in two different ways: as a separate data source to study unsolved crimes (Q2); and as a supplement to the police-recorded crime data in order to study both known and unknown offenders and their behaviour (Q3 and Q4). In summary, two important general findings can be deduced from these studies. First, there is clear selectivity in the DNA database that limits its use as the *only* data source for scientific research. Although the most-registered offence types in the DNA database were studied in this dissertation, the prevalence of unsolved crimes in the database was sometimes too low to carry out meaningful analysis, especially for lethal violence and sexual offences. Second, the DNA database clearly offers added value when it is used in *addition* to the police-recorded crime data. The combination of the police-recorded crime data and the DNA data makes it possible to study more crimes and offenders in comparison with the police-recorded crime data alone. Links can be identified between the two datasets that are not visible in each of the two separate databases. Moreover, there are not only more crimes and more offenders to study, but also the characteristics of the offending behaviour changes when both known and unknown offenders are taken into account.

The general public often see DNA as 'the silver bullet' to solve crime (Dahl, 2010). Popular TV shows like *CSI: Crime Scene Investigation* have undoubtedly reinforced this view (Ley, Jankowski, & Brewer, 2012). But can we conclude that it is also the silver bullet for criminologists who want to overcome the limitations of traditional data sources? In other words: **How can DNA databases contribute to the study of unknown offenders and their criminal behaviour?** The answer to this question is threefold, as this dissertation clearly illustrated – the use of DNA databases in criminological research is of methodological, theoretical and practical importance.

13.1. Methodological importance

Although many researchers use official recorded crime data, these data have significant limitations. Self-report studies do not always offer a solution, as both over-reporting and under-reporting can occur, when compared to the official recorded crime data. Occasionally, researchers combine different data sources. Wiles and Costello (Costello & Wiles, 2001; Wiles & Costello, 2000), for example, integrated three different data sources in their study on the mobility of high-volume crime offenders in and around Sheffield: geocoded police-recorded crime data, DNA database hits and offender interviews. They used the DNA database to study the travel patterns of offenders, irrespective of whether they were known to the police or not. There was no one-to-one match between the police-recorded data, the DNA data and the interviews. The data related to different periods (police data were from 1995 to June 1996, and DNA database hits were from June to December 1997) and referred to different offenders. The analyses of the datasets were made separately and integrated afterwards in a general conclusion. This dissertation used a **combination of different data sources**, and the DNA data supplements the missing data in police-recorded crime data. It used an **integrated dataset** in which the information about the same crimes from the Belgian General Police Database and the NGDB were linked one-to-one. This is unique in comparison with previous research in which various data sources are combined without a one-to-one link. This approach enables more offenders (i.e., unknown offenders) and their behaviour to be studied, and also supplements the available information about the offenders stored in the police-recorded data. The **network analysis** used in this dissertation further emphasized the advantages of an integrated dataset of both police-recorded crime data and DNA data. It makes it possible to study the serial and co-offending behaviour of both known and unknown offenders. It therefore makes significant methodological advancements over previous research on serial and co-offending behaviour.

13.2. Theoretical importance

By including unknown offenders, this study contributes to research into the 'dark figure' of offenders. But the integration of police-recorded crime data and DNA data not only provided a larger sample of (linked) offenders and their crimes. Principally, the DNA data contributed to the research into registered unsolved crimes, as the findings derived from this larger sample also deviated from the findings based only on police-recorded crime data. This dissertation is therefore in line with earlier research that pointed to the difference between known and unknown offenders (Lammers, 2013; Lynam, Piquero, & Moffitt, 2004). However, it also encourages further reflection on the role and impact of unknown offenders and their behaviour in criminological research.

It is indeed difficult to study unknown offenders and their behaviour, but this dissertation has demonstrated that it is certainly possible if DNA databases are used. Although the selectivity of DNA databases limits the number and types of crimes that can be studied, common crimes such as burglaries can be studied. But does this selectivity also have an impact on the findings on unknown offenders and their unsolved crimes? In other words, do the unknown offenders contained in DNA databases differ from the unknown offenders that are not contained in DNA databases? It is difficult to answer this question. However, offenders are often generalists and it seems unlikely that all the crimes committed by the offenders registered in the police data are known. As with co-offending, a mixed known and unknown offending behaviour seems likely. This dissertation also showed that known and unknown offenders are not two totally isolated groups. Statements can be made about the difference between known and unknown offenders, and the findings about known offenders (may) change when unknown offenders are included in a study. The research demonstrated not only the difference, but also the relation, between known and unknown offenders. Known and unknown offenders do not seem to have mutually exclusive features. Taking this a step further, perhaps the most relevant question is not the one that gauges the difference between known and unknown offenders, but the one that gauges the difference between solved and unsolved crimes.

This dissertation also clearly demonstrated the (importance of the) **link between forensic science and forensic intelligence**, and between operational and strategic or scientific research. In terms of forensic science and intelligence, the research showed that DNA profiles can only be obtained at a minority of crime scenes (forensic science), leading to selectivity in the DNA database that may have an impact on what can be studied and the findings based on the DNA database (forensic intelligence). In terms of operational and strategic/scientific research, the findings from the SPPT performed in Chapter 7 of this dissertation (i.e., forensic intelligence) provided insights into Belgian DNA retention policies (i.e., forensic science) (**Figure 27**). This awareness is important for all researchers making use of forensic DNA data.

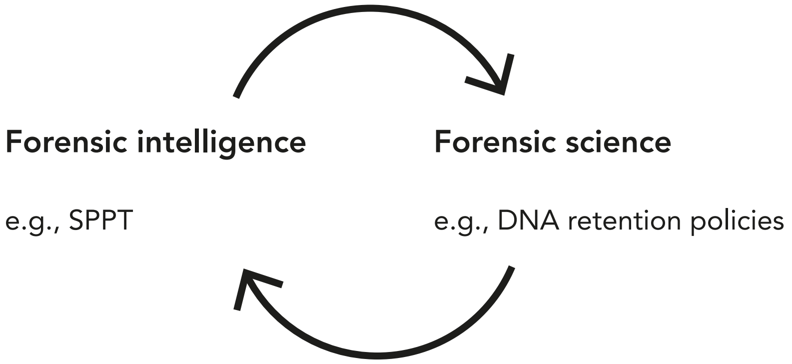


Figure 27: The relation between forensic intelligence and forensic science

13.3. Practical importance

This dissertation is also of practical importance, as guidelines for crime prevention and control can be derived from the improved insights on unknown offenders. The lower betweenness centrality of unknown offenders (Chapter 10), and offending networks having a larger spatial and temporal spread when unsolved crimes are integrated (Chapter 9), have implications for prevention and law enforcement strategies. Some specific guidelines were formulated above (see for example p. 91). However, the implications go far beyond this. The integration of police-recorded crime data and DNA data may also have an impact on how operational research can be organized. This is further discussed in Chapter 16 on the policy implications that follow from this dissertation.

14. Limitations

No study is without its limitations, and a dissertation is subject to that general rule. Limitations encountered in answering the different research questions were described in the relevant chapters. In addition, various limitations of the two data sources used in this study, namely the NGDB and the Belgian General Police Database, were discussed in Chapter 4. Conceivably, the most important limitation of both data sources is that they only include **registered crimes**. Crimes must be detected by or reported to the police before they can be registered in the NGDB and/or the Belgian General Police Database. However, this limitation is not unique to this dissertation. Any other research making use of official recorded crime data encounters this problem. On the contrary, this research was able to limit this limitation, as more offenders and their crimes became visible. Using DNA data enabled meaningful analyses to be carried out on both unsolved crimes and unknown offenders.

This dissertation also has a number of more specific limitations. First of all, this is a Belgian story, which may limit the **external validity** of the research. Almost one in three countries has an operational national DNA database (Forensic Genetics Policy Initiative) storing unknown forensic profiles extracted from biological samples found at crime scenes (e.g., a drop of blood), and/or reference profiles obtained from samples taken directly from known individuals (e.g., a buccal swab from a suspect or victim) (Home Office, 2014). Many of these countries also follow the previously mentioned European Standard Set when analysing forensic DNA samples. Although the main lines are the same, the way in which a national DNA database is interpreted can differ from country to country. Different inclusion criteria (especially regarding the reference profiles) and retention periods have an impact on the size and composition of the DNA database. Consequently, the representativeness of registered crime in a DNA database can vary from country to country. However, the same remark can be made with regard to police-recorded crime data. These databases also differ from country to country, although the external validity of police-recorded crime data is hardly questioned. Nevertheless, to meet this potential limitation of DNA databases, a clearly stated explanation of how and when DNA traces are collected at crime scenes, which profiles are (not) stored in the DNA database and under what circumstances a match can be concluded, may help to interpret and compare the findings from criminological research based on DNA data and to prevent false conclusions (De Moor et al., 2017).

A second limitation concerns not being able to involve **known offenders** stored in the NGDB in this investigation. The identity of the donor of a reference profile stored in the NGDB is managed by the National Cell. The legislator has consciously created a third authority as administrator to protect the privacy of suspects and offenders. It is therefore not possible to access these data for scientific research, and consequently it was not possible to check whether a known offender from the Belgian General Police Database and a known offender from the NGDB is one and the same person. A link between the entire

NGDB and the entire Belgian General Police Database always means an uncertainty. Therefore, the DNA dataset in this dissertation was limited to (crimes committed by) unknown offenders. Since the focus of this dissertation was on unknown offenders (and their crimes), this is not necessarily a problem. However, as links are formed between unknown offenders from the NGDB and known offenders from the Belgian General Police Database, it seems obvious that links can be made between known and unknown offenders within the NGDB. The involvement of known offenders from the NGDB would obviously have no influence on our research into the validity of the NGDB in representing unsolved crimes per se (Chapter 7), but could have supplemented, changed or reinforced the findings about (the difference between known and) unknown offenders and their crimes (Chapters 9 and 10).

A final limitation is inherent in the status of an unknown offender, namely that **the DNA database contains only limited information available about unknown offenders**. The ability to study the characteristics and behaviour of unknown offenders is limited. Except for gender, for example, no personal data are available in the DNA database. Only non-coding parts of DNA may be analysed in a forensic context. Characteristics about the behaviour of the unknown offenders, as applied in this dissertation, are derived from the information available about the crimes: the location and time of the crime. Other information, such as the age of the offender (which may be relevant in the context of studying the criminal career), is not (yet?) available. Nevertheless, the NGDB remains the only database containing information about offenders of unsolved crimes stored in the police-recorded crime data. Moreover, the 'limited' information available in the NGDB already made it possible to compare (the criminal behaviour of) unknown offenders with (the criminal behaviour of) known offenders and to study both the offenders and their behaviour in a network context.

15. Future research

As has already been mentioned, to date little criminological research has been carried out using DNA databases as a data source. Following on from this study, there are opportunities for further research both from a methodological and theoretical point of view. Some of these are specifically related to the research questions and were discussed in the relevant chapters. This section deals with future research opportunities that transcend the various research questions.

15.1. Criminal career research and SNA

As criminal career research has, to date, mainly been based on (individual) known offenders, a first research path that results from this dissertation is to further consider the possibilities of social network analysis for the study of both known and unknown offenders in criminal career research. Two improvements can be made to the research conducted here. The first is to use a broader time frame and to include all offence types present in the DNA database. Research shows that a criminal career can be of long duration, and an interruption of several years does not necessarily mean the end of the career (Metcalfe & Baker, 2014). Various circumstances can cause a temporary hiatus, for example imprisonment or illness. Another finding from the literature is that many offenders are generalists rather than specialists and therefore commit various types of crimes (Guerette, Stenius, & McGloin, 2005; Townsley et al., 2005).

Second, future criminal career research might build on the work presented here by applying a structural approach. This dissertation used the formal approach of SNA. Only the network structure is described, and not the impact the relation between offenders may have on the occurrence of crime, for example. A structural approach would enable the impact of the interdependencies among known and unknown offenders on the four key dimensions of a criminal career (i.e. prevalence, frequency, specialization, and desistance; see Piquero et al., 2003) also to be studied. More insights into the difference between the behaviour of known and unknown offenders could then be obtained.

15.2. Prüm and the international context

As mentioned when discussing the limitations of this dissertation, this is a Belgian story using Belgian data sources. However, crime is not a phenomenon that stops at national borders. It is therefore important to look beyond national borders, and in this context the earlier mentioned Prüm Convention and the transnational exchange of forensic data between Prüm member states is of great importance. The network of national DNA databases makes it possible to link an offender to different crimes, regardless of the country in which the crimes were committed. In addition, unsolved crimes committed in one country can be linked to (un)solved crimes in another country (Prainsack & Toom, 2013). This could result in breakthroughs in previously unsolved crimes such as drug cases and murder cases (NICC/INCC, 2018).

It also offers some interesting research opportunities for criminologists. For example, several authors have already studied cross-border offending based on the results of the exchange of forensic DNA data between Prüm members (see for example F. Santos & Machado, 2017; Taverne & Broeders, 2017). However, the studies are often limited to descriptive analyses. For example: which (regions of which) countries provide the most mutual matches? Again, an SNA could provide a different view of offending behaviour and whether or not there are differences between offenders who are internationally active and those who commit their offences within national borders.

15.3. Inclusion and retention criteria of DNA databases

The findings of this dissertation may stimulate the debate and scientific study on the inclusion and retention criteria of DNA databases. Many countries explicitly state in legislation which reference profiles are to be stored in their national DNA database, and for how long. For crime scene profiles, this is often limited to the retention period and no other retention criteria. But how do authorities determine whose DNA profile should be stored in their DNA database, and for how long? After all, there is a broad continuum between 'no DNA database' and a 'universal DNA database' (**Figure 28**).

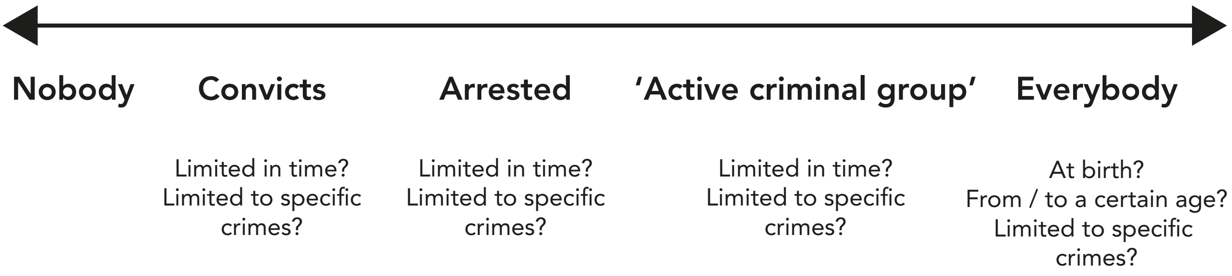


Figure 28: Spectrum of possible target groups stored in DNA databases

In order to decide which DNA profiles should be stored, the purposes of such databases must be considered. The primary aim of a DNA database is usually to solve and prevent crimes, taking into account the recidivism of offenders, as is the case in Belgium (Belgische Kamer van volksvertegenwoordigers, 1998). Research into, among other things, recidivism, criminal careers and the age–crime curve often provide guidance in determining (and extending) the inclusion conditions of a DNA database (Kazemian et al., 2011; Tseloni & Pease, 2011).[43]

Although research has already been conducted in this area, this dissertation may stimulate further the

---

[43] Although the 'elimination' and 'missing persons' databases have other purposes, a similar discussion can be held on the content of these DNA databases. Establishing a universal DNA database could also imply that both of these databases will become redundant.

debate on which type of DNA profiles should be stored and for how long, both reference profiles and forensic profiles. First, current research into recidivism, criminal careers and the age–crime curve is mainly based on known offenders and their crimes. Kazemian et al. (2011), for example, used data from convicted offenders participating in the Cambridge Study in Delinquent Development. Tseloni and Pease (2011) also only used known (i.e., arrested) offenders to evaluate DNA retention policies. However, a difference in criminal behaviour between known and unknown offenders can shed a light on inclusion and retention policies. Second, current research only focuses on reference profiles. This dissertation has demonstrated the importance of crime scene profiles, for both operational and strategic research. Unknown offenders must therefore be included together with known offenders when assessing DNA retention policies of both reference profiles and crime scene profiles. This future research track is therefore in line with the future research track on criminal careers discussed above.

