Forensic Science International 262 (2016) 1–10

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

To analyse a trace or not? Evaluating the decision-making process in the criminal investigation

Sonja Bitzer [a,*], Olivier Ribaux [a], Nicola Albertini [b], Olivier Delémont [a]

[a] Ecole des Sciences Criminelles, Université de Lausanne, Batochime, CH-1015 Lausanne, Switzerland
[b] Police Cantonale Vaudoise, Service d'Identité Judiciaire Centre Blécherette, 1014 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 18 November 2015
Received in revised form 22 January 2016
Accepted 10 February 2016
Available online 18 February 2016

Keywords:
Forensic science
DNA analysis
Decision tree
Classification

ABSTRACT

In order to broaden our knowledge and understanding of the decision steps in the criminal investigation process, we started by evaluating the decision to analyse a trace and the factors involved in this decision step. This decision step is embedded in the complete criminal investigation process, involving multiple decision and triaging steps.

Considering robbery cases occurring in a geographic region during a 2-year-period, we have studied the factors influencing the decision to submit biological traces, directly sampled on the scene of the robbery or on collected objects, for analysis. The factors were categorised into five knowledge dimensions: strategic, immediate, physical, criminal and utility and decision tree analysis was carried out.

Factors in each category played a role in the decision to analyse a biological trace. Interestingly, factors involving information available prior to the analysis are of importance, such as the fact that a positive result (a profile suitable for comparison) is already available in the case, or that a suspect has been identified through traditional police work before analysis. One factor that was taken into account, but was not significant, is the matrix of the trace. Hence, the decision to analyse a trace is not influenced by this variable.

The decision to analyse a trace first is very complex and many of the tested variables were taken into account. The decisions are often made on a case-by-case basis.

1. Process and decision-making steps

The criminal investigation process needs to be considered as a whole, with the traces collected at the crime scene and the reconstruction (i.e. micro-sequence of events [1]) going back to the events at this same crime scene [2]. Kind [3] and Brodeur [4] suggest two similar models of the criminal investigation process. The latter has a more detailed view of the investigative phase, dividing it into a triplet of identification of the author of the crime, locating the suspect, concluding with structuring of the evidence. The former differentiates three "chapters": (1) the problem to find, (2) refinement, checking and preparation for trial, and, finally (3) the problem to prove. This distinction is mainly based on a difference in inferential reasoning within each "chapter" and gives rise to different ways of using traces. In the first "chapter", the logical process starts from the traces leading to the suspect(s), a mainly abductive approach. Whereas, once a suspect is apprehended, the reasoning process becomes mainly deductive, starting from the case in order to explain the occurrence of these particular traces. This distinction in different phases of the investigation contributes to the manifold roles forensic science plays in the criminal investigation process [5].

We propose to adopt a complementary perspective considering the practice of forensic science within the investigation process as a decision-making process. The complete process can be divided into several decision steps, some of which are closely linked or even intertwined. In our view, the following key decision steps should be recognised:

1. the decision to attend the crime scene and search for traces
2. the decision to collect traces
3. the decision to analyse traces
4. the decision to use traces in the inquiry
5. the decision to collate trace-related information in a structured database
6. the decision to use traces in court

* Corresponding author. Tel.: +41 21 692 46 21; fax: +41 21 692 46 05.
E-mail address: sonja.bitzer@unil.ch (S. Bitzer).

http://dx.doi.org/10.1016/j.forsciint.2016.02.022
0379-0738/

Before the search begins, the question of whether a crime scene investigator *attends a crime scene* or not has to be answered (point 1). A first and more often than not latent triaging step occurs already at this stage, as crime scenes that are not attended cannot be sources of traces, and thus, clues or evidence [6,7]. The first phase of the criminal investigation consists of the problem to "find", or rather, the search for traces (point 2). This search needs to be systematic, based on cognitive skills such as observation and understanding of the criminal and immediate environment and the traces [5,8]. This is another step that undergoes triaging; traces that are not known will not be looked for, detected nor collected. Hence, limiting the search for traces to only certain types of traces, excludes other types of traces from the investigation process. The result of the search of traces leads at best to their *detection*. In order to detect traces, the recognition as such is crucial. The detection of traces is not a decision step as such; there is no conscious decision-making regarding the detection of traces. However, the recognition of traces and their anticipation rely on cognitive abilities that are linked to personal skills, training and experience and these are factors influencing other decision steps. Then follows the *collection of traces*. Concerning visible traces, this decision is often based on the quality of the trace. A certain triaging is undertaken at this point, knowing, however, that you generally cannot go back to the crime scene at a latter point. This step concludes the crime scene investigation.

The next decision-making step is the question of whether to *analyse a trace* or not (point 3); this includes in-house treatment of traces (e.g. shoe marks, fingermarks) and submission to external laboratories (mainly for biological traces). One could argue that this decision is obsolete, as the collection of traces already serves as triaging step and the reason that a trace is collected is that it will be analysed and further exploited. For some traces this might be the case, or rather, the decision to analyse a trace is already anticipated at the moment of its collection. This would merely constitute a shift in the moment when the decision is made. However, in many cases, all collected traces are not analysed, or not all in the first instance. It is then necessary to decide which traces to analyse, even if it is the question of which traces to analyse first.

Using a trace, or rather a clue–the information gained from the trace by analysing it–in the investigation is closely linked to the question of whether to analyse a trace or not (point 4). An assumption is made that anticipating the use of a clue guides the decision to analyse a trace. Assuming this, the probability of obtaining a profile suitable for comparison (for the sake of simplicity, hereafter called "Positive result") will also influence the decision about which trace to analyse. Then follows the decision to *collate trace-related information in a structured database* in order to use the potential of this information for intelligence on criminal phenomena, repetitive crimes, etc. (point 5). This database is a representation of what is known, at a certain time, on the crime environment. This is why it is generally called the memory. It organises information on specific cases and on their relations in the perspective of making the best use of its potential for providing intelligence on repetitive crimes and crime problems. Indeed, beyond using traces in investigations, as a reaction to the occurrence of each single event (point 4), this information can also be used in more proactive style of policing. This means that traces contribute to the development of knowledge on crime problems (e.g. by aggregating cases through linking), allowing to anticipate further occurrence, and devise a global response. This is typical of intelligence-led policing framework [9,10]). The final step of the use of forensic science in the criminal justice process consists of its *use for court purposes* (point 6). In this context, constraints, reasoning, and decisions to be taken are of a very different nature. This is why the clues used in court are not necessarily the same as the ones that were used in the investigation. For instance, depending on legal systems, in the investigative phase, the standard that needs to be reached for the trace to be used as information does not need to be as high as when it has to be accepted for court.