15.4. The impact of DNA policies and DNA databases

Several authors have highlighted the lack of systematic empirical research into the actual impact or effectiveness of forensic DNA (Murphy, 2018; F. Santos et al., 2013; Wilson, McClure, & Weisburd, 2010), although they acknowledge that such research is conceptually and methodologically complex (F. Santos et al., 2013). Indeed, there are different ways to interpret and evaluate the utility or impact of DNA databases. Once again, when evaluating the impact of DNA databases it is important to keep the purpose of the database in mind. In addition to the match rate, the deterrence effect on convicted offenders and the impact on the general crime level, the utility of DNA databases can also be assessed by their impact on public safety or by their cost efficiency (Bieber, 2006; p. 230; Doleac, 2017).

Currently, no utility research evaluating the NGDB has been conducted.[44] In addition, it is not easy to generalize research results from foreign studies as the different inclusion and retention criteria can limit their external validity. Moreover, existing utility studies only take into account the DNA database. This dissertation showed that the link between the NGDB and the Belgian General Police Database can identify links between unknown and known offenders, so that the NGDB can also indirectly contribute to crime solving. In addition to the match rate, this indirect impact must also be taken into account when evaluating the utility of the NGDB. Research specifically focused on the Belgian context is therefore urgently needed. New insights obtained from criminological research based on an integrated dataset of both the NGDB and the Belgian General Police Database could add weight to arguments about the effectiveness and efficiency of the NGDB for crime prevention and law enforcement (De Moor et al., 2018b). In other words: *"In an environment of limited resources, knowing which technologies provide the biggest benefit for solving crimes provides policy-makers with knowledge to support the expansion of certain*

---

[44] The research by Renard and Jeuniaux (2012) was limited to an evaluation of the court costs for DNA research over 11 years (2000–2010).

practices or the reduction or elimination of others, as well as providing a base of information for criminal investigators weighing the opportunity costs of one choice over another" (Wilson et al., 2010, p. 468).

16. Policy implications

A number of policy implications arise from this dissertation. As the selectivity and representativity of the DNA data are not important for operational research, all the implications relate to identifying unknown offenders and endorsing the operational research. **Figure 29** schematically represents the three proposed policy implications.

16.1. Reinforce information exchange ANG and trace databases

The Belgian government recognizes the importance of information exchange between judicial and administrative services. The Integral Security Framework Document (ISFD) for closer cooperation between the police and other government and security services, and the National Security Plan (NSP), the strategic policy plan of the integrated police, define ten security phenomena and seven cross-cutting themes to achieve an effective safety strategy in Belgium. One of the cross-cutting themes is enforcing administrative law intensively as part of the approach to tackle crime, while ensuring an efficient exchange of information with the various stakeholders (Federale Politie, 2016; Geens & Jambon, 2016). These stakeholders can be administrative enforcement bodies such as tax administration, or private actors such as energy companies, insurance companies or car repair services.

It is therefore surprising that the information exchange between the NGDB and the Belgian General Police Database is so limited. In the context of operational research, no one has used an integrated version of the two databases, as applied in this dissertation. The NGDB and the Belgian General Police Database are two separate databases and each has its own manager (respectively the DIS service of the NICC / INCC, and the DRI of the federal police), and access to the stored data is legally protected. This dissertation showed that by merging crime data from the NGDB with crime data from the Belgian General Police Database, links can be identified between offenders and between crime that are not visible in each of the individual databases. The integration of the unknown offenders stored in the NGDB makes it possible to make links visible between unknown offenders / unsolved crimes, and also between unknown offenders / unsolved crimes with known offenders / solved crimes. The latter gets closer to an offender, without knowing the identity of the offender (yet). From this perspective, Jeuniaux et al. (2016) studied co-offending networks present in the NGDB. They acknowledged the operational use of network intelligence from both the forensic profiles and the reference profiles, as the likelihood of identifying unknown offenders is increased when they are linked to known offenders. The authors only used the NGDB as the data source for their research, so not all information available within the police and judiciary was fully utilized. Other academics have also advocated the integration of different (trace) data sources to enhance the forensic intelligence process (Legrand & Vogel, 2014; Ribaux et al., 2003).

In other words, strengthening the information exchange between the NGDB (or other trace databases) and the Belgian General Police Database can only enhance operational research. The terms in which this information exchange should ideally take place should be examined. The advantages and disadvantages of maintaining the databases or integrating the two databases into a single database must be weighed up. Consideration should also be given to the management of the databases. For example, the Fichier National Automatisé des Empreintes Génétiques (FNAEG), the French national DNA database, is managed by a sub-directorate of the technical and scientific departments of the French police force. French police officers can compare DNA profiles in the database (Ribbers, 2015). In the Netherlands, a special unit of the Dutch police receives weekly DNA-based network data from the Netherlands Forensic Institute. After adding further intelligence, for example fingerprints, the information is provided to all police forces in the Netherlands. As a result, connections between cases from different police forces can be made (Jeuniaux et al., 2016).

The current Belgian legislation will need to be to be adapted if an integrated database is to be used for operational police investigations, as different rules and laws apply to each database. The NGDB is managed by the NICC/INCC, and the National DNA Cell manages the DNA code numbers of the reference profiles – two separate agencies to protect the privacy of individuals whose DNA is stored in the NGDB. Access to the Belgian General Police Database is also legally restricted. The ministerial guideline MFO-3 describes in detail the modalities for access to and consultation of databases accessible to members of the police services.[45] Moreover, the protection of the privacy of those involved must also be taken into account when intensifying the exchange of information between the two different databases. This obviously has implications for the extent to which the police have access to the stored DNA data. For example, not all information stored in the NGDB is useful for operational research. Police officers, for example, should not have access to the DNA profiles, the actual genetic information.

16.2. Link forensic science and forensic intelligence

Intensifying the exchange of information, whether or not this results in a single database, raises three other issues. First, the integration of information from various databases implies making full use of the available data. This means that the raw data is used to extract information and knowledge. Second, an intensification of information exchange implies that the clear distinction between purely operational research and purely strategic research is weakened. Third, the exchange of information implies that the way in which the NGDB (as well as the Belgian General Police Database and the other traces databases) is fed needs to be re-examined. These three elements are inextricably linked.

---

[45] Gemeenschappelijke richtlijn MFO-3 van de Ministers van Justitie en van Binnenlandse Zaken betreffende het informatiebeheer inzake gerechtelijke en bestuurlijke politie, BS 14 juni 2002

Comparative DNA research in Belgium only focuses on the direct or indirect identification of a person involved in a crime, or on proving his innocence (Belgisch Staatsblad, 2011a). If the NGDB is to be optimally utilized, one should also consider the possibilities of the NGDB regarding orientation (e.g., mapping out the broader environment of a known or an unknown offender). Since the police currently have no access to the data concerning unknown offenders and their crimes that is stored in the NGDB, much information is lost. Indeed, this dissertation highlights the importance of integrating unknown offenders and unsolved crimes in network research. Orientation is best expressed when all available information is used, not only within a specific case but also in the context of forensic intelligence. After all, a thorough knowledge of crime phenomena or offender groups promotes operational research. The strict distinction between operational (case-based) and strategic research or forensic intelligence could disappear, and the importance and mutual influence of both dimensions should be recognized. This of course also requires an intelligence approach when the information is collected and registered in both the NGDB and other trace databases and the Belgian General Police Database. Specifically with regard to the trace databases, this coincides with a shift from crime scene investigation (CSI) to crime scene reconstruction (CSR). When carrying out CSR the central question is not what DNA traces are present, but what may have happened. Different scenarios are considered, and these guide the search for traces. Being able to rely on forensic intelligence is of course essential here. Within CSR, for example, DNA traces of the victim are also important to be able to reconstruct the crime.

16.3. Awareness and education

Magistrates who cannot estimate the cost of a DNA analysis, or reports by DNA experts that are difficult for the police and judges to interpret (de Keijser, Malsch, Luining, Weulen Kranenbarg, & Lenssen, 2016; Malsch, de Keijser, Luining, Weulen Kranenbarg, & Lenssen, 2016; Malsch et al., 2013), are just two examples that illustrate the need for information and education. But the need goes beyond this. Before there can be any question of developing thorough forensic intelligence and strengthening the exchange of information between the various databases, it is necessary to raise awareness of the importance of forensic trace research and appropriate training for all the actors involved. This applies both to the police and the magistracy. Actors within the police and judiciary must be informed about the possibilities of forensic science and forensic intelligence: what is possible with forensic DNA, but also, and perhaps especially, what is not. On the one hand, people in the field set high expectations in terms of DNA as forensic science to solve a specific crime, but on the other hand they do not or insufficiently acknowledge the possibilities for forensic intelligence (Renard et al., 2018). For example, the NGDB must, as it were, be regarded as the ‘LinkedIn for the criminal career’. People use LinkedIn, the online social network for professionals, to get in touch with others via their own network: a friend brings you into contact with the CEO of a company where you want to work, for example. Getting in touch with the CEO is much easier if you know his friends. So both the first and second line contact are of high value. The same principle can be applied for criminal networks: approaching unknown offenders and getting to know them becomes easier when they can be approached through known co-offenders.

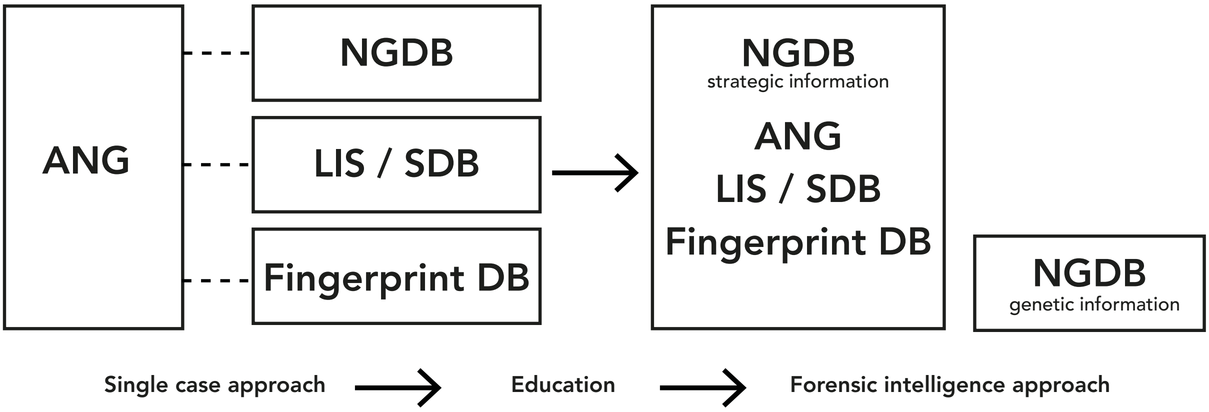


Figure 29: Schematic representation of policy implications

17. The ethics of forensic DNA and DNA databases

The many fascinating ‘cold hits’ and ‘exonerations’ cases that have been reported (see for example Hampikian, West, & Akselrod, 2011; Wade, 2015) encourage a belief in the importance and usefulness of forensic DNA and DNA databases, among both policy makers and the general public. This is often accompanied by a demand for more extensive DNA inclusion criteria and retention policies. However, forensic DNA databases are not free of controversy. In a dissertation that uses such a database, some attention should be paid to this. The debate goes far beyond what is described below, but this brief explanation begins to place DNA databases in a broader context of rights and freedoms.

The fault line between the defenders and the opponents of (extensive) DNA databases often coincides with political and judicial representatives on the one side, and academics on the other. This is an exaggerated black-and-white position, but actors within politics and justice often emphasize the benefits of DNA databases in the fight against and prevention of crime, while the academic world raises critical questions about such databases (Machado, Silva, & Cunha, 2012). In addition to the many proponents of DNA databases, critics also make their voices heard (Levitt, 2007; Wallace, 2006; Williams & Johnson, 2004, 2005). Critics express their concern that an extension of DNA databases to include more reference profiles would be an example of excessive government control (Gamero et al., 2008; Williams & Johnson, 2005). Moreover, it would create a list of suspects in a way that is discriminatory (GeneWatch UK, 2006; McCartney, 2004; Wallace, 2006) and violates the presumption of innocence (De Gorgey, 1990; Machado et al., 2012). Thus (extended) DNA databases would pose a threat to human rights, in particular “*liberty, autonomy, privacy, informed consent, moral and physical integrity and the presumption of innocence*” (Machado & Silva, 2014; p.133), which are anchored in Article 3 (prohibition of torture), Article 6 (the right to a fair trial) and Article 8 (the right to respect for privacy) of the European Convention on Human Rights (ECHR). The (long-term) storage of forensic profiles is also not without controversy. In addition, it is not always possible to identify whether the DNA profile is of an offender, a casual witness or a victim.

It is striking, however, that fingerprint databases generate much less controversy than DNA databases, even though fingerprints are often taken from broader target groups. In Belgium, for example, fingerprints are part of the so-called triptych of an arrestee, the other items being images and a description. Fingerprints are taken automatically from every arrestee, while a DNA reference profile of a suspect is only taken at the request of the public prosecutor. Both databases contain biometric information obtained from reference profiles or crime scene profiles, which leads to the individualization of a person (Stevens, 2001). In contrast to a fingerprint, DNA also contains hereditary information from which all kinds of characteristics of the donor can be derived, but forensic DNA analysis is limited to the non-coding part of the DNA. A forensic DNA profile thus contains no more information than a fingerprint (Kaye & Smith, 2003). Moreover, the implication of storing information is the same for both databases: the donor is considered to be a potential offender of (future) crimes.

So it seems that the controversy is not so much about what one actually does with the biometric information, but what one *could* do with it. Indeed, although DNA laboratories only examine 'junk DNA' to draw up a DNA profile, some speculate on what would happen if, in the future, this junk DNA actually turned out to be coding DNA (Hochschild & Sen, 2012). Scientific developments also create pressure on the social, legal and ethical aspects of these innovations. Wienroth, Morling, and Williams (2014) have described four successive waves of technological innovation in forensic genetics: the use of DNA as evidence; the creation of DNA databases; analysis of the coding part of the DNA; and the search for genetic similarity in criminal cases. So-called *predictive techniques* and *phenotypic DNA testing*, where personal characteristics or medical data of a potential offender are drawn up only on the basis of a forensic DNA profile, do not seem very far off (Staley, 2005). In Belgium, KULeuven is already conducting research into visualizing a human face only on the basis of DNA (see for example Claes et al., 2014).

Will these developments lead to coded DNA also being analysed in the future, if their usefulness in the fight against crime is recognized? Will the samples also be stored (as they are in the UK), in addition to the profiles, so that they can be re-analysed later when techniques have evolved (Rothstein & Talbott, 2006) or so that they can be checked for errors (Wallace, 2006)? Other developments, such as rapid forensic DNA analysis where a DNA profile can be produced from a sample in just a few hours (Hopwood et al., 2010), also seem to pave the way for an increase in DNA analysis. In the United States, for example, the Rapid DNA Act was signed into law in 2017, making it possible to carry out real-time searches of the national database using rapidly generated DNA profiles outside of the laboratory context (Murphy, 2018).

A specific recommendation is to open the debate about databases to all biometric information. Such debate should consider both the differences and the similarities between the various types of bio-metric information and should differentiate between the use of the information for forensic intelligence and for forensic science. The positive and useful aspects of DNA should not be lost because of the dangers of abuse. Increasing public awareness will be central to conducting a useful and well-founded debate.

REFERENCES

Aebi, M. F., & Linde, A. (2010). Is There a Crime Drop in Western Europe? *European Journal on Criminal Policy and Research, 16*(4), 251-277. doi:10.1007/s10610-010-9130-y

Alarid, L. F., Burton Jr., V. S., & Hochstetler, A. L. (2009). Group and solo robberies: do accomplices shape criminal form? *Journal of Criminal Justice, 37*, 1-9.

Almog, J. (2014). Forensics as a proactive science. *Science and Justice, 54*, 325-326.

Andresen, M. A. (2009). Testing for similarity in area-based spatial patterns: A nonparametric Monte Carlo approach. *Applied Geography, 29*(3), 333-345. doi:10.1016/j.apgeog.2008.12.004

Andresen, M. A. (2016). An area-based nonparametric spatial point pattern test: The test, its applications, and the future. *Methodological Innovations, 9*, 205979911663065. doi:10.1177/2059799116630659

Andresen, M. A., & Felson, M. (2010). The impact of co-offending. *British Journal of Criminology, 50*, 66-81.

Andresen, M. A., & Linning, S. J. (2012). The (in)appropriateness of aggregating across crime types. *Applied Geography, 35*(1-2), 275-282. doi:10.1016/j.apgeog.2012.07.007

Andresen, M. A., Linning, S. J., & Malleson, N. (2017). Crime at Places and Spatial Concentrations: Exploring the Spatial Stability of Property Crime in Vancouver BC, 2003–2013. *Journal of Quantitative Criminology, 33*, 255-275. doi:10.1007/s10940-016-9295-8

Andresen, M. A., & Malleson, N. (2011). Testing the Stability of Crime Patterns: Implications for Theory and Policy. *Journal of Research in Crime and Delinquency, 48*(1), 58-82. doi:10.1177/0022427810384136

Babinski, L. M., Hartsough, C. S., & Lambert, N. M. (2001). A comparison of self-report of criminal involvement and official arrest record. *Aggressive Behavior, 27*, 44-54.

Balding, D. (1999). When can a DNA profile be regarded as unique? *Science & Justice, 39*, 257-260.

Baranowski, A. M., Burkhardt, A., Czernik, E., & Hecht, H. (2017). The CSI-education effect: Do potential criminals benefit from forensic TV series? *International Journal of Law, Crime and Justice, in press.* doi:10.1016/j.ijlcj.2017.10.001

Barash, M., Reshef, A., & Brauner, P. (2010). The use of adhesive tape for recovery of DNA from crime scene items. *Journal of Forensic Science, 55*(4), 1058-1064.

Bates, D., & Maechler, M. (2017). Matrix: Sparse and Dense Matrix Classes and Methods. R package version 1.2-12.

Beauregard, E., & Bouchard, M. (2010). Cleaning up your act: Forensic awareness as a detection avoidance strategy. *Journal of Criminal Justice, 38*(6), 1160-1166.

Beaver, K. M. (2014). The promises and pitfalls of forensic evidence in unsolved crimes. *Criminology & Public Policy, 9*(2), 405-410.

Belgisch Staatsblad. (1999). Wet van 22 maart 1999 betreffende de identificatieprocedure via DNA-analyse in strafzaken.

Belgisch Staatsblad. (2011a). Wet van 7 november 2011 houdende wijziging van het Wetboek van strafvordering en van de wet van 22 maart 1999 betreffende de identificatieprocedure via DNA onderzoek in strafzaken.

Belgisch Staatsblad. (2011b). Wetsontwerp houdende wijziging van het Wetboek van strafvordering en van de wet van 22 maart 1999 betreffende de identificatieprocedure via DNA-onderzoek in strafzaken.

Belgisch Staatsblad. (2013a). Omzendbrief DNA: identificatieprocedure via DNA-onderzoek in stafzaken.

Belgisch Staatsblad. (2013b). Wet tot wijziging van het Wetboek van strafvordering en van de wet van 22 maart 1999 betreffende de identificatieprocedure via DNA-onderzoek in strafzaken, met het oog op de oprichting van een DNA-gegevensbank " Vermiste personen ".

Belgisch Staatsblad. (2015). Koninklijk besluit houdende uitvoering van artikel 6 van de programmawet (II) van 27 december 2006 tot bepaling van de tarieven in strafzaken voor het deskundigenonderzoek in genetische analyse in opdracht van een gerechtelijke overheid.

Belgisch Staatsblad. (2017a). Koninklijk besluit van 11 augustus 2017 tot integratie in het federaal parket van de nationale cel, bedoeld in de wet van 22 maart 1999 betreffende de identificatieprocedure via DNA-onderzoek in strafzaken.

Belgisch Staatsblad. (2017b). Omzendbrief DNA: Identificatieprocedure via dna-onderzoek in strafzaken, Addendum IV – contactpersonen.

Belgisch Staatsblad. (2017c). Wet van 9 april 2017 tot wijziging van de wet van 21 december 2013 betreffende de oprichting van een DNA-gegevensdatabank "Vermiste Personen".

Belgisch Staatsblad. (2017d). Wet van 17 mei 2017 tot wijziging van het Wetboek van strafvordering en van de wet van 22 maart 1999 betreffende de identificatie-procedure via DNA-onderzoek in strafzaken, met het oog op de oprichting van een DNA-gegevensbank « Intervenanten ».

Belgische Kamer van volksvertegenwoordigers. (1998). Wetsvoorstel betreffende de identificatieprocedure via genetische analyses in strafzaken. *1047/2, 26*, 1-29.

Berlusconi, G. (2013). Do all the pieces matter? Assessing the reliability of law enforcement data sources for the network analysis of wire taps. *Global Crime, 14*(1), 61-81. doi:10.1080/17440572.2012.746940

Berlusconi, G. (2017). Social Network Analysis and Crime Prevention. In B. LeClerc & E. U. Savona (Eds.), *Crime Prevention in the 21st Century: Insightful Approaches for Crime Prevention Initiatives* (pp. 129-141). Cham: Springer International Publishing.

Bernasco, W. (2006). Co-offending and the choice of target areas in burglary. *Journal of Investigative Psychology and Offender Profiling, 3*, 139–155.

Bernasco, W. (2008). Them Again? Same-offender involvement in repeat and near repeat burglaries. *European Journal of Criminology, 5*(4), 411-431.

Bernasco, W. (2014). *A DNA cartography of the Prüm hits between the Netherlands and the EU.* Paper presented at the PIES workshop, Brussels.

Bernasco, W., & Steenbeek, W. (2017). More Places than Crimes: Implications for Evaluating the Law of Crime Concentration at Place. *Journal of Quantitative Criminology, 33*(3), 451-467. doi:10.1007/s10940-016-9324-7

Biderman, A. D., & Reiss, A. J. (1967). On Exploring the “Dark Figure” of Crime. *The Annals of the American Academy, 374*(1), 1-15.

Bieber, F. R. (2006). Turning Base Hits into Earned Runs: Improving the Effectiveness of Forensic DNA Data Bank Programs. *The Journal of Law, Medicine & Ethics, 34*(2), 222-233.

Bloor, M., & Wood, F. (2006). *Keywords in Qualitative Methods. A Vocabulary of Research Concepts*. London: Sage.

Blumstein, A., & Cohen, J. (1987). Characterizing criminal careers. *Science, 237*, 985-991.

Bond, J. W. (2007). Value of DNA evidence in detecting crime. *Journal of Forensic Science, 52*(1), 128-136.

Borgatti, S. P., Carley, K. M., & Krackhardt, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social Networks, 28*(2), 124-136. doi:10.1016/j.socnet.2005.05.001

Boslaugh, S. E. (2010). Secondary data source. In N. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 1330-1331). Thousand Oaks, California: SAGE Publications Ltd.

Bouchard, M., & Malm, A. (2016). Social Network Analysis and Its Contribution to Research on Crime and Criminal Justice. *Oxford Handbooks Online*. Retrieved from www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935383.001.0001/oxfordhb-9780199935383-e-21 doi:10.1093/oxfordhb/9780199935383.013.21

Bramley, B. (2009). DNA databases. In J. Fraser & R. Williams (Eds.), *Handbook of forensic science* (pp. 309-336). Devon: Willian Publishing.

Brantingham, P. L., & Brantingham, P. J. (1993). Nodes, paths and edges: considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology, 13*, 3-28.

Bright, D. A., Hughes, C. E., & Chalmers, J. (2012). Illuminating dark networks: a social network analysis of an Australian drug trafficking syndicate. *Crime, Law and Social Change, 57*(2), 151-176. doi:10.1007/s10611-011-9336-z

Bright, J.-A., Curran, J. M., & Buckleton, J. S. (2014). The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation. *Forensic Science International: Genetics, 12*, 208-214.

Brooks, L. W. (2015). Police Discretionary Behavior. In R. Dunham & G. P. Alpert (Eds.), *Critical issues in policing: contemporary readings* (7th ed., pp. 122-142). Prospect Heights (Ill.): Waveland press.

Burcher, M., & Whelan, C. (2015). Social network analysis and small group ‘dark’ networks: an analysis of the London bombers and the problem of ‘fuzzy’ boundaries. *Global Crime, 16*(2), 104-122. doi:10.1080/17440572.2015.1005363

Burrows, J., & Tarling, R. (2004). Measuring the impact of forensic science in detecting burglary and autocrime offences. *Science and Justice, 44*(4), 217-222.

Butler, J. M. (2010). *Fundamentals of forensic DNA typing*. Amsterdam Academic Press/Elsevier.

Campana, P., & Varese, F. (2011). Listening to the wire: criteria and techniques for the quantitative analysis of phone intercepts. *Trends in Organized Crime, 15*(1), 13-30. doi:10.1007/s12117-011-9131-3

Campbell, R., Pierce, S. J., Sharma, D. B., Feeney, H., & Fehler-Cabral, G. (2016). Should Rape Kit Testing Be Prioritized by Victim-Offender Relationship? *Criminology & Public Policy, 15*(2), 555-583. doi:10.1111/1745-9133.12205

Cantor, D., & Lynch, J. P. (2000). Self-report surveys as measures of crime and criminal victimization.

Criminal Justice, 4, 85-138.

Carcach, C. (1997). Reporting crime to the police. *Trends & issues in crime and criminal justice, 68*, 1-6.

Carrington, P. J. (2002). Sex homogeneity in co-offending groups. In J. Hagberg (Ed.), *Contributions to Social Network Analysis, Information Theory and Other Topics in Statistics. A Festschrift in Honour of Ove Frank.* (pp. 101-116). Stockholm: Stockholm University.

Carrington, P. J. (2014). Co-offending. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of Criminology and Criminal Justice* (pp. 548-558). New York: Springer.

Chattoe, E., & Hamill, H. (2005). It's Not Who You Know—It's What You Know About People You Don't Know That Counts. Extending the Analysis of Crime Groups as Social Networks. *British Journal of Criminology, 45*(6), 860-876.

Cheatwood, D., & Block, K. J. (1980). Youth and homicide: an investigation of the age factor in criminal homicide. *Justice Quarterly, 7*(2), 265-292.

Claes, P., Liberton, D. K., Daniels, K., Matthes Rosana, K., Quillen, E. E., Pearson, L. N., . . . Shriver, M. D. (2014). Modeling 3D Facial Shape from DNA. *PLOS Genetics, 10*(3), 1-14.

Coleman, C., & Moynihan, J. (1996). *Understanding crime data: haunted by the dark figure*. Buckingham Open university press.

Coles, N. (2001). It's Not What You Know—It's Who You Know That Counts. Analysing Serious Crime Groups as Social Networks. *British Journal of Criminology, 41*(4), 580-594.

Cope, N. (2004). Intelligence led policing or policing led intelligence? Integrating volume crime analysis into policing. *British Journal of Criminology, 44*, 188-203.

Costello, A., & Wiles, P. (2001). GIS and the journey to crime. An analysis of patterns in South Yorkshire. In A. Hirschfield & K. Bowers (Eds.), *Mapping and analysing crime data. Lessons from research and practice* (pp. 27-60). London: Taylor & Francis.

Costenbader, E., & Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Social Networks, 25*(4), 283-307. doi:10.1016/s0378-8733(03)00012-1

Council of Europe. (1992). Recommendation No. R (92) 1 of the committee of ministers to member states on the use of analysis of deoxyribonucleic acid (DNA) within the framework of the criminal justice system.

Council of the European Union. (2005). Prüm Convention. 1-44.

Council of the European Union. (2008a). Council Decision 2008/615/JHA of 23 June 2008 on the stepping up of cross-border cooperation, particularly in combating terrorism and cross-border crime. *OJ L, 210*, 1-11.

Council of the European Union. (2008b). Council Decision 2008/616/JHA of 23 June 2008 on the implementation of Decision 2008/615/JHA on the stepping up of cross-border cooperation, particularly in combating terrorism and cross-border crime. *OJ L, 210/12*, 12-72.

Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems, 1695*(5).

Dahl, J. Y. (2010). DNA the Nor-way: black-boxing the evidence and monopolising the key. In R. Hindmarsh & B. Prainsack (Eds.), *Genetic Suspects: Global Governance of Forensic DNA Profiling and Databasing* (pp. 197-217). Cambridge: Cambridge University Press.

Davies, J. B., & Hoy, M. (1994). *Making Inequality Comparisons When Lorenz Curves Intersect*. Department of Economics Research Reports. London.

De Gorgey, A. (1990). The advent of DNA databanks: implications for information privacy. *American Journal of Law and Medicine, 16*(3), 381-398.

de Keijser, J. W., Malsch, M., Luining, E. T., Weulen Kranenbarg, M., & Lenssen, D. J. H. M. (2016). Differential reporting of mixed DNA profiles and its impact on jurists' evaluation of evidence. An international analysis. *Forensic Science International: Genetics, 23*, 71-82. doi:10.1016/j.fsigen.2016.03.006

de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). *The international handbook of survey methodology*. New York: Taylor & France Group.

de Melo, S. N., Matias, L. F., & Andresen, M. A. (2015). Crime concentrations and similarities in spatial crime patterns in a Brazilian context. *Applied Geography, 62*, 314-324. doi:10.1016/j.apgeog.2015.05.012

De Moor, S., Vander Beken, T., & Van Daele, S. (2017). DNA Databases as Alternative Data Sources for Criminological Research. *European Journal on Criminal Policy and Research, 23*(2), 175-192. doi:10.1007/s10610-016-9327-9

De Moor, S., Vandeviver, C., & Vander Beken, T. (2018a). Are DNA data a valid source to study the spatial behavior of unknown offenders? *Science & Justice, 58*(5), 315-322. doi:doi.org/10.1016/j.scijus.2018.04.003

De Moor, S., Vandeviver, C., & Vander Beken, T. (2018b). Integrating police recorded crime data and DNA data to study serial co-offending behaviour. *European Journal of Criminology, 15*(5), 632-651. doi:10.1177/1477370817749499

De Wree, E., Vermeulen, G., & Christiaens, J. (2006). *(Strafbare) overlast door jongerengroepen in het kader van openbaar vervoer*. Antwerpen: Maklu.