The decision to analyse a trace, which is under scrutiny in this paper, is embedded in the described decision process. Understanding how this decision is made through determining which factors are involved in the decision to analyse a trace is the main objective of this paper and will be discussed first through a literature review. Subsequently, the methodology and data employed are specified and a model is suggested. Finally, the influencing variables are presented and discussed and tested through statistical analysis. This study has been undertaken in order to empirically determine which factors contribute to the decision to analyse a trace, and thus, raise questions about existing assumptions in the literature regarding effectiveness and efficiency as key drivers in decisions to analyse traces.

2. Factors affecting the decision to analyse traces

In several effectiveness measurement studies [11–13], the authors examined the contribution of forensic science at five different stages: crime scene attendance, evidence submission, analysis, identification, and arrest. They generally used success rates (in terms of number of cases where forensic evidence was present) and lead times as effectiveness indicators, and compared these indicators for fingermarks and biological traces. All or most of the filtering was done at the crime scene with the decision to collect traces, and (almost) 100% of these traces were then forwarded to the laboratory for analysis. Hence, no factors relating to the decision to analyse were studied in these organisational systems.

Often, strategic guidelines are established by policy makers, police or forensic managers. They are implemented through protocols and procedures for deciding which traces are analysed, stemming from financial and performance pressure [14]. These guidelines focus mainly on qualitative aspects of the trace: "rich" biological traces (blood, saliva, etc.) are preferably chosen for analysis over contact traces, especially in high-volume crimes [14]. However, when considering the case circumstances of these types of cases, "rich" biological traces are not the most recurrent traces, and also not necessarily the most promising in terms of utility (added value of information to the case, see [15]).

When police investigators were asked why they would use forensic science, the main reason given was to strengthen the case against a suspect [16–18]. Similarly, when asked about why traces were not submitted for analysis, the lack of a suspect was mentioned recurrently [19,20]. Furthermore, the presence of traces is not statistically significant for the arrest of a suspect, which is not surprising, as the studied group of crime scene investigators does not recognize the utility of the DNA database [16,21,22].

Ribaux et al. [7] proposed a deconstruction and formalisation of the first steps impacting on the decision process. In their model encompassing four environments, they outlined some of the factors that affect the first two decision steps preceding the analysis of a trace: the decision to attend a crime scene and to search for traces. The proposed model reunites the *strategic, criminal, physical and immediate* knowledge dimensions. These incorporate constraints on, and facilitators of, the decision steps, factors that either limit or promote attendance at the crime scene and the collection of traces. In our view, this model could be extended to the decision to analyse a trace. Indeed, some of the environments (i.e. the strategic and the physical environment) have already been mentioned to influence this decision, but such an holistic view has never been adopted. We suggest the addition of the *utility* dimension [15], which includes factors relating to information available to the decision-maker at the moment of the decision and the added-value of the information retrievable through the analysis of the trace. This adapted model will be tested on data about robbery cases and the variables are explained in the following section.

3. Data and method

3.1. Study data and methodology

We studied robbery cases[1] occurring in the canton Vaud (a region of Switzerland of 3212 km^2 and 760 861 inhabitants) between January 2012 and December 2013. More precisely, we examined the data recorded by the forensic unit of the police (state police) and considered robbery cases where at least one biological trace was collected. These cases can be very diverse, from the street robbery of a handbag to the highly organised and planned robbery of a jewellery store or bank. Robbery cases were chosen as compromise between high volume crimes, where the decision to analyse a trace does not involve much of a triaging decision based on reasoning about the usefulness of the clue (criteria used are much more global, in accordance with financial limits), and homicide cases. For such serious cases, all resources are invested and thus, again, the decision to analyse a trace is not considered important because all collected traces are analysed. The canton Vaud was chosen as study area for proximity reasons.

First, we performed an extraction of information from the police database: 102 cases were registered in the database of the forensic police unit for the type of event and the selected period. This corresponded to a total of 410 biological traces. In 12 and respectively 38 cases, fingermarks and shoe marks were also collected.

This dataset was completed with qualitative data through participant observation that was performed during 5 months at the forensic unit in the beginning of 2015. It was possible to follow two cases at the crime scene and another 20 were observed during the decision process at the laboratory. During this period, additional information about the cases was collected. This allowed integrating pieces of information that are typically not included in the structured scheme of the database, such as the case progress before submission for analysis (e.g. if a suspect was identified before the analysis of the trace) or the contamination of the crime scene by the victim after the events. This mixed methodology allowed for a more comprehensive account of the complete decision-making process, and particularly the decision to analyse a trace. The principal aim of the participant observation was to help further our understanding of why a trace was analysed (or not). Furthermore, the results could be more easily interpreted by having seen and followed cases from the start to the decision to submit traces for analysis.

3.2. Variables

The dichotomous dependent variable considered is the *Analysis of a trace*: whether the biological trace was analysed or not (1 = trace was analysed, 0 = trace was not analysed). In this study, we considered that the analysis of a biological trace consisted in its DNA profiling performed by an external DNA laboratory; visual examination or presumptive testing to inform the collection or the type of biological trace were not considered an analysis.

More often that not, the different traces collected in a given case are processed in successive batches of analyses: they are not all submitted at once to the external laboratory for DNA profiling. Therefore, we decided to adopt a double perspective in our study: a first model was considered taking into account the dependent variable (analysed traces) at the end of the investigation, after all analyses were performed. A second model was considered using the first batch of analysed traces in the case as dependent variable (the traces that were analysed in the first batch were coded 1, all the others were coded 0, even if they were analysed in a second, third, ... batch). This dual perspective was chosen in order to follow the sequence of the analyses and get an understanding of what factors were affecting the composition of the first batch of analyses.

The independent variables were separated into the four environments defined by Ribaux et al. [7] with the addition of a utility knowledge dimension, including previously available information (see Table 1). Some variables are case-specific, and thus the same for all the traces in the case, whereas others are trace-specific.

3.3. Analytical strategy

Classification modelling was performed in order to create models that best explain our data. Recursive partitioning methods, including decision tree- and rule-based models, were used to follow the decision steps and highlight and select the important factors in our model. Details about the chosen statistical procedure can be found in Kuhn and Johnson [23]. These algorithms split the data in multiple steps in order to discriminate a maximum of the observations. Each dataset was split until the remaining subsamples consisted mainly of one group (i.e. classification). These classification modelling algorithms had several advantages, such as the possibility of integrating various types of variables (binary, ordinal, continuous) into the model and little influence by extreme values or by missing data. Simple decision tree modelling algorithms, like CART, j48, and single decision tree C5.0, were used and compared with performance results of more complex models, like boosted C5.0, rule-based PART, bagged trees (treebag) and boosted trees (gbm). The raw data were extracted from the operational database of the forensic police unit and were completed with information retrieved from audition reports. Microsoft® Excel was used to collect and codify the data and the open source software R® was used for further statistical modelling.