Develtere, R. (2014). *Vingersporen versus DNA*. Paper presented at the DNA Symposium, Brussels.

Devroe, E., Beyens, K., & Enhus, E. (Eds.). (2006). *Zwart op wit? Duiding van cijfers over onveiligheid en strafrechtsbedeling in België. Handboek criminografische basisinformatie*. Brussel: VUBPRESS.

Dillman, D. A. (1991). The design and administration of mail surveys. *Annual Review of Sociology, 17*, 225-249.

Dillman, D. A. (2000). *Mail and internet surveys: the tailored design method* (2nd ed.). London: Wiley.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail and mixed-mode surveys. The tailored design method* (4th ed.). New York: Wiley

Doleac, J. L. (2016). *The effects of DNA databases on crime*. Retrieved from http://dx.doi.org/10.2139/ssrn.2556948

Doleac, J. L. (2017). The Effects of DNA Databases on Crime. *American Economic Journal: Applied Economics, 9*(1), 165–201.

Duboccage, L. (2008). Nationale DNA databanken. In P. Boel, V. De Cloet, J. De Kinder, J. Mahieu, & D. Van Varenbergh (Eds.), *Handboek Forensisch Onderzoek. De mogelijkheden van het forensisch onderzoek* (pp. 157-162). Brussel: Politeia.

DuRant, R. H., Cadenhead, C., Pendergrast, R. A., Slavens, G., & Linder, C. W. (1994). Factors associated with the use of violence among urban black adolescents. *American Journal of Public Health, 84*(4), 612-617.

Eck, J. E., Lee, Y., O, S., & Martinez, N. (2017). Compared to what? Estimating the relative concentration of crime at places using systematic and other reviews. *Crime Science, 6*(8). doi:10.1186/s40163-017-0070-4

Eck, J. E., & Weisburd, D. (1995). Crime places in crime theory. *Crime and place, crime prevention studies, 4*, 1-33.

Everson, S. (2003). Repeat victimisation and prolific offending: chance or choice? *International Journal of Police Science & Management, 5*(3), 180-194.

Farrell, G., Tilley, N., & Tseloni, A. (2014). Why the Crime Drop? *Crime and Justice, 43*(1), 421-490. doi:10.1086/678081

Farrington, D. P. (2001). *What has been learned from self-reports about criminal careers and the causes of offending?* Retrieved from https://www.researchgate.net/publication/255603763_What_Has_Been_Learned_from_Self-Reports_About_Criminal_Careers_And_the_Causes_of_Offending

Farrington, D. P., Coid, J. W., Harnett, L. M., Jolliffe, D., Soteriou, N., Turner, R. E., & West, D. J. (2006). *Criminal careers up to age 50 and life success up to age 48: new findings from the Cambridge Study in Delinquent Development*. Retrieved from www.crim.cam.ac.uk/people/academic_research/david_farrington/hors299.pdf

Farrington, D. P., Jolliffe, D., Hawkins, J. D., Catalano, R. F., Hill, K. G., & Kosterman, R. (2003). Comparing delinquency careers in court records and self-reports. *Criminology, 41*(3), 933-958.

Farrington, D. P., & Ttofi, M. M. (2014). Criminal careers in self-reports compared with official records. *Criminal Behaviour and Mental Health, 24*(4), 225-228.

FBI. (2017). Crime in the U.S. 2016. *Uniform Crime Report*. Retrieved from https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016

Federale Politie. (2016). *Nationaal Veiligheidsplan 2016-2019*. Retrieved from https://www.politie.be/5998/nl/over-ons/geintegreerde-politie/het-nationaal-veiligheidsplan

Felson, M. (2003). The process of co-offending. *Crime Prevention Studies, 16*, 149-167.

Ferwerda, H. (2013). Serieplegers. In A. van Wijk (Ed.), *Serieplegers. Kenmerken, achtergronden en opsporing* (pp. 9-14). Den Haag: Boom Lemma Uitgevers.

Forensic Genetics Policy Initiative. (2014). About us. Retrieved from http://dnapolicyinitiative.org/

Fox, C. (2010). Developing an offender problem profile. *Safer Communities, 9*(3), 17-27.

Francis, B., Humphreys, L., Kirby, S., & Soothill, K. (2013). *Understanding criminal careers in organised crime*. Retrieved from https://www.gov.uk/government/publications/understanding-criminal-careers-in-organised-crime

Frantz, T. L., Cataldo, M., & Carley, K. M. (2009). Robustness of centrality measures under uncertainty: Examining the role of network topology. *Computational and Mathematical Organization Theory, 15*(4), 303-328. doi:10.1007/s10588-009-9063-5

Freeman, L. F. (1979). Centrality in Social Networks: Conceptual Clarification. *Social Networks, 1*, 215-239.

Fuxman Bass, J. I., Diallo, A., Nelson, J., Soto, J. M., Myers, C. L., & Walhout, A. J. (2013). Using networks to measure similarity between genes: association index selection. *Nat Methods, 10*(12), 1169-1176.

doi:10.1038/nmeth.2728

Galaskiewicz, J. (1991). Estimating point centrality using different network sampling techniques. *Social Networks, 13*(4), 347-386. doi:https://doi.org/10.1016/0378-8733(91)90002-B

Gamero, J.-J., Romero, J.-L., Peralta, J.-L., Corte-Real, F., Guillén, M., & Anjos, M.-J. (2008). A study of Spanish attitudes regarding the custody and use of forensic DNA databases. *Forensic Science International: Genetics, 2*(2), 138-149.

Gartner, R., & Macmillan, R. (1995). The effect of victim-offender relationship on reporting crimes of violence against women. *Canadian Journal of Criminology, 393*, 393-429.

Geens, K., & Jambon, J. (2016). *Kadernota integrale veiligheid 2016-2019*. Retrieved from https://www.politie.be/5998/nl/over-ons/geintegreerde-politie/het-nationaal-veiligheidsplan

Genetics Home Reference. (2017). What is DNA?   Retrieved from https://ghr.nlm.nih.gov/primer/basics/dna

GeneWatch UK. (2006). The DNA expansion programme: reporting real achievement? *GeneWatch UK Briefing*.

Gideon, L. (Ed.) (2012). *Handbook of survey methodology for the social sciences*. New York: Springer

Gilleir, F. (2013). *Discretionaire ruimte bij de Belgische lokale politie. Een explorerend, kwalitatief onderzoek naar uitvoeringswerk in de frontlijn* (Vol. 8). Antwerpen-Apeldoorn: Maklu.

Goethals, J., Ponsaers, P., Beyens, K., Pauwels, L., & Devroe, E. (2002). Criminografisch onderzoek in België. In K. Beyens, J. Goethals, P. Ponsaers, & G. Vervaeke (Eds.), *Criminologie in actie. Handboek criminografisch onderzoek* (pp. 137-188). Brussel: Politeia.

Golafshani, N. (2003). Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report, 8*(4), 597-607.

Gold, R. L. (1958). Roles in Sociological Field Observations. *Social Forces, 36*(3), 217-223. doi:10.2307/2573808

Goldweber, A., Dmitrieva, J., Cauffman, E., Piquero, A. R., & Steinberg, L. (2011). The development of criminal style in adolescence and young adulthood: separating the lemmings from the loners. *Journal of Youth and Adolescence, 40*, 332–346.

Goodwin, W., Linacre, A., & Hadi, S. (2007). *An Introduction to Forensic Genetics*. West Sussex: John Wiley And Sons Ltd.

Gottfredson, M. R., & Hindelang, M. J. (1977). A consideration of telescoping and memory decay biases in victimization surveys. *Journal of Criminal Justice, 5*(3), 205-216.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford Stanford university press.

Goudriaan, H., Lynch, J. P., & Nieuwbeerta, P. (2003). Aangifte doen bij de politie? De invloed van de nationale context op het aangiftegedrag van slachtoffers van criminaliteit in zestien westerse landen. *Tijdschrift voor Criminologie, 45*(2), 35-52.

Grapendaal, M., & van Tilburg, W. (2002). Veelplegers in Nederland. *Tijdschrift voor Criminologie, 44*(3), 214-230.

Greely, H. T., Riordan, D. P., Garrison, N. A., & Mountain, J. L. (2006). Family Ties: The Use of DNA Offender Databases to Catch Offenders? *Journal of Law, Medecine & Ethics, 34*(2), 248-262.

Guerette, R. T., Stenius, V. M. K., & McGloin, J. M. (2005). Understanding offense specialization and versatility: A reapplication of the rational choice perspective. *Journal of Criminal Justice, 33*(1), 77-87. doi:https://doi.org/10.1016/j.jcrimjus.2004.10.005

Haen Marschall, I. (1996). De methode van zelfrapportage. Aanzet tot rationele benadering. *Tijdschrift voor Criminologie, 38*(1), 2-20.

Hallauer, R. (2014). *The Swiss situation concerning the comparison of DNA data within Prüm and other frameworks* Paper presented at the PIES Workshop, Brussels.

Hampikian, G., West, E., & Akselrod, O. (2011). The Genetics of Innocence: Analysis of 194 U.S. DNA Exonerations. *Annual Review of Genomics and Human Genetics, 12*(1), 97-120. doi:10.1146/annurev-genom-082509-141715

Harrison, K. (2006). Is Crime Scene Examination science, and does it matter anyway? *Science & Justice, 46*(2), 65-68.

Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange. *Library & Information Science Research, 18*(4), 323-342. doi:doi.org/10.1016/S0740-8188(96)90003-1

Hibdon, J., Telep, C. W., & Groff, E. R. (2016). The Concentration and Stability of Drug Activity in Seattle, Washington Using Police and Emergency Medical Services Data. *Journal of Quantitative Criminology, 33*(3), 497-517. doi:10.1007/s10940-016-9302-0

Hindelang, M. J., Hirschi, T., & Weis, J. G. (1979). Correlates of delinquency: the illusion of discrepancy between selfreport and official measures. *American Sociological Review, 44*(6), 995-1014.

Hindelang, M. J., Hirschi, T., & Weis, J. G. (1981). *Measuring delinquency*. Beverly Hills (Calif.): Sage.

Hirschi, T. (1969). *Causes of delinquency*. Berkeley University of California press.

Hochschild, J. L., & Sen, M. (2012). *DNA and criminal justice: public opinion on a new policy*. Paper presented at the Midwest Political Science Association Conference, Chicago.

Home Office. (2004). *The Home Office Strategic Plan 2004-2008*. Retrieved from https://www.gov.uk/government/publications/the-home-office-strategic-plan-2004-to-2008

Home Office. (2005). *DNA Expansion Programme 2000–2005: Reporting achievement*. Retrieved from www.statewatch.org/news/2006/jan/uk-DNA-database.pdf

Home Office. (2009). *Keeping the right people on the DNA database. Science and public protection*. Retrieved from http://library.college.police.uk/docs/homeoffice/cons-2009-dna-response.pdf

Home Office. (2014). *National DNA Database strategy board. Annual report 2013-14*. Retrieved from https://www.gov.uk/government/publications/national-dna-database-annual-report-2013-to-2014

Hopwood, A. J., Hurth, C., Yang, J., Cai, Z., Moran, N., Lee-Edghill, J. G., . . . Tully, G. (2010). Integrated Microfluidic System for Rapid Forensic DNA Analysis: Sample Collection to DNA Profile. *Analytical Chemistry, 82*(16), 6991-6999. doi:10.1021/ac101355r

Huisman, M. (2009). Imputation of missing network data. *Journal of Social Structure, 10*(1), 1-28.

Huizinga, D., Esbensen, F.-A., & Weiher, A. W. (1991). Are there multiple paths to delinquency? *The Journal of Criminal Law & Criminology, 82*(1), 83-118.

Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist, 11*(2), 37-50.

doi:10.1111/j.1469-8137.1912.tb05611.x

Jamieson, A. (2004). A rational approach to the principles and practice of crime scene investigation: I. Principles. *Science & Justice, 44*(1), 3-7.

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Individual-specific 'fingerprints' of human DNA. *Nature, 316*, 76-79.

Jeuniaux, P. (2014). *A statistical analysis of the Belgian DNA database.* Paper presented at the PIES workshop, Brussels.

Jeuniaux, P., De Moor, S., Robert, L., Renard, B., Stappers, C., & Vanvooren, V. (2018). Reconstruction and study of offending trajectories through forensic evidence. An illustration using a forensic DNA database. In Q. Rossy, D. Décary-Hétu, O. Delémont, & M. Mulone (Eds.), *The Routledge International Handbook of Forensic Intelligence and Criminology* (pp. 177-190). Oxford: Routledge.

Jeuniaux, P., Duboccage, L., Renard, B., Van Renterghem, P., & Vanvooren, V. (2016). Establishing networks in a forensic DNA database to gain operational and strategic intelligence. *Security Journal, 29*(4), 584-602. doi:10.1057/sj.2015.31

Jeuniaux, P., Renard, B., Duboccage, L., Steuve, S., Stappers, C., Gallala, I., . . . Vanvooren, V. (2015). Managing forensic DNA records in a divided world: the Belgian case. *Records Management Journal, 25*(3), 269-287. doi:10.1108/RMJ-05-2015-0018

Jobling, M. A., & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics, 5*, 739-752.

Johnson, D. (2013). The space/time behaviour of dwelling burglars: finding near repeat patterns in serial offender data. *Applied Geography, 41*, 139-146.

Junger-Tas, J., & Haen Marshall, I. (1999). The self-report methodology in crime research. *Crime and Justice, 25*, 291-367.

Juodis, M., Woodworth, M., Porter, S., & Ten Brinke, L. (2009). Partners in crime: a comparison of individual and multi-perpetrator homicides. *Criminal Justice and Behavior, 36*(8), 824-839.

Kaye, D. H., & Smith, M. E. (2003). DNA identification databases: legality, legitimacy, and the case for population-wide coverage. *Wisconsin Law Review*, 413-459.

Kazemian, L., Pease, K., & Farrington, D. P. (2011). DNA retention policies: the potential contribution of criminal career research. *European Journal of Criminology, 8*(1), 48-64.

Kirk, D. S. (2006). Examining the divergence across self-report and official data sources on inferences about the adolescent life-course of crime. *Journal of Quantitative Criminology, 22*, 107-129.

Kirk, P. L. (1963). The ontogeny of criminalistics. *Journal of Criminal Law and Criminology, 54*(2), 235-238.

Kivivuori, J. (2011). *The discovery of hidden crime: self-report delinquency surveys in criminal policy context*. New York: Oxford University Press.

Kloosterman, A., Sjerps, M., & Quak, A. (2014). Error rates in forensic DNA analysis: definition, numbers, impact and communication. *Forensic Science International: Genetics, 12*, 77-85.

Knoke, D., & Yang, S. (2008). *Social Network Analysis* (Vol. 154). Thousand Oaks: Sage Publications.

Kocsis, R. N., & Irwin, H. J. (1998). The psychological profile of serial offenders and a redefinition of the misnomer of serial crime. *Psychiatry, psychology and law, 5*(2), 197-213.

Kossinets, G. (2006). Effects of missing data in social networks. *Social Networks, 28*(3), 247-268. doi:10.1016/j.socnet.2005.07.002

Lammers, M. (2013). *Catch me if you can. Using DNA traces to study the influence of offending behaviour on the probability of arrest.* Vrije Universiteit Amsterdam, Amsterdam.

Lammers, M. (2014a). Are arrested and non-arrested serial offenders different? A test of spatial offending patterns using DNA found at crime scenes. *Journal of Research in Crime and Delinquency, 51(2)*, 143-167.

Lammers, M. (2014b). *Using DNA traces to study offending behaviour.* Paper presented at the PIES Workshop, Brussels.

Lammers, M., & Bernasco, W. (2013). Are mobile offenders less likely to be caught? The influence of the geographical dispersion of serial offenders' crime locations on their probability of arrest. *European Journal of Criminology, 10*(2), 168-186.

Lammers, M., Bernasco, W., & Elffers, H. (2012). How long do offenders escape arrest? Using DNA traces to analyse when serial offenders are caught. *Journal of Investigative Psychology and Offender Profiling, 9*, 13-29.

Lammers, M., Bernasco, W., & van de Beek, K. (2011). Criminologisch onderzoek met DNA-sporen. *Panopticon, 2*, 39-53.

Lantz, B., & Ruback, R. B. (2016). The Relationship Between Co-Offending, Age, and Experience Using a Sample of Adult Burglary Offenders. *Journal of Developmental and Life-Course Criminology, 3*(1), 76-97. doi:10.1007/s40865-016-0047-0

Laumann, E. O., Marsden, P. V., & Prensky, D. (1987). The Boundary Specification Problem in Network Analysis. In L. C. Freeman, D. R. White, & A. K. Romney (Eds.), *Research Methods in Social Network Analysis* (pp. 61-79). New Brunswick: Transaction Publishers.

Leary, D., & Pease, K. (2003). DNA and the active criminal population. *Crime Prevention and Community Safety: An International Journal, 5*, 7-12.

Lee, J. W., Lee, H.-S., Park, M., & Hwang, J.-J. (2001). Evaluation of DNA match probability in criminal case. *Forensic Science International, 116*, 139-148.

Legrand, T., & Vogel, L. (2012). Forensic intelligence. *Briefing Paper*(9), 1-5.

Legrand, T., & Vogel, L. (2014). The landscape of forensic intelligence research. *Australian Journal of Forensic Sciences*, 1-11.

Levitt, M. (2007). Forensic databases: benefits and ethical and social costs. *British Medical Bulletin, 83*, 235-248.

Ley, B. L., Jankowski, N., & Brewer, P. R. (2012). Investigating CSI: portrayals of DNA testing on a forensic crime show and their potential effects. *Public Understanding of Science, 21*(1), 51-67. doi:10.1177/0963662510367571

Linning, S. J., Andresen, M. A., & Brantingham, P. J. (2016). Crime Seasonality: Examining the Temporal Fluctuations of Property Crime in Cities With Varying Climates. *International Journal of Forecasting Offender Therapy and Comparative Criminology*. doi:10.1177/0306624X16632259

Locard, E. (1923). *Manuel de technique policière : (enquête criminelle)*: Paris : Payot.

Lovell, R., Luminais, M., Flannery, D. J., Overman, L., Huang, D., Walker, T., & Clark, D. R. (2017).

Offending patterns for serial sex offenders identified via the DNA testing of previously unsubmitted sexual assault kits. *Journal of Criminal Justice, 52*, 68-78. doi:10.1016/j.jcrimjus.2017.08.002

Lundrigan, S., Czarnomski, S., & Wilson, M. (2010). Spatial and environmental consistency in serial sexual assault. *Journal of Investigative Psychology and Offender Profiling, 7*(1), 15-30. doi:10.1002/jip.100

Lynam, D. R., Piquero, A. R., & Moffitt, T. E. (2004). Specialization and the propensity to violence. Support from self-reports but not official records. *Journal of Contemporary Criminal Justice, 20*(2), 2015-2228.

Lynch, M. (2003). God's signature: DNA profiling, the new gold standard in forensic science. *Endeavour, 27*(2), 93-97.

Machado, H., & Silva, S. (2014). "Would you accept having your DNA profile inserted in the National Forensic DNA database? Why?" Results of a questionnaire applied in Portugal. *Forensic Science International: Genetics, 8*, 132-136.

Machado, H., Silva, S., & Cunha, M. (2012). Multiple views of DNA surveillance: the surveilled, the surveillants and the academics. *Crime, Security and Surveillance*, 177-192.

Malm, A., & Bichler, G. (2011). Networks of collaborating criminals: assessing the structural vulnerability of drug markets. *Journal of Research in Crime and Delinquency, 48*(2), 271-297.

Malsch, M., de Keijser, J. W., Luining, E. T., Weulen Kranenbarg, M., & Lenssen, D. J. H. M. (2016). Hoe hard is DNA-bewijs? Internationaal-vergelijkend onderzoek naar de interpretatie van DNA-profielen. *Nederlands Juristenblad, 18*, 1261-1266.

Malsch, M., Taverne, M. D., Elffers, H., de Keijser, J. W., & Kranendonk, P. R. (2013). *DNA rapporten: makkelijker kunnen we het niet maken, begrijpelijker wel*. Den Haag.

Marin, A., & Wellman, B. (2011). Social network analysis: an introduction. In J. Scott & P. J. Carrington (Eds.), *The Sage handbook of social network analysis* (pp. 11-25). London: Sage.

Martin, P. D. (2004). National DNA databases—practice and practicability. A forum for discussion. *International Congress Series, 1261*, 1-8.

Martin, P. D., Schmitter, H., & Schneider, P. M. (2001). A brief history of the formation of DNA databases in forensic science within Europe. *Forensic Science International, 119*, 225-231.

Mason, J. (2003). *Qualitative researching* (2nd ed.). London Sage.

Matza, D. (1967). *Delinquency and drift*. New York: Transaction Publishers.

Maxfield, M. G., Luntz Weiler, B., & Spatz Widom, C. (2000). Comparing self-reports and official records of arrests. *Journal of Quantitative Criminology, 13*(1), 87-110.

Mazerolle, P., Brame, R., Paternoster, R., Piquero, A. R., & Dean, C. (2000). Onset age, persistence, and offending versatility: comparisons across gender. *Criminology, 38*(4), 1143-1172.

McCartney, C. (2004). Forensic DNA sampling and the England and Wales National DNA Database: a sceptical approach. *Critical Criminology, 12*, 157-178.

McCartney, C. (2012). Of Weighty Reasons and Indiscriminate Blankets: The Retention of DNA for Forensic Purposes. *The Howard Journal of Criminal Justice, 51*(3), 245-260. doi:10.1111/j.1468-2311.2012.00717.x

McCartney, C., Wilson, T. J., & Williams, R. (2011). Transnational exchange of forensic DNA: viability, legitimacy, and acceptability. *European Journal on Criminal Policy and Research, 17*, 305-322.

McGloin, J. M. (2005). Policy and Intervention Considerations of a Network Analysis of Street Gangs. *Criminology and Public Policy, 4*(3), 607-636.

McGloin, J. M. (2009). Delinquency balance: revisiting peer influence. *Criminology, 47*(2), 439-477. doi:10.1111/j.1745-9125.2009.00146.x

McGloin, J. M., & Kirk, D. S. (2010). Social network analysis. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 209-224). New York: Springer.

McGloin, J. M., Sullivan, C. J., Piquero, A. R., & Pratt, T. C. (2007). Local life circumstances and offending specialization/versatility: comparing opportunity and propensity models. *Journal of Research in Crime and Delinquency, 44*(3), 321-346.

Meijer, R. F., van Panhuis, P., Siero, S., & Smit, P. (2002). Elf procent verdachten verantwoordelijk voor 20 of 60 procent van de criminaliteit? *Tijdschrift voor Criminologie, 44*(3), 282-284.

Metcalfe, C. F., & Baker, T. (2014). The drift from convention to crime: Exploring the relationship between co-offending and intermittency. *Criminal Justice and Behavior, 41*(1), 75-90. doi:10.1177/0093854813500775

Meulenbroek, A. J. (2009). *De essenties van forensisch biologisch onderzoek. Humane biologische sporen en DNA*. Paris: Zutphen.

Ministerraad. (2016). *Kadernota Integrale Veiligheid 2016-2019*. Retrieved from https://justitie.belgium.be/nl/nieuws/andere_berichten_29

Morelato, M., Baechler, S., Ribaux, O., Beavis, A., Tahtouh, M., Kirkbride, P., . . . Margot, P. (2014). Forensic intelligence framework—Part I: Induction of a transversal model by comparing illicit drugs and false identity documents monitoring. *Forensic Science International, 236*, 181-190.

Morelato, M., Beavis, A., Tahtouh, M., Ribaux, O., Kirkbride, P., & Roux, C. (2013). The use of forensic case data in intelligence-led policing: The example of drug profiling. *Forensic Science International, 226*, 1-9.

Morselli, C. (2009). *Inside criminal networks* (Vol. 8). New York: Springer.

Morselli, C., Giguère, C., & Petit, K. (2007). The efficiency/security trade-off in criminal networks. *Social Networks, 29*(1), 143-153. doi:10.1016/j.socnet.2006.05.001

Mortelmans, D. (2007). *Handboek kwalitatieve onderzoeksmethoden*. Leuven Acco.

Murphy, E. (2018). Forensic DNA Typing. *Annual Review of Criminology, 1*(1), 497-515. doi:10.1146/annurev-criminol-032317-092127

National DNA Database. (2004) The National DNA Database annual report 03/04. London.

Nazaretian, Z., & Merolla, D. M. (2013). Questioning Canadian Criminal Incidence Rates: A Re-analysis of the 2004 Canadian Victimization Survey. *Canadian Journal of Criminology and Criminal Justice, 55*(2), 239-261.

NICC/INCC. (2014). *Wat U moet weten over forensisch DNA-onderzoek*. Retrieved from https://nicc.fgov.be/upload/files/OverhetNICC/nicc-dna-magazine-nl-lr.pdf

NICC/INCC. (2018). De uitwisseling van DNA-gegevens tussen Spanje en België zeer succesvol. Retrieved from https://nicc.fgov.be/de-uitwisseling-van-dna-gegevens-tussen-spanje-en-belgie-zeer-succesvol

Nieuwbeerta, P., Blokland, A. A. J., Piquero, A. R., & Sweeten, G. (2011). A life course analysis of offense specializaton across age: introducing a new method for studying individual specialization over the life course. *Crime & Delinquency, 57*(1), 3-28. doi:10.1177/0011128710376336

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Norwich: Geo Books.

Papachristos, A. V. (2011). The coming of a networked criminology? In J. M. MacDonald (Ed.), *Measuring crime & criminality* (Vol. 17, pp. 101-140). New Brunswick: Transaction Publishers.

Papachristos, A. V., & Wildeman, C. (2014). Network Exposure and Homicide Victimization in an African American Community. *American Journal of Public Health, 104*(1), 143–150.

Paternoster, R. (2014). Career criminals and criminological theory. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of Criminology and Criminal Justice* (pp. 276-285). New York: Springer.

Pauwels, L. (2017). *Kwantitatieve criminologie: basishandboek kwantitatieve methoden van criminologisch onderzoek*. Gent: Academia Press.

Penacino, G., Sala, A., & Corach, D. (2003). Are DNA tests infallible? *International Congress Series, 1239*, 873-877.

Piquero, A. R., Farrington, D. P., & Blumstein, A. (2003). The criminal career paradigm. In M. Tonry & N. Morris (Eds.), *Crime and Justice: A Review of Research* (Vol. 30, pp. 359-506). Chicago: University of Chicago Press.

Prainsack, B., & Toom, V. (2013). Performing the Union: The Prüm Decision and the European dream. *Studies in History and Philosophy of Biological and Biomedical Sciences, 44*, 71-79.

Prell, C. (2013). *Social network analysis: History, Theory & Methodology*. Los Angeles: Sage.

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Ratcliffe, J. H. (2004a). Geocoding Crime and a First Estimate of an Acceptable Minimum Hit Rate. *International Journal of Geographical Information Science, 18*(1), 61-72. doi:10.1080/13658810310001596076

Ratcliffe, J. H. (2004b). The hotspot matrix: a framework for the spatio-temporal targeting of crime reduction. *Police Practice and Research, 5*(1), 5-23.

Ratcliffe, J. H. (2007). *Integrated intelligence and crime analysis: enhanced information management for law enforcement leaders*. Retrieved from http://www.policefoundation.org/docslibrary.html and http://www.jratcliffe.net/wp-content/uploads/Ratcliffe-2007-Integrated-intelligence-and-crime-analysis.pdf

Ratcliffe, J. H. (2010). The spatial dependency of crime increase dispersion. *Security Journal, 23*(1), 18-36. doi:10.1057/sj.2009.16

Raymond, J. J., van Oorschot, R. A. H., Gunn, P. R., Walsh, S. J., & Roux, C. (2004). Trace evidence characteristics of DNA: A preliminary investigation of the persistence of DNA at crime scenes. *Forensic Science International: Genetics, 4*, 26-33.

Reid, A. A., Tayebi, M., & Frank, R. (2013). *Exploring the Structural Characteristics of Social Networks in a Large Criminal Court Database*. Paper presented at the IEEE International Conference on Intelligence and Security Informatics.

Reiss, A. J. (1988). Co-offending and criminal careers. *Crime and Justice, 10*, 117-170.

Reiss, A. J., & Farrington, D. P. (1991). Advancing knowledge about co-offending: results from a prospective longitudinal survey of London males. *The Journal of Criminal Law & Criminology, 82*(2), 360-395.

Renard, B., Duboccage, L., Jeuniaux, P., & Vanvooren, V. (2013). Les banques nationales de données génétiques en Belgique. Un premier bilan de 10 ans d'activité. *Revue de Droit Pénal et de Criminologie, 12*, 927-961.

Renard, B., & Jeuniaux, P. (2012). *De kosten en het gebruik van DNA-onderzoek in strafzaken*. Retrieved from https://nicc.fgov.be/upload/files/ODcriminologie/Expertises/brochure_cout_ADN_INCC_2012-03-09_NL.pdf

Renard, B., Vander Beken, T., De Hert, P., Stappers, C., De Moor, S., & Gallala, I. (2018). *Be-Gen eindrapport*. Retrieved from www.belspo.be

Ribaux, O., Baylon, A., Lock, E., Delémont, O., Roux, C., Zingg, C., & Margot, P. (2010). Intelligence-led crime scene processing. Part II: Intelligence and crime scene examination. *Forensic Science International, 199*(1-3), 63-71.

Ribaux, O., Baylon, A., Roux, C., Delémont, O., Lock, E., Zingg, C., & Margot, P. (2010). Intelligence-led crime scene processing. Part I : Forensic intelligence. *Forensic Science International, 195*(1-3), 10-16.

Ribaux, O., Crispino, F., & Roux, C. (2014). Forensic intelligence: deregulation or return to the roots of forensic science. *Australian Journal of Forensic Sciences*, 1-11.

Ribaux, O., Girod, A., Walsh, S. J., Margot, P., Mizrahi, S., & Clivaz, V. (2003). Forensic intelligence and crime analysis. *Law, Probability and Risk, 2*, 47-60.

Ribaux, O., & Margot, P. (2003). Case based reasoning in criminal intelligence using forensic case data. *Science and Justice, 43*(3), 135-143.

Ribaux, O., & Talbot Wright, B. (2014). Expanding forensic science through forensic intelligence. *Science and Justice, 54*, 494-501.