For model comparison purposes, the dataset was partitioned in a training set and a test set (split ratio = 0.8), which were recursively resampled ($n = 100$), in order to evaluate the models. A correlation matrix was constructed, and the highly correlated variables (threshold = >0.75) were removed. The performance of the models was assessed through ROC (*receiver operating characteristic*) curves and their corresponding area under the curve (AUC) values, sensitivity and specificity. The chosen classification algorithm was then applied to the complete dataset.

4. Results and discussion

4.1. Descriptive analysis of the dataset

Overall, the analysis rate of collected biological traces (i.e. the rate of biological traces submitted for analysis per collected traces) was .74, with .56 of biological traces analysed in a first batch. By comparison, for fingermarks and shoe marks, the analysis rates reached .90 (.87 and .93% respectively). The analysis decisions for fingermarks and shoe marks were not studied in detail, as around

[1] Robbery is legally characterised by Art. 140 Swiss Criminal Code as follows: “Any person who commits theft by using force on another, threatening another with imminent danger to life or limb, or making another incapable of resistance”. We use it as a definition for delineating the kind of events we aim at covering. More precisely, cases that were first reported as being robbery cases were studied. In fact, in 10% of these cases, the legal classification changed later in the course of the investigation and the suspect was finally charged for a different offense (e.g. burglary, aggression, misleading the judicial authorities).

Table 1
Description of independent variables. Variables in italic are not available for the second model (first batch of analyses). For the sake of simplicity, the term "identification" of the suspect will be used here. However, this should not be understood in an evaluative/interpretative manner. For the nominal variables, dummy variables were created.

Variable	Type	Case or trace-specific	Description
Strategic			
Type of intervention	binary	case	Crime scene investigated by forensic unit (predictor) or objects brought to laboratory by police first responder
Prosecutor in charge of the case	binary	case	A prosecutor was in charge of the case, and thus in charge of DNA analyses. In the investigated area, the police start the inquiry. In some situations, a prosecutor is introduced in the procedure.
Team of inspector in charge	nominal	case	Crime scene investigators at this forensic unit are separated into 5 teams
Number of collected biological traces	ordinal	case	Number of collected biological traces registered in the database for the case
Immediate			
Type of target	nominal	case	Three categories: business (i.e. jewellery store, post office), service (i.e. café, bar, restaurant) and private (i.e. apartment, street, parking lot)
Surveillance camera	binary	case	Predictor: surveillance camera images available, no information about the quality of these images
Witness report	binary	case	Predictor: witness report available, no information about the quality of this report
Armed robbery	binary	case	Predictor: a weapon (gun, knife, ...) was used
Violence against victim	binary	case	Predictor: the victim was hurt by the offender(s)
Number of offenders	ordinal	case	Number of offenders described in the summary of the case
Number of collected shoe marks	ordinal	case	Number of collected shoe marks registered in the database for the case
Number of collected fingermarks	ordinal	case	Number of collected fingermarks registered in the database for the case
Physical			
Type of biological trace	binary	trace	"Rich" biological trace (blood or saliva; predictor) or contact trace
Presumptive testing	binary	trace	Predictor: presumptive testing was performed
Matrix of the trace	continuous	trace	Probability of obtaining a positive result (which can be used for comparison) on a specific matrix in the year previous to the decision about analysis of this trace, compiled from all analyses of biological traces in the database (for all types of crimes)
Criminal			
Known crime link	binary	trace	Predictor: crime link is known, as registered in the database, the link is known before the analysis of the trace
Utility			
Suspect identification through police inquiry	binary	case	Police inquiry led to the identification of a suspect before analysis of the trace
Number of biological traces analysed previously in the case	*ordinal*	*trace*	*Number of biological traces analysed in the case previously to the trace in question*
Positive result available	*binary*	*trace*	*Positive result (a profile suitable for comparison) available within the specific case before analysis of the trace*
Identification available through biological trace analysis	*binary*	*trace*	*Suspect identification by DNA analysis before analysis of the trace*
Identification available through other traces	binary	trace	Suspect identification by other types of traces before analysis of the trace
Identification available total	*binary*	*trace*	*Identification by DNA or other types of traces before analysis of the trace*
Suspect identification available before analysis	*binary*	*trace*	*Suspect identification through DNA analysis available before analysis of the trace*

90% turn-around was determined and these analyses were performed in-house. The main reasons for the non-analysis of fingermarks or shoe marks were quality and/or redundancy: either the trace was considered to be of poor quality or the trace existed already in the case and was thus not further processed. Due to the high analysis rates for these types of traces, the variables *Number of collected fingermarks* and *Number of collected shoe marks* can be considered to be equivalent to the *Number of analysed fingermarks* and *Number of analysed shoe marks*.

The first step of our study consisted of a descriptive statistical analysis of the raw data for each variable that considered the perspectives of both dependent variables: all the analysed traces (*General set*) and only the traces that were part of the first submitted batch (*First batch analysis*). Table 2 presents the breakdown of the analysed and non-analysed traces.

4.2. Comparison and selection of classification modelling

Two groups of variables were identified as being correlated. The first group (indicated † in Table 2) was constituted of two variables from the physical dimension and the second group (indicated # in Table 2) of 5 variables from the utility dimension. From these groups, the least correlated variable(s) remain(s) for the construction of the models.

Different classification models were applied to the resampled training and the corresponding test sets. The application of different classification models aimed at assessing whether the simpler models could compare in terms of performance and accuracy with more complex ones. Overall, the performance values (AUROC; Area under ROC measured through a combination of specificity and sensitivity) were very good for all the models, especially with the general set (see Fig. 1). The best classification performance of our data, i.e. highest AUROC values, were obtained with the boosted trees (mean of AUROC: 0.9729) for the general dataset, and the boosted C5.0 trees (mean of AUROC: 0.8299) for the first batch analysis.

As the performance of the simpler C5.0 single decision tree model compares well to the more complex models, the C5.0 single tree decision model was finally chosen for further analysis (mean of AUROC: 0.9328 resp. 0.7556), as the outcome is more easily

Table 2
Descriptive statistics on the number of trace analyses and non-analyses ($N = 410$). Two groups of variables were identified as being correlated (indexed with † and #); of these groups only those marked with (*) remain for the construction of the models.