Ribaux, O., Walsh, S. J., & Margot, P. (2006). The contribution of forensic science to crime analysis and investigation: Forensic intelligence. *Forensic Science International, 156*, 171–181.

Ribbers, P. (2015). DNA-databanken voor strafzaken in perspectief. In D. van Beelen, C. Klostermann, G. Noordeloos, & P. Ribbers (Eds.), *Privacy en gegevensbescherming* (pp. 89-106). Apeldoorn-Antwerpen: Maklu.

Robert, M. (2012). *Forensic intelligence*. Boca Raton: CRC Press.

Rocchini, C. (2007). Hue scale representing node betweenness on a graph

Roman, J. K., Reid, S., Reid, J., Chalfin, A., Adams, W., & Knight, C. (2008). *The DNA field experiment: Cost-effectiveness analysis of the use of DNA in the investigation of high-volume crimes*. Retrieved from http://www.urban.org/url.cfm?id=411697&RSSFeed=UI_Crime/Justice.xml

Rossy, Q., Ioset, S., Dessimoz, D., & Ribaux, O. (2013). Integrating forensic information in a crime intelligence database. *Forensic Science International, 230*(1-3), 137-146.

Rossy, Q., & Morselli, C. (2018). The contribution of forensic science to the analysis of crime networks. In Q. Rossy, D. Décary-Hétu, O. Delémont, & M. Mulone (Eds.), *The Routledge International*

Handbook of Forensic Intelligence and Criminology (pp. 191-204). Oxford: Routledge.

Rothstein, M. A., & Talbott, M. K. (2006). The Expanding Use of DNA in Law Enforcement. What Role for Privacy? *Journal of Law, Medicine & Ethics, 34*(2), 153-164.

Roux, C., Crispino, F., & Ribaux, O. (2012). From forensics to forensic science. *Current Issues in Criminal Justice, 24*(1), 7-24.

Roux, C., Talbot-Wright, B., Robertson, J., Crispino, F., & Ribaux, O. (2015). The end of the (forensic science) world as we know it? The example of trace evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1674). doi:10.1098/rstb.2014.0260

Santos, F., & Machado, H. (2017). Patterns of exchange of forensic DNA data in the European Union through the Prüm system. *Science & Justice, 57*(4), 307-313. doi:http://doi.org/10.1016/j.scijus.2017.04.001

Santos, F., Machado, H., & Silva, S. (2013). Forensic DNA databases in European countries: is size linked to performance? *Life Sciences, Society and Policy, 9*(1), 12. doi:10.1186/2195-7819-9-12

Santos, R. B. (2013). *Crime analysis with crime mapping*. Los Angeles: Sage.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students* (5th ed.). New York: Prentice Hall.

Schneider, P. M., & Martin, P. D. (2001). Criminal DNA databases: the European situation. *Forensic Science International, 119*, 232-238.

Scott, J. (2013). *Social network analysis* (3rd ed.). Los Angeles: Sage.

Sellin, T. (1931). The basis of a crime index. *Journal of Criminal Law and Criminology, 22*, 335-356.

Sheptycki, J. (2004). Organizational pathologies in police intelligence systems. Some contributions to the lexicon of intelligence-led policing. *European Journal of Criminology, 1*(3), 307-332.

Simpson, S., Lyday, R., Hayasaka, S., Marsh, A., & Laurienti, P. (2013). A permutation testing framework to compare groups of brain networks. *Frontiers in Computational Neuroscience, 7*(171). doi:10.3389/fncom.2013.00171

Skogan, W. G. (1984). Reporting crimes to the police: The status of world research. *Journal of Research in Crime and Delinquency, 21*(2), 113-137.

Smith, J. A., & Moody, J. (2013). Structural Effects of Network Sampling Coverage I: Nodes Missing at Random. *Social Networks, 35*(4). doi:10.1016/j.socnet.2013.09.003

Smith, J. A., Moody, J., & Morgan, J. (2017). Network sampling coverage II: The effect of non-random missing data on network measurement. *Social Networks, 48*, 78-99. doi:10.1016/j.socnet.2016.04.005

Smith, M. E. (2006). Let's Make the DNA Identification Database as Inclusive a Possible. *Journal of Law, Medecine & Ethics, 34*(2), 385-389.

Song, Y. S., Patil, A., Murphy, E. E., & Slatkin, M. (2009). Average probability that a "cold hit" in a DNA database search results in a erroneous attribution. *Journal of Forensic Science, 54*(1), 22-27.

Sparrow, M. K. (1991). The application of network analysis to criminal intelligence: An assessment of the prospects. *Social Networks, 13*, 251-274.

Spradley, J. P. (2016). *Participant Observation*. Long Grove: Waveland Press.

Staley, K. (2005). *The Police National DNA Database: Balancing Crime Detection, Human Rights and Privacy*. Retrieved from http://www.genewatch.org/uploads/f03c6d66a9b354535738483c1c3d49e4/NationalDNADatabase.pdf

Stappers, C. (2018). *DNA in strafzaken: een analyse van het gebruik*. Paper presented at the Journée Nationale ADN – Nationale Studiedag DNA, Brussels.

Stappers, C., De Moor, S., Gallala, I., Jeuniaux, P., Vanvooren, V., & Renard, B. (2016). DNA-databank veroordeelden: ethische kwesties en juridische keuzes. *Fatik, 33*(149), 5-11.

Steenbeek, W., & Bernasco, W. (2018). lorenzgini: generalized Gini for sparse data situations. R package version 0.1.1. Retrieved from https://github.com/wsteenbeek/lorenzgini

Steenbeek, W., Vandeviver, C., Andresen, M. A., Malleson, N., & Wheeler, A. (2018). sppt: Spatial Point Pattern Test. R package version 0.2.1. Retrieved from https://github.com/wsteenbeek/sppt

Steenbeek, W., & Weisburd, D. (2015). Where the Action is in Crime? An Examination of Variability of Crime Across Different Spatial Units in The Hague, 2001–2009. *Journal of Quantitative Criminology, 32*(3), 449-469. doi:10.1007/s10940-015-9276-3

Stevens, A. P. (2001). Arresting crime: expanding the scope of DNA databases in America. *Texas Law Review, 79*, 921-960.

Struijk, S. (2009). De nieuwe ISD-maatregel. Een kritische verkenning van wetgeving en praktijk. *Proces. Maandblad voor Berechting en Reclassering*, 1-12.

Sutherland, E. H., & Cressey, D. R. (1960). *Principles of criminology* (6th ed.). Chicago: Lippincott.

Taverne, M. D., & Broeders, A. P. A. (2017). Cross-border patterns in DNA matches between the Netherlands and Belgium. *Science and Justice, 57*(1), 28-34. doi:10.1016/j.scijus.2016.08.008

Thompson, W. C. (2008). *The potential for error in forensic DNA testing (and how that complicates the use of DNA databases for criminal identification)* Paper presented at the Forensic DNA Databases and Race: Issues, Abuses and Actions, New York.

Thornberry, T. P., & Krohn, M. D. (2000). The self-report method for measuring delinquency and crime. *Criminal Justice, 4*, 33-83.

Thornberry, T. P., & Krohn, M. D. (2003). Comparison of Self-Report and Official Data for Measuring Crime. In N. R. Council (Ed.), *Measurement Problems in Criminal Justice Research: Workshop Summary*. Washington, DC: The National Academies Press.

Tilley, N., & Townsley, M. (2009). Forensic science in UK policing: strategies, tactics and effectiveness. In J. Fraser & R. Williams (Eds.), *Handbook of forensic science* (pp. 359-379). Cullompton: Willian Publishing.

Tilstone, W. J., Savage, K. A., & Clark, L. A. (2006). *Forensic Science: An Encyclopedia of History, Methods, and Techniques*. California: ABC-CLIO Ltd.

Tompson, L., Johnson, S., Ashby, M., Perkins, C., & Edwards, P. (2014). UK open source crime data: accuracy and possibilities for research. *Cartography and Geographic Information Science, 42*(2), 97-111. doi:10.1080/15230406.2014.972456

Tontodonato, P. (1996). *Group offending and criminal careers: violence among juvenile delinquents and adult offenders*. Retrieved from https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=172228

Townsley, M., & Sidebottom, A. (2010). All offenders are equal, but some are more equal than others:

variation in journeys to crime between offenders. *Criminology, 48*(3), 897-917. doi:10.1111/j.1745-9125.2010.00205.x

Townsley, M., Smith, C., & Pease, K. (2005). *Using DNA to catch offenders quicker: serious detections arising from criminal justice samples*. Retrieved from https://dspace.lboro.ac.uk/2134/852

Townsley, M., Smith, C., & Pease, K. (2006). First impressions count: serious detections arising from criminal justice samples. *Genomics, Society and Policy, 2*(1), 28-40.

Tracy, P. E., & Morgan, V. (2000). Big Brother and his science kit: DNA databases for 21st centry crime control. *Journal of Criminal Law and Criminology, 90*(2), 635-690.

Tremblay, P. (1993). Searching for suitable co-offenders. In R. V. Clarke & M. Felson (Eds.), *Routine activity and rational choice (Advances in criminological theory, vol. 5)* (pp. 17-36). New Brunswick: Transaction Publishers.

Tseloni, A., & Pease, K. (2011). DNA retention after arrest: balancing privacy interests and public protection. *European Journal of Criminology, 8*(1), 32-47.

UNODC. (2009). *Crime scene and physical evidence awareness for non forensic personnel*. Retrieved from http://www.unodc.org/unodc/en/scientists/crime-scene-and-physical-evidence-awareness-for-non-forensic-personnel.html

Van Daele, S., & Vander Beken, T. (2011). Outbound offending: the journey to crime and crime sprees. *Journal of Environmental Psychology, 31*, 70-78.

van der Beek, K. (2011). Forensic DNA Profiles Crossing Borders in Europe (Implementation of the Treaty of Prüm). Retrieved from www.promega.com/resources/profiles-in-dna/2011/forensic-dna-profiles-crossing-borders-in-europe/?activeTab=0

van der Hulst, R. C. (2009). Introduction to Social Network Analysis (SNA) as an investigative tool. *Trends in Organized Crime, 12*(2), 101-121. doi:10.1007/s12117-008-9057-6

van Kesteren, J., Mayhew, P., & Nieuwbeerta, P. (2000). *Criminal victimisation in seventeen industrialised countries*. Retrieved from https://pure.uvt.nl/portal/en/publications/criminal-victimisation-in-seventeen-industrialised-countries%284763395c-6496-4a7c-9d99-2ec5f7581987%29.html

van Kesteren, J., van Dijk, J., & Mayhew, P. (2014). The international crime victims surveys: a retrospective. *International Review of Victimology, 20*(1), 49-69.

van Mastrigt, S. B. (2014). Co-offending and offender attributes. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of Criminology and Criminal Justice* (pp. 559-570). New York: Springer.

van Mastrigt, S. B., & Carrington, P. J. (2014). Sex and age homophily in co-offending networks: opportunity or preference? In C. Morselli (Ed.), *Crime and Networks* (pp. 28-51). New York: Routledge.

van Mastrigt, S. B., & Farrington, D. P. (2009). Co-offending, age, gender and crime type: implications for criminal justice policy. *British Journal of Criminology, 49*, 552-573.

van Sleeuwen, S. E. M., Ruiter, S., & Menting, B. (2018). A Time for a Crime. *Journal of Research in Crime and Delinquency*, 002242781876639. doi:10.1177/0022427818766395

Vandeviver, C., Neutens, T., Van Daele, S., Geurts, D., & Vander Beken, T. (2015). A discrete spatial choice model of burglary target selection at the house-level. *Applied Geography, 64*, 24-34.

Vandeviver, C., & Steenbeek, W. (2017). The (In)Stability of Residential Burglary Patterns on Street Segments: The Case of Antwerp, Belgium 2005–2016. *Journal of Quantitative Criminology*. doi:10.1007/s10940-017-9371-8

Versteegh, P., Janssen, J., & Bernasco, W. (2003). Beginners, doorstromers en veelplegers. Carrièrecriminaliteit in de politieregio Haaglanden. *Tijdschrift voor Criminologie, 45*(2), 127-139.

Vlaemynck, M. (2014). Sociale netwerkanalyse in de criminologie: een perspectief met toekomst. *Panopticon, 35*(1), 65-70.

Vlaemynck, M. (2016). *Disentangling 'social supply': a personal network study into the social world of recreational cannabis use and its supply.* Ghent University, Ghent.

Voultsos, P., Njau, S., Tairis, N., Psaroulis, D., & Kovatsi, L. (2011). Launching the Greek forensic DNA database. The legal framework and arising ethical issues. *Forensic Science International: Genetics, 5*, 407-410.

Wade, L. (2015). Cold Hit: New Technology Reveals Old Evidence. *Forensic Magazine*. Retrieved from https://www.forensicmag.com/article/2015/01/cold-hit-new-technology-reveals-old-evidence

Wallace, H. M. (2006). The UK National DNA Database. Balancing crime detection, human rights and privacy. *EMBO reports, 7*.

Wallace, H. M., Jackson, A. R., & Thibedeau, A. D. (2014). Forensic DNA databases: Ethical and legal standards: A global review. *Egyptian Journal of Forensic Sciences, 4*(3), 57-63.

Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data*. Hoboken, New Jersey: Wiley.

Walsh, S. J., Curran, J. M., & Buckleton, J. S. (2010). Modeling forensic DNA database performance. *Journal of Forensic Science, 55*(5), 1174-1183.

Wang, D. J., Shi, X., McFarland, D. A., & Leskovec, J. (2012). Measurement error in network data: A re-classification. *Social Networks, 34*(4), 396-409. doi:10.1016/j.socnet.2012.01.003

Warr, M. (1996). Organization and instigation in delinquent groups. *Criminology, 34*(1), 11-38.

Warr, M. (2002). *Companions in crime. The social aspects of criminal conduct*. Cambridge Cambridge University Press.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications.* New York: Cambridge University Press.

Weedn, V. W., & Hicks, J. W. (1997). The unrealized potential of DNA testing. *National Institute of Justice Journal*(234), 16-23.

Weerman, F. M. (2001). *Samenplegen. Over criminele samenwerking en groepsvorming*. Nijmegen: Ars Aequi Libri.

Weerman, F. M. (2003). Co-offending as social exchange: Explaining characteristics of co-offending. *British Journal of Criminology, 43*, 398–416.

Weerman, F. M. (2014). Theories of co-offending. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of Criminology and Criminal Justice* (pp. 5173-5184). New York: Springer.

Weerman, F. M., & Kleemans, E. (2002). Criminele groepen en samenwerkingsverbanden. *Tijdschrift voor Criminologie, 44*(2), 114-127.

Westerberg, K., Grant, T., & Bond, J. W. (2007). Triangulation mobility of auto-theft offenders. *Journal of Investigative Psychology and Offender Profiling, 4*, 109-120.

Wheeler, A., Steenbeek, W., & Andresen, M. A. (2018). Testing for Similarity in Area-Based Spatial Patterns: Alternative Methods to Andresen's Spatial Point Pattern Test. *Transactions in GIS*. doi:10.1111/tgis.12341

Wienroth, M., Morling, N., & Williams, R. (2014). Technological Innovations in Forensic Genetics: Social, Legal and Ethical Aspects. *Recent Advances in DNA and Gene Sequences, 8*(2), 98-103.

Wiles, P., & Costello, A. (2000). *The 'road to nowhere': the evidence for travelling criminals*. London: Home Office.

Williams, R., & Johnson, P. (2004). Wonderment and dread: representations of DNA in ethical disputes about forensic DNA databases. *New Genetics and Society, 23*(2), 205-223.

Williams, R., & Johnson, P. (2005). Inclusiveness, effectiveness and intrusiveness: issues in the developing uses of DNA profiling in support of criminal investigations. *Journal of Law, Medecine & Ethics,* 545-558.

Wilson, D. B., McClure, D., & Weisburd, D. (2010). Does Forensic DNA Help to Solve Crime? The Benefit of Sophisticated Answers to Naive Questions. *Journal of Contemporary Criminal Justice, 26*(4), 458-469. doi:10.1177/1043986210377231

Wilson, D. B., Weisburd, D., & McClure, D. (2011). Use of DNA testing in police investigative work for increasing offender identification, arrest, conviction and case clearance. *Campbell Systematic Reviews, 7,* 53.

Wittebrood, K. (2005). Van delictmeting tot officiële aangifte. Sprake van sociale ongelijkheid? In G. J. N. Bruinsma, W. Huisman, & R. van Swaaningen (Eds.), *Basisteksten in de criminologie I: aard, omvang en verklaringen* (pp. 90-103). Den Haag: Boom Juridische Uitgevers.

Xu, J., & Chen, H. (2005). Criminal network analysis and visualization. *Communications of the ACM, 48*(6), 101-107.

Xu, J., & Chen, H. (2008). The topology of dark networks. *Communications of the ACM, 51*(10), 58. doi:10.1145/1400181.1400198

Zaitch, D., Mortelmans, D., & Decorte, T. (2009). Participerende observatie in de criminologie. In T. Decorte & D. Zaitch (Eds.), *Kwalitatieve methoden en technieken in de criminologie* (pp. 257-309). Leuven: Acco.

APPENDIXES

Appendix 1:
Decision diagram for the selection and classification of trace carriers for possible DNA testing, and list of classification of the DNA carriers into categories (in Dutch)

MODEL 14 – BESLISSINGSDIAGRAM VOOR DE SELECTIE EN RANGSCHIKKING VAN DE SPOREN

	1. DNA waarschijnlijk afkomstig van de dader/ slachtoffer *	2. Afkomst DNA onbekend	3. DNA waarschijnlijk niet afkomstig van de dader/ slachtoffer *
A. Weefsel of lichaamsvocht (bv. bloed, sperma, speeksel,..)	Grote kans op bruikbaar DNA-profiel. + Grote kans op profiel dader/slachtoffer *	Grote kans op bruikbaar DNA-profiel. + Kans op profiel onbekende of mengprofiel	Grote kans op bruikbaar DNA-profiel. + Grote kans profiel niet-dader/ slachtoffer *
B. Goede dragers (cfr. lijst)	Kans op bruikbaar DNA-profiel + Grote kans op profiel dader/slachtoffer *	Kans op bruikbaar DNA-profiel + Kans op profiel onbekende of mengprofiel	Kans op bruikbaar DNA-profiel + Grote kans profiel niet-dader/ slachtoffer *
C. Minder goede dragers (bv. door kort contact, DNA gedegradeerd door weersomstandigheden, water,...)	Weinig kans op bruikbaar DNA-profiel + Grote kans op profiel dader/slachtoffer *	Weinig kans op bruikbaar DNA-profiel + Kans op profiel onbekende of mengprofiel	Weinig kans op bruikbaar DNA-profiel + Grote kans profiel niet-dader
X.	Geen codering mogelijk wegens het ontbreken van de nodige informatie i.v.m. sporenvrijwaring en oorsprong		

* schrappen wat niet past

Legende:

1 → 3: link met de op te sporen persoon (kan zowel de dader als het slachtoffer zijn) via de tactische gegevens
A → C: hoeveelheid/kwaliteit van DNA via wetenschappelijke en contextuele gegevens

Lijst met indeling van de DNA-dragers in categorieën

Hoofdgroep	Subgroep	Categorie
Biologisch materiaal	Bloed	A
Biologisch materiaal	Faeces	C
Biologisch materiaal	Haar (met wortel)	B
Biologisch materiaal	Speeksel	A
Biologisch materiaal	Sperma	A
Boorgaatjes	Boorgaatje	C
Contactsporen	Deurbel	C
Contactsporen	Deuropener	C
Contactsporen	Oor / kaakspoor	C
Contactsporen	Vinger / handpalmsporen	C
Contactsporen	Andere	C
Drankverpakkingen	Blik	A
Drankverpakkingen	Brik	A
Drankverpakkingen	Fles	A
Drankverpakkingen	Rietje	A
Etensresten	Etensresten	B
Kledij	Bril	B
Kledij	Broek	B
Kledij	Handschoenen – rubber – latex	A
Kledij	Handschoenen – fleece – leder – wol	B
Kledij	Handschoenen – werkhandschoen	B
Kledij	Andere	B
Kledij	Hoofddeksel	B
Kledij	Jas	B
Kledij	Kledij (nat)	C
Kledij	Kledij verpakt in plastiek	C
Kledij	Motorhelm	A
Kledij	Onderbroek	B
Kedij	Schoenen	C
Kledij	Sokken	B
Kledij	Trui	B
Kledij	T-shirt	B
Knevelmateriaal	Colsonbandjes	C

Knevelmateriaal	Handboeien	C
Knevelmateriaal	Tape	C
Knevelmateriaal	Touw	C
Peuken	Peuk	A
Peuken	Peuk (nat)	C
Referentie	Wangslijmvlies-Swab	A
Voertuig	Airbag	B
Voertuig	Dashboard passagierszijde	C
Voertuig	GPS	C
Voertuig	Handrem	C
Voertuig	Hendel motorkap	C
Voertuig	Nummerplaat	C
Voertuig	Portierkruk	C
Voertuig	Rand achteruitkijkspiegel	C
Voertuig	Stuurmidden	B
Voertuig	Stuurwiel	B
Voertuig	Versnellingspook	B
Voertuig	Verstelhendel bestuurszijde	C
Voertuig	Verstelhendel passagierszijde	C
Wapens	Mes	B
Wapens	Pistool / revolver	B
Wapens	Andere	B
Werktuigen	Beitel, breekijzer, koevoet, schroevendraaier, tang, zaklamp,..	B

Appendix 2:
Additional analyses of data registered in 2010, 2011, 2012, 2013 and 2015

Table 20: S-index proportion difference test

	2010	2011	2012	2013	2014	2015
Violent theft	0.7777778	0.8518519	0.8148148	0.8888889	0.8518519	0.8518519
Aggravated burglary	0.4444444	0.3333333	0.1481481	0.1851852	0.2222222	0.4074074
Lethal violence	1	0.962963	0.962963	1	1	1
Sexual offences	0.962963	1	0.962963	0.962963	0.962963	0.962963

2010

Table 21: Description of datasets for 2010

Crime type	**Police dataset**		**DNA dataset**	
	N	**%**	**N**	**%**
Violent theft	316	21.31	18,160	9.64
Aggravated burglary	1,033	69.66	162,862	86.42
Lethal violence	32	2.16	416	0.22
Sexual offences	102	6.88	7,010	3.72
Total	1,483	100	188,448	100

Table 22: Generalized Gini coefficients for 2010

2010	**Police dataset**				**DNA dataset**		
	G'	**S.E. G'**	**Lower CI**	**Upper CI**	**G'**	**S.E. G'**	**p**
4 crime types	0.6012937	0.07905201	0.4629325	0.7308341	0.5836018	0.05318141	0.468
Violent theft	0.7267784	0.08394231	0.5821693	.8423719	0.6551805	0.05410582	0.361
Aggravated burglary	0.5949841	0.07756618	0.4636374	0.7248930	0.6100893	0.05670676	0.282
Lethal violence	0.5692664	0.05737039	0.4891034	0.6548383	0.7800926	0.16651	< 0.001
Sexual offences	0.4463042	0.05601607	0.3556075	0.5715478	0.6746550	0.05943583	< 0.001

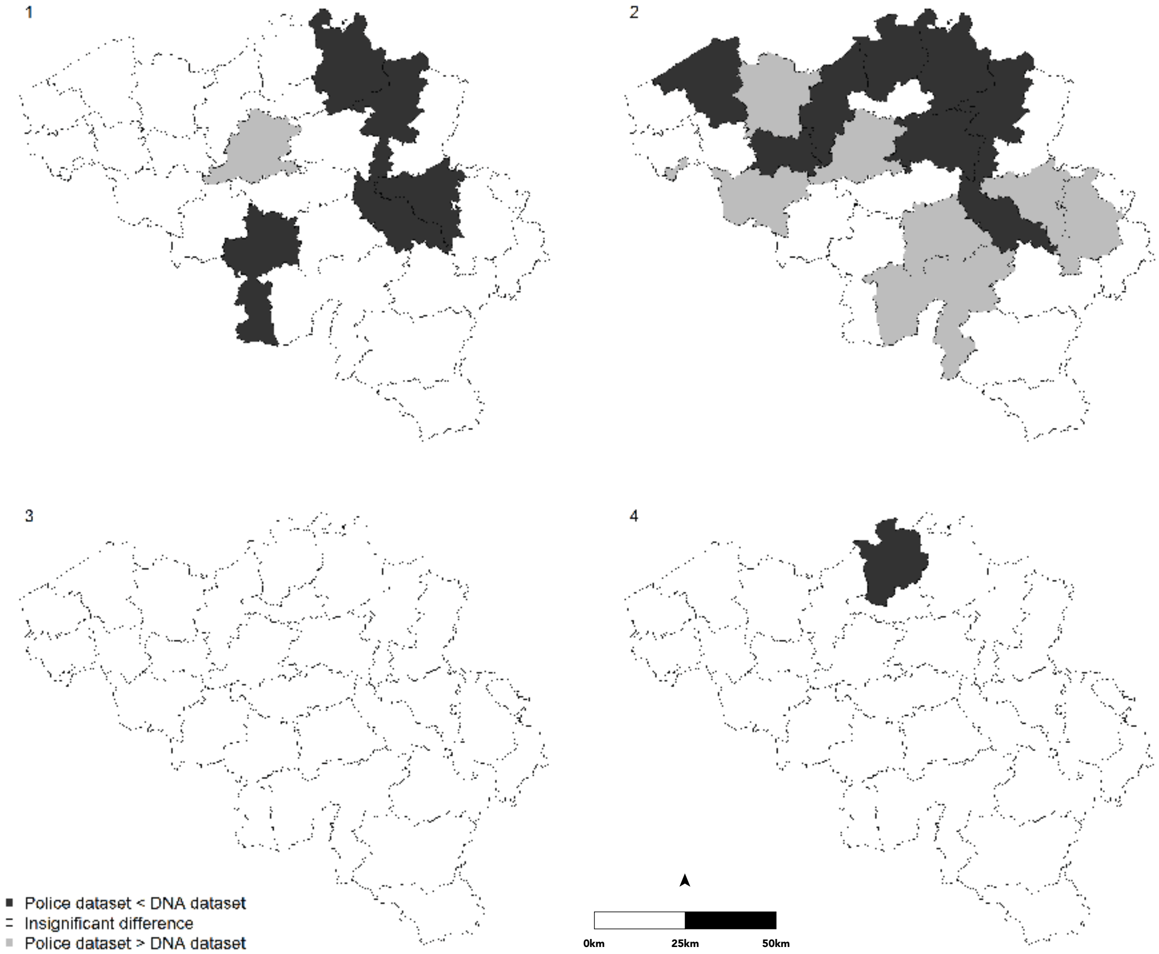


Figure 30: Local S-indices for violent theft (1), aggravated burglary (2), lethal violence (3), and sexual offences (4) for 2010

2011

Table 23: Description of datasets for 2011

Crime type	Police dataset		DNA dataset	
	N	%	N	%
Violent theft	372	18.09	21,580	10.71
Aggravated burglary	1,547	75.24	172,009	85.40
Lethal violence	33	1.61	406	0.20
Sexual offences	104	5.06	7,418	3.68
Total	2,056	100	201,413	100

Table 24: Generalized Gini coefficients for 2011

2011	Police dataset				DNA dataset		
	G'	S.E. G'	Lower CI	Upper CI	G'	S.E. G'	p
4 crime types	0.5853780	0.08169602	0.4399430	0.7180088	0.5812437	0.05514125	0.396
Violent theft	0.7465966	0.08803778	0.5929382	0.8586898	0.6716448	0.05368266	0.362
Aggravated burglary	0.5722758	0.07894417	0.4336265	0.7044098	0.5923532	0.06798178	0.265
Lethal violence	0.5714286	0.05693238	0.4842206	0.6744491	0.7003367	0.09071918	< 0.001
Sexual offences	0.4456927	0.05787306	0.3531753	0.5783612	0.6737892	0.05658388	< 0.001

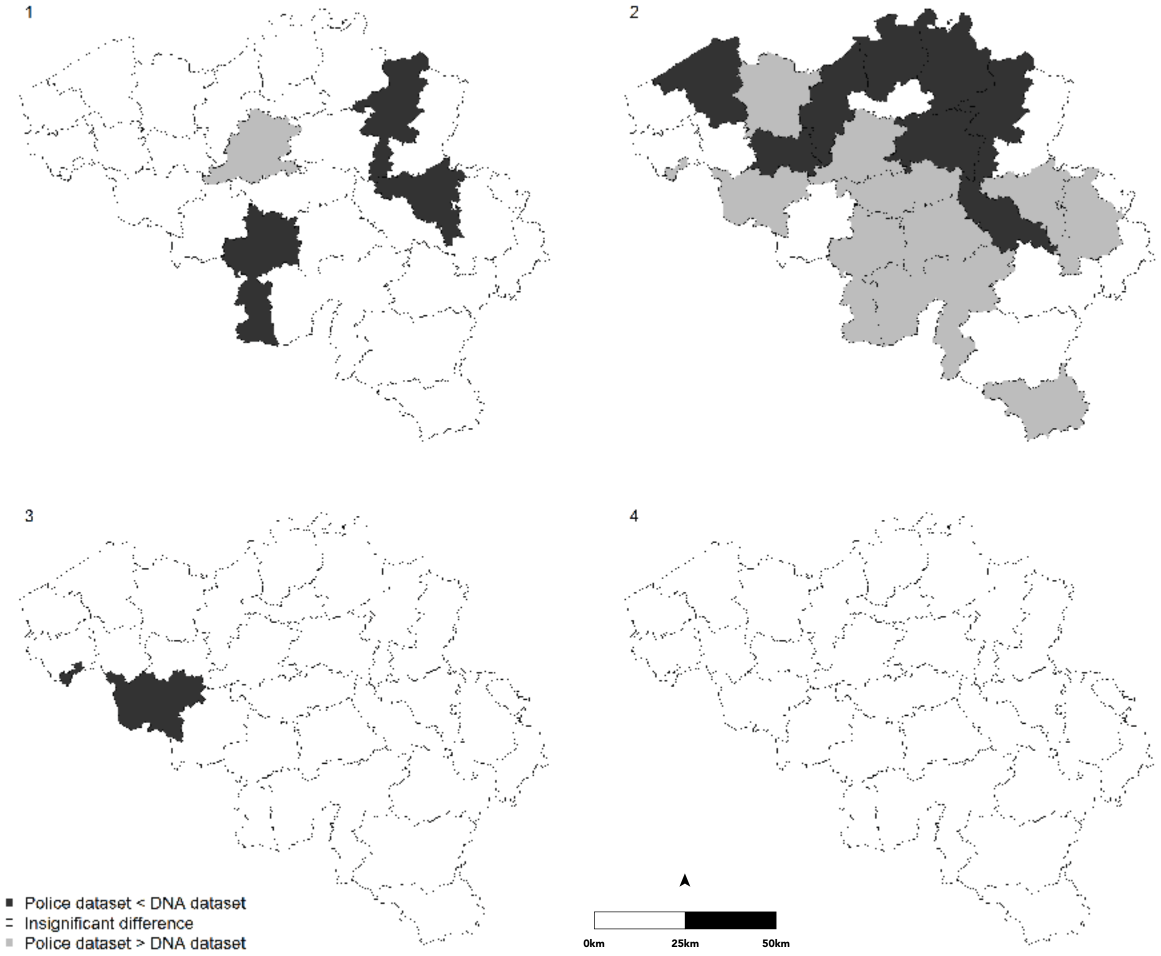


Figure 31: Local S-indices for violent theft (1), aggravated burglary (2), lethal violence (3), and sexual offences (4) for 2011