Variable	n	General set: Analysis of trace n (%)	General set: Non analysis of trace n (%)	First batch analysis: Analysis of trace n (%)	First batch analysis: Non analysis of trace n (%)
Strategic					
Type of intervention					
Crime scene intervention	353	258 (73.09)	95 (26.91)	187 (52.97)	166 (47.03)
Objects brought to lab	57	44 (77.51)	13 (22.81)	42 (73.68)	15 (26.32)
Prosecutor in charge of the case					
yes	345	247 (71.59)	98 (28.41)	184 (53.33)	161 (46.67)
no	65	55 (84.62)	10 (15.38)	45 (69.23)	20 (30.77)
Inspector in charge – Team					
a	91	79 (86.81)	12 (13.19)	63 (69.23)	28 (30.77)
b	67	31 (46.27)	36 (53.73)	27 (40.30)	40 (59.70)
c	162	113 (69.75)	49 (30.25)	76 (46.91)	86 (53.09)
d	60	53 (88.33)	7 (11.67)	39 (65.00)	21 (35.00)
e	30	26 (86.67)	4 (13.33)	24 (80.00)	6 (20.00)
Number of collected biological traces					
1	36	33 (91.67)	3 (8.33)	33 (91.67)	3 (8.33)
2	48	36 (75.00)	12 (25.00)	31 (64.58)	17 (35.42)
3	27	23 (85.19)	4 (14.81)	20 (74.07)	7 (25.93)
4	12	7 (58.33)	5 (41.67)	3 (25.00)	9 (75.00)
5	34	24 (70.59)	10 (29.41)	23 (67.65)	11 (32.35)
> 5	253	179 (70.75)	74 (29.25)	119 (47.04)	134 (52.96)
Immediate					
Target					
Business	163	121 (74.23)	42 (25.77)	91 (55.83)	72 (44.17)
Service	73	53 (72.60)	20 (27.40)	30 (41.10)	43 (58.90)
Private	174	128 (73.56)	46 (26.44)	108 (62.07)	66 (37.93)
Surveillance camera					
yes	102	71 (69.61)	31 (30.39)	57 (55.88)	45 (44.12)
no	308	231 (75.00)	77 (25.00)	172 (55.84)	136 (44.16)
Witness report					
yes	335	261 (77.91)	74 (22.09)	202 (60.30)	133 (39.70)
no	75	41 (54.67)	34 (45.33)	27 (36.00)	48 (74.00)
Armed robbery					
yes	272	190 (69.85)	82 (30.15)	136 (50.00)	136 (50.00)
no	138	112 (81.16)	26 (18.84)	93 (67.39)	45 (32.61)
Violence against victim					
yes	214	144 (67.29)	70 (32.71)	114 (53.27)	100 (46.73)
no	196	158 (80.61)	38 (19.39)	115 (58.67)	81 (41.33)
Number of offenders					
1	91	63 (69.23)	28 (30.77)	57 (62.64)	34 (37.36)
2	134	107 (79.85)	27 (20.15)	82 (61.19)	52 (38.81)
3	157	109 (69.43)	48 (30.57)	76 (48.41)	81 (51.59)
> 3	28	23 (82.14)	5 (17.86)	14 (50.00)	14 (50.00)
Number of collected shoe marks					
0	150	121 (80.67)	29 (19.33)	103 (68.67)	47 (31.33)
1	89	53 (59.55)	36 (40.45)	37 (41.57)	52 (58.43)
2	66	39 (59.09)	27 (40.91)	32 (48.48)	34 (51.52)
> 2	105	89 (84.76)	16 (15.34)	57 (54.29)	48 (45.71)
Number of collected fingermarks					
0	324	262 (80.86)	62 (19.14)	198 (61.11)	126 (38.89)
1	8	4 (50.00)	4 (50.00)	2 (25.00)	6 (75.00)
2	44	17 (38.64)	27 (61.36)	16 (36.36)	28 (63.64)
> 2	34	19 (55.88)	15 (44.12)	13 (38.24)	21 (61.76)
Physical					
Type of biological trace†,*					
"Rich" biological trace	91	43 (47.25)	48 (52.75)	30 (32.97)	61 (68.03)
Contact trace	319	259 (81.19)	60 (18.81)	199 (62.38)	120 (37.62)
Presumptive testing†					
yes	74	29 (39.19)	45 (70.81)	21 (28.38)	53 (71.62)
no	336	273 (81.25)	63 (18.75)	208 (61.90)	128 (38.10)
Matrix of the trace					
Propitious (≥ 0.5)	296	209 (70.61)	87 (29.39)	155 (52.36)	141 (47.64)
Non propitious (< 0.5)	114	93 (81.58)	21 (18.42)	74 (64.91)	40 (35.09)
Criminal					
Known crime link					
yes	63	54 (85.71)	9 (14.29)	26 (41.27)	37 (58.73)
no	347	248 (71.47)	99 (28.53)	203 (58.50)	144 (41.50)
Utility					
Suspect identification through police inquiry					
yes	76	33 (43.42)	43 (56.58)	26 (34.21)	50 (65.79)
no	334	269 (80.54)	65 (19.46)	206 (61.68)	128 (38.32)
Number of biological traces analysed previously in the case#,*					
0	256	232 (90.63)	24 (9.37)	*NA*	*NA*

Table 2 (*Continued*)

Variable	n	General set: Analysis of trace n (%)	General set: Non analysis of trace n (%)	First batch analysis: Analysis of trace n (%)	First batch analysis: Non analysis of trace n (%)
> 0	154	70 (45.45)	84 (54.55)	*NA*	*NA*
Positive result available#,*					
yes	107	29 (27.10)	78 (72.90)	*NA*	*NA*
no	303	273 (90.10)	30 (9.90)	*NA*	*NA*
Identification available through biological trace analysis#					
yes	89	22 (24.72)	67 (72.28)	*NA*	*NA*
no	321	280 (87.23)	41 (12.78)	*NA*	*NA*
Identification available through other traces					
yes	51	23 (45.10)	28 (54.90)	11 (21.57)	40 (78.43)
no	359	279 (77.72)	80 (22.28)	218 (60.72)	141 (39.28)
Identification available total#					
yes	116	42 (36.21)	74 (63.79)	*NA*	*NA*
no	294	260 (88.44)	34 (11.56)	*NA*	*NA*
Suspect identification available before analysis#					
yes	101	35 (34.65)	66 (65.35)	*NA*	*NA*
no	309	267 (86.41)	42 (13.59)	*NA*	*NA*

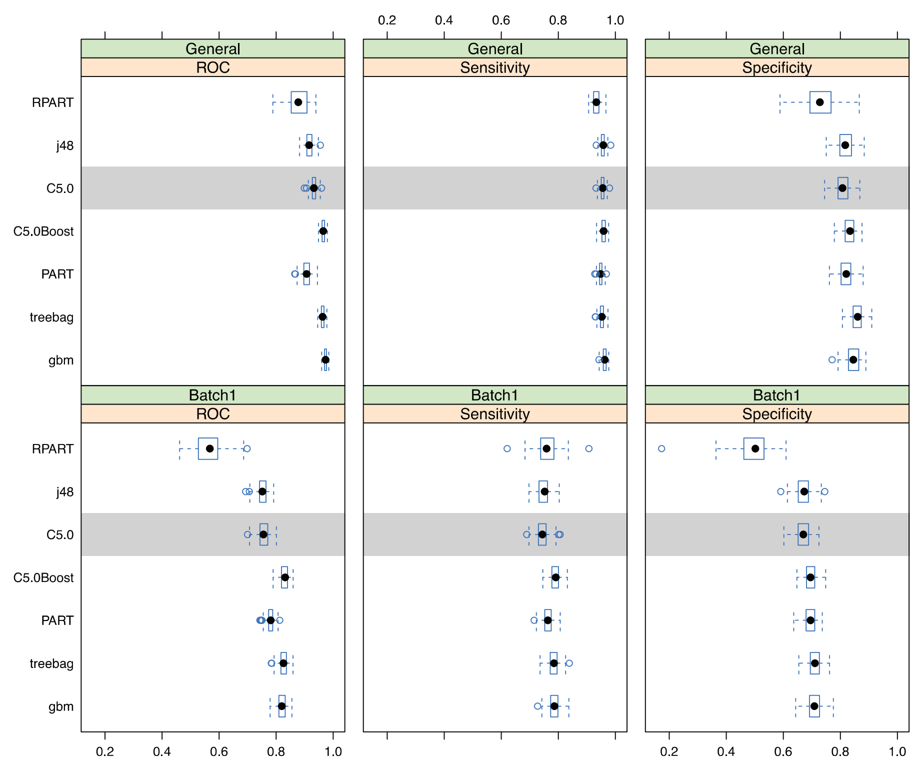

Fig. 1. Comparison of performance values (AUROC, sensitivity and specificity values) for all tested models using resampled training and test sets ($n = 100$) for general set and first batch analysis. The highlighted models were chosen for classification modelling, as a compromise between good performance results and simplicity of interpretation.

interpretable. This choice was adopted as the main aim of the study was to understand and model the decision process and we are not striving for excellent predictive power. Another advantage of the C5.0 decision tree model is that a visualisation of the model can be extracted, as opposed to the bagged or boosted trees for instance.

4.3. General set of data

From the 25 independent variables that were considered, 9 were used to construct the model (see Fig. 2). Thus, these 9 factors contribute to the decision to analyse a trace, in the sense that the splits created by these factors lead to the constitution of smaller, more homogeneous groups. In the first model, considering all the analysed traces, these are mainly variables related to previously available information, such as the factor that a positive result is already available within the specific case (which is correlated to other factors involving previous knowledge, as shown in Table 2), or that the suspect has been identified through police inquiry before analysis. However, forensic factors also contribute to the decision to analyse a trace, such as forensic intelligence and the number of collected biological traces. Furthermore, the decision appears to also depend upon individuals as the team of investigators in charge of the case influences the model.

It is interesting to emphasise that factors including previous information affect the decision to analyse a trace. Thus, the decision is not solely based on purely qualitative factors such as the type of biological trace. The knowledge of traces and of their analysis outcomes plays a significant role. The first factor is the

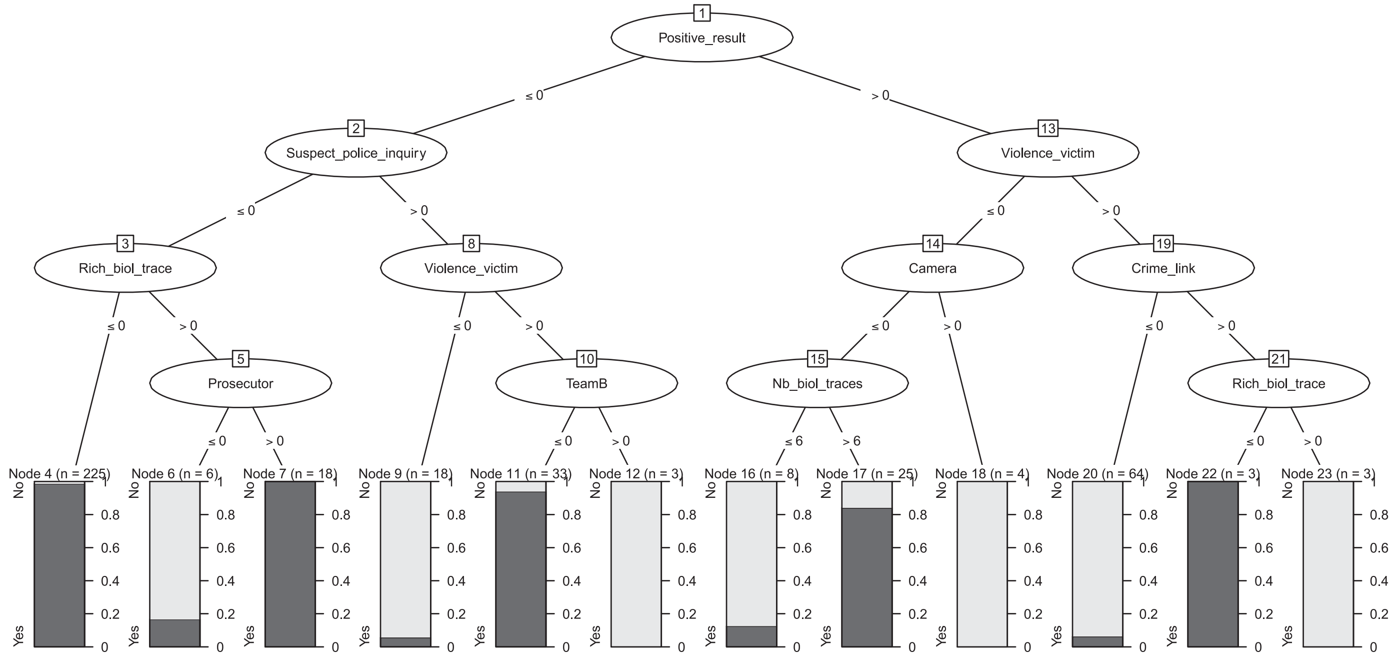

Fig. 2. C5.0 decision tree model of the general data set.

knowledge about a previous positive result (a DNA profile suitable for comparison) already available within the specific case (node 1, *Positive_result* in Fig. 2); hence, crime scene investigators take into account the results of previous analyses performed in the same case in their decision to analyse a biological trace. As this is the model for the "looking back" scenario, it is not surprising that some traces remain unanalysed when others have delivered a positive result. In a case with multiple biological traces, some traces might have delivered a usable profile or even an identification, hence the analysis of the remaining traces becomes less prone to provide new, useful information, and will thus not be performed. However, it needs to be emphasised that this positive result did not necessarily lead to an identification of a suspect. These two variables are correlated, and as a consequence, the latter was not included in the model, however, they are not fully equivalent.