2012

Table 25: Description of datasets for 2012

Crime type	Police dataset		DNA dataset	
	N	%	N	%
Violent theft	365	16.86	21,120	10.58
Aggravated burglary	1,646	76.03	170,659	85.49
Lethal violence	31	1.43	450	0.23
Sexual offences	123	5.68	7,390	3.70
Total	2,165	100	199,619	100

Table 26: Generalized Gini coefficients for 2012

2012	Police dataset				DNA dataset		
	G’	S.E. G’	Lower CI	Upper CI	G’	S.E. G’	p
4 crime types	0.5654806	0.07613442	0.4397684	0.6967595	0.5386708	0.05907063	0.48
Violent theft	0.7173295	0.09153266	0.5541703	0.8422600	0.6015221	0.04519617	0.277
Aggravated burglary	0.5538508	0.07266402	0.4396091	0.6838069	0.5372846	0.06888008	0.464
Lethal violence	0.5405761	0.05736531	0.4496932	0.6374776	0.7861410	0.1349444	< 0.001
Sexual offences	0.4282163	0.0564156	0.3351881	0.5615682	0.6702800	0.06443557	< 0.001

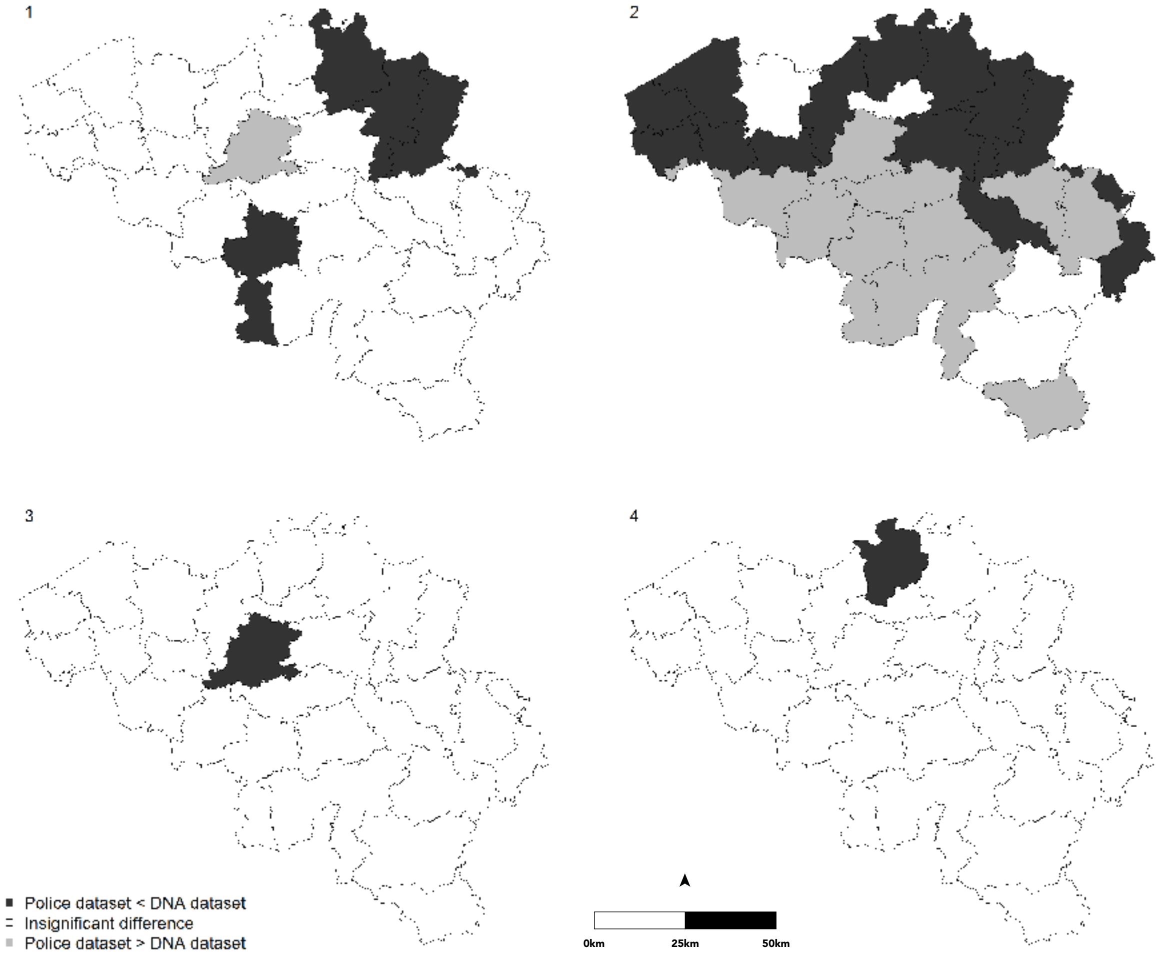


Figure 32: Local S-indices for violent theft (1), aggravated burglary (2), lethal violence (3), and sexual offences (4) for 2012

2013

Table 27: Description of datasets for 2013

Crime type	Police dataset N	 %	DNA dataset N	 %
Violent theft	267	13.17	18,366	9.78
Aggravated burglary	1,619	79.87	161,122	85.81
Lethal violence	34	1.68	428	0.23
Sexual offences	107	5.28	7,855	4.18
Total	2,027	100	187,771	100

Table 28: Generalized Gini coefficients for 2013

2013	Police dataset				DNA dataset		
	G’	**S.E. G’**	**Lower CI**	**Upper CI**	**G’**	**S.E. G’**	**p**
4 crime types	0.5600502	0.07286328	0.4400864	0.6946747	0.5363153	0.05334956	0.493
Violent theft	0.7081523	0.0764994	0.5865243	0.8305804	0.6084062	0.05128875	0.237
Aggravated burglary	0.5508074	0.0710741	0.4365470	0.6861854	0.5367740	0.05523458	0.454
Lethal violence	0.5503634	0.06477753	0.4412698	0.6738913	0.7712418	0.2044554	< 0.001
Sexual offences	0.4330056	0.05851932	0.3388157	0.5645880	0.6964348	0.06668315	< 0.001

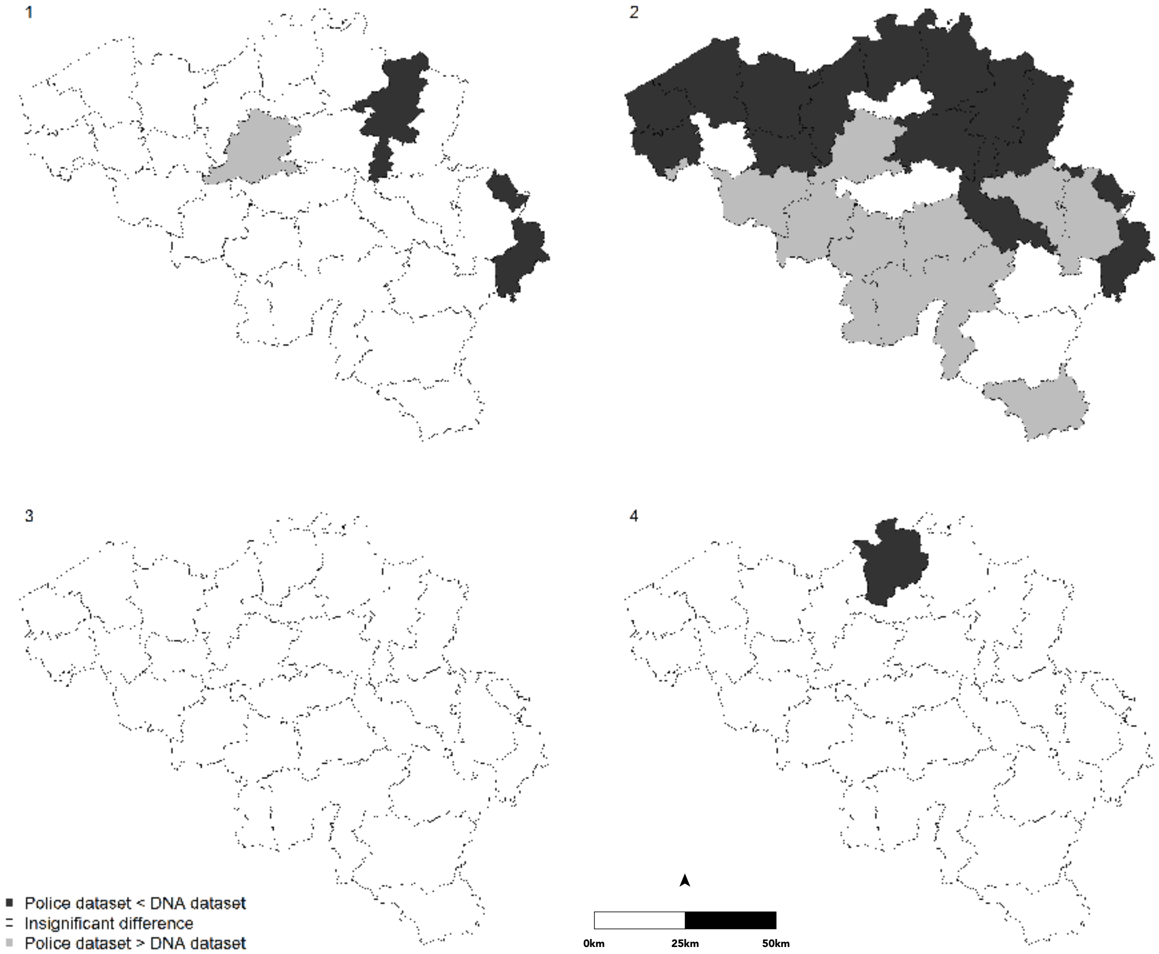


Figure 33: Local S-Indices for violent theft (1), aggravated burglary (2), lethal violence (3), and sexual offences (4) for 2013

2015

Table 29: Description of datasets for 2015

Crime type	Police dataset		DNA dataset	
	N	%	N	%
Violent theft	140	12.95	14,851	11.16
Aggravated burglary	848	78.45	112,017	84.17
Lethal violence	19	1.76	443	0.33
Sexual offences	74	6.85	5,771	4.34
Total	1,081	100	133,082	100

Table 30: Generalized Gini coefficients for 2015

2015	Police dataset				DNA dataset		
	G’	S.E. G’	Lower CI	Upper CI	G’	S.E. G’	p
4 crime types	0.5583269	0.06972365	0.4461923	0.6908841	0.5786823	0.04755035	0.243
Violent theft	0.7206249	0.08128673	0.5838221	0.8361241	0.6380952	0.0602213	0.32
Aggravated burglary	0.5462193	0.06645975	0.4395158	0.6769351	0.5993187	0.05451599	0.088
Lethal violence	0.5494524	0.06572644	0.4384630	0.6676429	0.7423823	0.152775	< 0.001
Sexual offences	0.4241771	0.05589142	0.3354737	0.5531660	0.6516517	0.06562113	< 0.001

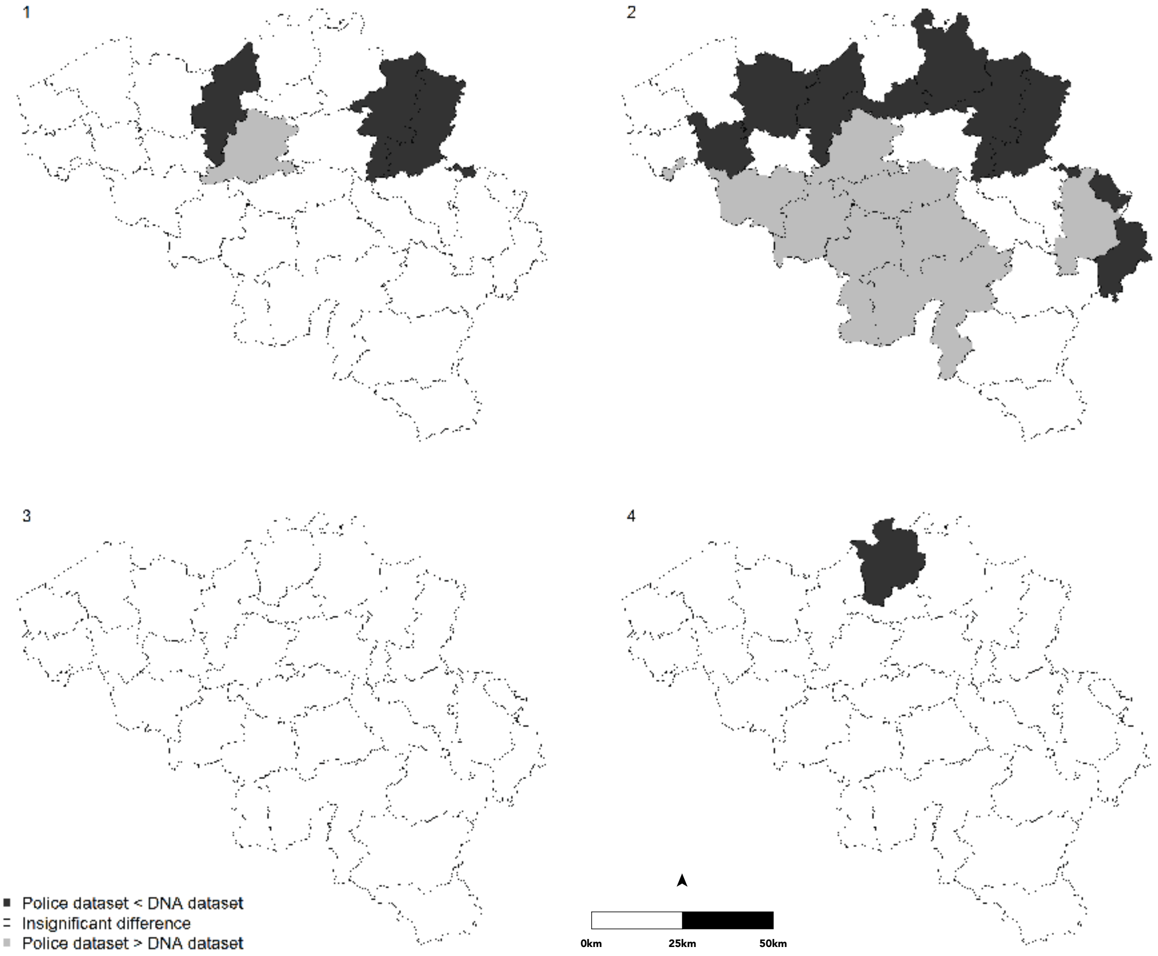


Figure 34: Local S-indices for violent theft (1), aggravated burglary (2), lethal violence (3), and sexual offences (4) for 2015