When 'no positive result' was available (i.e. either no DNA analysis has been performed, or, the result of the analysis did not yield a profile that is of sufficient quality to be used for comparison), traditional police work was of importance (node 2, *Suspect_police_inquiry* in Fig. 2). Indeed, this could be observed during the participant observation, as a crime scene investigator preferred to wait for the police inquiry and their results, before deciding whether to analyse a trace or not. These results contradict one of the predominantly mentioned reasons for analysing a trace, which was to build a case against the suspect [14–16]. In the studied forensic unit, this reason could not be corroborated. If a suspect was identified prior to the analysis, the analysis rate of biological traces dropped (see Table 2). The main utility of the trace appears to be to gather intelligence rather than to produce evidence. The potential of a biological trace–by comparison with a reference database–to "provide" a name is widely understood. Contrary to the reasons found in the literature for the non-submission for analysis of a trace, the crime scene investigators exploited the trace's potential to "provide" a lead in the inquiry. Interestingly, the causal link appears to be in the opposite direction, compared to the studies that analysed the predictive effect of the presence of traces on arrest. It is not the presence of traces that is predictive of arrest, but arrest that is predictive of the analysis of traces. Hence, it appears very likely that the crime scene investigators consider the utility of the clue, in context with the available information, to form their decision about the traces to analyse [15].

The descriptive model highlights that contact traces are preferably chosen for analysis over "rich" biological traces (nodes 3 and 21, *Rich_biol_trace* in Fig. 2). At first, this finding appears counterintuitive as "rich" biological traces are more likely to yield a positive result in purely analytical terms. This is also contradictory to some of the analysis strategies in place: often the type of biological trace is the triaging factor (e.g. "rich" biological traces are preferably chosen for analysis over contact traces) in order to maximise the rate of profiles suitable for comparison [14], more often than not in response to financial pressure. However, when looking at the case level and the information conveyed by the traces, the preference for contact biological traces makes a lot of sense. The reason for this seems to be that if "rich" biological traces are found (blood or saliva in these cases), the investigators inferred that these traces more likely originated from the victim (who was hurt in some cases) and are thus of lower utility to their case. By consequence, and very logically, they were not submitted for analysis. Furthermore, on recovered bottles or cans, the investigators infer the presence of saliva without performing presumptive testing. As the results of these tests directly define the assertion of presence of a "rich" biological trace, in such situation this latter variable was coded as 1 for *Rich_biol_trace* but 0 for *Presumptive_test*.

In inquisitorial justice systems, the collaboration between the police and a 'prosecutor' is of high importance. This is because prosecutors have a broad range of competencies in deciding how investigations are conducted, and to commission expertise and analysis. As noticed during the participant observation, the involvement of a prosecutor was perceived very differently between the crime scene investigators. One crime scene investigator decided to analyse 3 out of 4 recovered biological traces in "priority mode" (analysed the same day as the others, but first on the list), in order to be able to quickly give the prosecutor results, and advice on the utility of the first 3 traces versus the fourth one, which he considered irrelevant. Other crime scene investigators work in close cooperation with the prosecutor and the decision

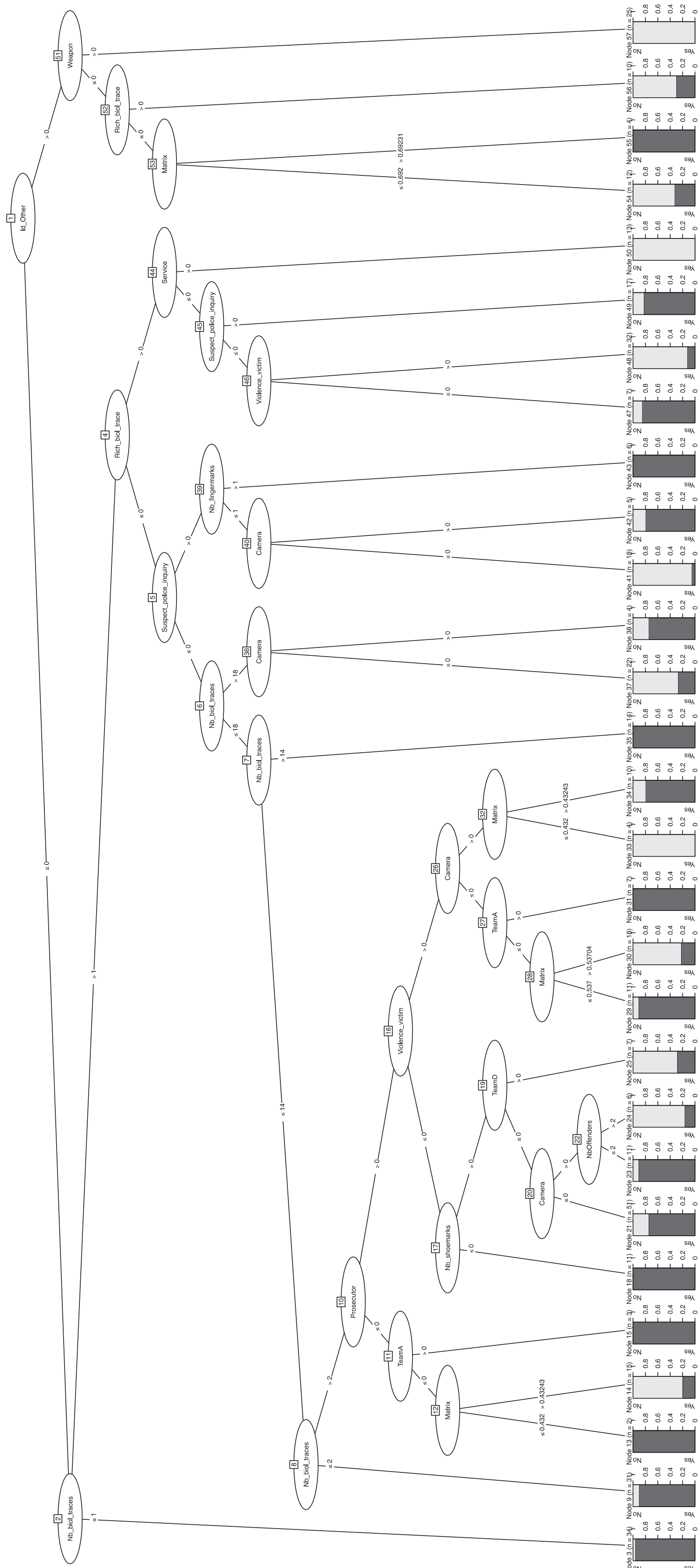

Fig. 3. C5.0 decision tree model of the first batch analysis.

about which traces to analyse arises though a collaborative approach. In the explicative model, the predictor *Prosecutor* is only important when no other results or police inquiry information are available, and when it is a "rich" biological trace. When the prosecutor is in charge of the case, and also of the financial part of the analyses, then all the biological traces are analysed. When there is no prosecutor in charge of the case, the proportion of analysed traces is much smaller. The same explanation as previously mentioned could be given: the prosecutor decides to analyse "rich" biological traces that the investigator inferred originated from the victim. At node 7, 17 out of 18 biological trace analyses resulted in a positive result, out of which 8 delivered a match with a suspect. The remaining 9 did not provide a match with the database, however, the victim cannot be excluded as being at the source of these traces, as a reference profile was not necessarily collected, analysed and compared.

Crime intelligence appears to play a role in the decision to analyse a trace in certain situations (node 19, *Crime_link* in Fig. 2). During the participant observation, it could be determined that the knowledge of a link with another crime would influence the decision to attend the scene, but also the analysis of traces. Indeed, a police investigator contacted the forensic unit in order to get information about traces that remained for analysis in the linked cases.

In addition to considering the variables that have been included in the decision tree, it is also noteworthy to emphasise the variables that were excluded by the algorithm, as no additional information would have been provided by the inclusion of these variables. One factor that is often used in analysis strategies is the matrix on which the trace was deposited, and hence the likelihood of obtaining a positive result from such a matrix. This variable was not considered important in the decision tree for the general dataset. Hence, the decision to analyse a trace is not influenced by this variable. It needs to be highlighted here, that most of the figures for the probability of obtaining a positive result are stable throughout the analysed period. However, for some matrices, it seems to be very difficult to make the right choice (decide for the analysis of that particular trace when the matrix is propitious for yielding a positive result), as the probability of obtaining a positive result varies a lot.

Similarly, the variable considering the number of offenders is not retained in the decision tree model. One would hypothesise that the higher the number of offenders the more traces would be analysed per case, and thus the higher the likelihood of analysis of the trace. However, this does not reflect the actual situation.

4.4. The decision to analyse a trace first

The model created with the dataset considering the first batch of traces analysed as dependent variables is more complex than the previous one (see Fig. 3). The very first decision regarding the analysis of biological traces, i.e. the decision for the first batch of analysis, does not seem to be straightforward; 15 out of the 23 available variables were used to construct the model. Again, factors related to individuals (for instance membership to a certain team of crime scene investigators) are affecting the decision to analyse a trace, as well as the knowledge about identification in the case, this time through a different type of forensic trace, and the police inquiry progress. In addition to the important factors highlighted in the previous model, the number of offenders, the type of robbery (armed robbery), the matrix of the trace, the number of collected fingermarks and shoe marks, and the targeted location (i.e. service) affected the decision regarding the first batch of analysis.

The variable in the first node is the presence of an identification through other traces in the case. This information can be available through fingermarks but also through a shoe mark comparison after the suspect has been identified through police inquiry. The decision to analyse a biological trace thus depends on the results delivered by other types of traces, often considered of lesser importance.

When no identification has been obtained and only one biological trace was collected in the case (node 3), almost all of them are analysed. This result is not surprising, as it is the sole biological trace they have in the case, and thus no triaging needs to be done.

In armed robberies (node 51, with an identification available through other traces), no biological traces were analysed. This result seems counterintuitive, as the case is considered more serious when a weapon is used. However, this finding is the reflection of the situation: it can be explained due to the circumstances of this kind of cases. When a weapon is used, more often than not, there is no contact between the offender and the victim/scene. Hence, fewer traces are considered useful to the case, and thus submitted for analysis.

However, some of the variables have different influences depending on the subsample. When comparing nodes 12 and 32, and their outcomes in nodes 13 and 14, respectively 33 and 34, it is noteworthy that the proportion of analysed biological traces is influenced in both directions by the same variable (the matrix of the trace) and a similar split criterion. The logical way would be to choose a higher number of traces for the analysis when the matrix is propitious for yielding a positive result. However, the decision to analyse a trace does not seem to always follow this logic (see also Table 2, a higher proportion of traces on less propitious matrices are analysed compared to the analysis rate of traces on propitious matrices), although this is one of the guidelines in this forensic unit.

In this model, information of about a known crime link was not considered to be important for the decision to analyse a first trace. It must however be considered that this variable is probably incomplete, in the sense that some observations were coded as 0, although a crime link was known, and thus, should have been coded as 1. However, this information was not available in the database. It has been however observed during the participant observation, that a series of cases with similar modus operandi is occasionally emphasised at daily meetings, but this information is not registered in the database, and can thus not be traced back.

5. Conclusion

To our knowledge, this is the first study of its kind that focuses on the decision-making process related to the decision to analyse a trace and tries to decipher the factors affecting it. For the sake of this study, we focused on robbery cases that are supposed not to be too deeply impacted by actuarial or managerial policy. A decision tree model was chosen to follow the decision paths to get an informed picture of the variables influencing this decision. When considering the decision tree and following the decision criteria through the tree, a variety of variables appear as important to the decision, for the general model but also, and especially, for the first batch of analysis. The results showed that all the suggested knowledge dimensions affect the decision to analyse a trace. Some, such as the criminal dimension, only act on a specific subsample within the decision tree. The utility dimension, referring to previous knowledge available in the case, is particularly important and interesting to note.

The results and the appreciation for the utility of the clues can be partly explained by the environment in which the analysed forensic unit is embedded: in this case, a decentralised police system with many contact points between scientific and police investigators, crime scene investigators who are more generalist than in other countries and thus more concerned about using police information and more sensitive to the question of utility.

In both settings, the general model and the first batch analysis, the knowledge of a positive result or of an identification through other traces affected the decision to analyse a trace in a statistically significant way. These findings are contradictory to the findings in the literature that emphasise that the main reason repeatedly mentioned for the analysis of a trace is to build a case against a suspect, and similarly, the main reason for the non-submission for analysis of a trace is that no suspect has been found. In the specific context of our study, we observed that the crime scene investigators took into account the knowledge about suspect identification through police inquiry or other types of traces for their decisions to submit biological traces for analysis, submitting fewer traces for analysis when a suspect was known. This suggests that they emphasise the potential of biological traces, in relation to a reference database, to gather intelligence and do not solely see the confirmation utility of biological traces as evidence to be presented for court purposes. Moreover, it appears from a detailed study of the cases that the biological traces were not solely used for identification purposes but also for the reconstruction of the event or the determination of the implication of different offenders.

One interesting finding relates to the nature of the trace itself. In both models, it appears that the matrix of the trace was not statistically significant in the decision to analyse the trace. Another factor relating to the trace itself is its type (''rich'' vs contact trace), assessed by the results of presumptive tests. While this factor was statistically significant, its influence was negative, in the sense that a ''rich'' biological trace is less likely to be submitted for analysis. Despite this observation seeming inconsistent at a first glance considering the possibility of obtaining a full DNA profile, it makes greater sense when taking into account the situation and the expected utility of the information conveyed by the trace. This is highly interesting from our perspective as both factors–the matrix of the trace and its ''quality''–are commonly used in guidelines or policy documents to support the decision to proceed to a DNA investigation. While these factors may indeed favour the chance to get an analytical result of good quality (DNA profile), they are not good predictors (on the contrary) of the utility of the information conveyed by the trace for the inquiry. In this sense, managerial policy, funded on apparent efficiency and taking into account only a narrow aspect of the multi-dimensional contribution of forensic science to the criminal investigation, while very popular and seemingly efficient, may have negative consequences.

Financial aspects, considered through the variable *Prosecutor*, appear to play only a limited role in the decision to analyse a trace, contrarily to the common belief that the decision is mostly economically driven. When the prosecutor is in charge of the case, he is also responsible for the financial side of the investigation, and hence the analyses of the traces. In the decision tree model of the general dataset, this variable appears only at the bottom, and hence only in very specific cases. Hence, efficiency is not considered a key variable in the decision-making process.

The decision to analyse (or not) a trace appears to be very complex, context dependent, limited by the situation and case-specific. Consequently, it seems very difficult, and scarcely relevant to establish rigid (managerial) guidelines for this decision step. On the basis of the results of our study, it appears that some variables are important under very specific circumstances, in subsamples in the trees. While it is of foremost importance to highlight that the explicative models and the results presented in this manuscript are highly dependent on the police structure where the study took place, and that they may not be valid in other environments, they however suggest that the different decision steps encompassed in the overall process of forensic science's contribution in the criminal investigation should be the subject of more scrutiny.

Acknowledgments

The authors would like to thank the forensic police unit of the canton Vaud (Switzerland) for their willingness and open-mindedness to collaborate for this study and their help in completing the dataset. In addition, the authors would like to thank the anonymous reviewers for their valuable comments.

References

[1] A.D. Barclay, Using forensic science in major crime inquiries, in: J. Fraser, R. Williams (Eds.), Handbook of Forensic Science, Cullompton, UK, 2009, pp. 337–358.
[2] Y. Schuliar, La coordination scientifique dans les investigations criminelles. Propositions d'organisation, aspects éthiques ou de la necessité d'un nouveau métier, Université Paris Descartes (2009).
[3] S.S. Kind, Crime Investigation and the Criminal Trial: A Three Chapter Paradigm of Evidence, J. Forensic Sci. Soc. 34 (3) (1994) 155–164.
[4] J.-P. Brodeur, L'enquête criminelle, Criminologie 38 (2) (2005) 39–64.
[5] O. Delémont, E. Lock, O. Ribaux, Forensic science and criminal investigation. Encyclopedia of Criminology and Criminal Justice, Springer, 2014, pp. 1754–1763.
[6] A. Baylon, L'utilisation du renseignement forensique pour guider les décisions liées à l'investigation de scène de crime, Rev. Suisse Criminol. 11 (40–45) (2012).
[7] O. Ribaux, A. Baylon, E. Lock, O. Delémont, C. Roux, C. Zingg, P. Margot, Intelligence-led crime scene processing. Part II: Intelligence and crime scene examination, Forensic Sci. Int. 199 (1–3) (2010) 63–71.
[8] O. Delémont, P. Esseiva, O. Ribaux, P. Margot, La violence laisse des traces: l'homicide dévoilé par la science forensique, in: M. Cusson, S. Guay, J. Proulx, F. Cortoni (Eds.), Traité des Violences Criminelles, Hurtubise, Montréal, 2012.
[9] O. Ribaux, Police Scientifique Le renseignement par la trace, Presses polytechniques et universitaires romandes 1 (2014).
[10] J.H. Ratcliffe, Intelligence-led policing and the problems of turning rhetoric into practice, Polic. Soc. 12 (1) (2002) 53–66.
[11] ANZPAA, End-to-end forensic identification process project. 2012.
[12] C. Brown, A. Ross, R.G. Attewell, Benchmarking Forensic Performance in Australia–Volume Crime, Forensic Sci. Policy Manag. Int. J. 5 (3-4) (Jan 2015) 91–98.
[13] Home Office, Summary Report of the Scientific Work Improvement (SWIM) Package. Home Office, 2007.
[14] A.A. Mapes, A.D. Kloosterman, C.J. de Poot, DNA in the Criminal Justice System: The DNA Success Story in Perspective, J. Forensic Sci. (2015).
[15] S. Bitzer, N. Albertini, E. Lock, O. Ribaux, O. Delémont, Utility of the clue —From assessing the investigative contribution of forensic science to supporting the decision to use traces, Sci. Justice 55 (6) (2015) 509–513.
[16] D. Baskin, I. Sommers, The influence of forensic evidence on the case outcomes of homicide incidents, J. Crim. Justice 38 (6) (2010) 1141–1149.
[17] F. Horvath, R. Meesig, The Criminal Investigation Process and the Role of Forensic Evidence: A Review of Empirical Findings, J. Forensic Sci. 41 (6) (1996) 963–969.
[18] M. Ramsay, The Effectiveness of the Forensic Science Service, Her Majesty's Statuary Office, London, 1987.
[19] N. Ritter, Untested evidence in sexual assault cases: using research to guide policy and practice, Sex. Assault Rep. 16 (3) (2013).
[20] K.J. Strom, M.J. Hickman, Unanalyzed evidence in law-enforcement agencies, Criminol. Public Policy 9 (2) (2010) 381–404.
[21] D. Baskin, I. Sommers, The influence of forensic evidence on the case outcomes of assault and robbery incidents, Crim. Justice Policy Rev. (Mar 2011) 1–25.
[22] J.L. Peterson, I. Sommers, D. Baskin, D. Johnson, The role and impact of forensic evidence in the criminal justice process, Natl. Inst. Justice (2010).
[23] M. Kuhn, K. Johnson, Applied Predictive Modeling, Springer, 2013.