



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

---

Year : 2019

## Evaluation des résultats ADN considérant des propositions au niveau de l'activité

Samie Foucart Lydie

Samie Foucart Lydie, 2019, Evaluation des résultats ADN considérant des propositions au niveau de l'activité

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB\_F5BBEDAE658D2

### **Droits d'auteur**

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

### **Copyright**

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

Université de Lausanne  
Faculté de droit, des Sciences criminelles et d'Administration publique  
Ecole des Sciences Criminelles

---

# **Evaluation des résultats ADN considérant des propositions au niveau de l'activité**

---

## **Thèse de doctorat**

Présentée à la Faculté de droit, des Sciences criminelles et d'Administration publique de  
l'Université de Lausanne

Pour l'obtention du grade de Docteur ès Sciences en Science forensique

Par

Lydie Samie-Foucart

Directeurs de thèse :  
Prof. Christophe Champod  
Prof. Franco Taroni

LAUSANNE

2019



*Unil*

UNIL | Université de Lausanne  
Ecole des sciences criminelles  
bâtiment Batochime  
CH-1015 Lausanne

## IMPRIMATUR

A l'issue de la soutenance de thèse, le Jury autorise l'impression de la thèse de Mme Lydie Samie-Foucart, candidate au doctorat en science forensique, intitulée

« Évaluation des résultats ADN considérant des propositions  
au niveau de l'activité »

Le Président du Jury



Professeur Andy Bécue

Lausanne, le 17 décembre 2019



## Remerciements

---

Je tiens à remercier Christophe Champod et Franco Taroni, Professeurs à l'École des Sciences Criminelles de l'Université de Lausanne, qui m'ont encadrée, encouragée et soutenue tout au long de cette thèse. A leurs côtés, j'ai pu m'épanouir durant ces cinq années de doctorat. Je remercie plus particulièrement Christophe Champod pour son aide et ses conseils dans l'écriture d'articles et Franco Taroni pour son enthousiasme et sa participation aux expériences menées dans le cadre de ce projet de thèse.

Je remercie également Andy Bécue, Professeur à l'École des Sciences Criminelles de l'Université de Lausanne et Président de ce jury, d'avoir accepté ce rôle ainsi que de son soutien et ses conseils lors de la préparation du concours MT180, qu'il m'a fait découvrir.

J'adresse tous mes remerciements à Diane SEGUIN, directrice de l'unité de Biologie/ADN du laboratoire de sciences judiciaires et de médecine légale de Montréal, à la Docteure Diana Hall, Professeure à l'École des Sciences Criminelles ainsi qu'au Docteur Patrick Basset, de l'honneur qu'ils m'ont fait en acceptant de faire partie du jury de cette thèse.

Je remercie également le bureau de l'égalité de l'UNIL ainsi que la fondation pour l'UNIL et l'association des diplômés en Science Criminelles de m'avoir soutenue financièrement pour que je sois en mesure de présenter les résultats de cette thèse en conférence.

Je tiens à remercier du fond du cœur Marco De Donno, sans qui cette thèse aurait pris beaucoup plus de temps ainsi que Duncan Taylor pour son implication et son intérêt pour ce projet.

J'exprime ma gratitude à mes parents, Michèle Foucart et René Samie, qui ont lu et corrigé cette thèse, tant dans la structure que dans le contenu. Cela constitue un travail impressionnant et je les remercie pour cela.

Je tiens à remercier Tacha Hicks Champod qui m'a soutenue tout au long de cette thèse. Je la remercie pour ses enrichissantes discussions, sa participation aux expériences, son enthousiasme et son intérêt immense pour ce domaine.

Egalement, un grand merci à Miguel Garcia, responsable de la plateforme technologique de cytométrie en flux à l'École polytechnique fédérale de Lausanne (EPFL), à Valérie Glutz, de cette même unité, ainsi qu'au Docteur Vincent Castella, responsable de l'Unité de Génétique Forensique de Lausanne, pour leur contribution dans la publication d'articles, leur aide et les discussions qui m'ont permis de faire avancer ce travail.

Je remercie Mélissa Bryois, Géraldine Gobat, Florian Honoré, Nicolas Schmitt, et tout particulièrement Valentin Carlier, Margaux Dupuy, Anne Céline Fléchet et Ambre Jeanneret-Grosjean, pour avoir donné de leur temps afin de participer aux expériences.

---

## Table des matières

---

<b>Remerciements .....</b>	<b>I</b>
<b>Table des matières .....</b>	<b>III</b>
<b>Résumé .....</b>	<b>VII</b>
<b>Abstract .....</b>	<b>VIII</b>
<b>Liste des abréviations.....</b>	<b>IX</b>
<b>Chapitre I - Introduction.....</b>	<b>1</b>
I.1 Contexte .....	1
I.2 Problématique .....	2
I.3 Objectifs de la thèse .....	9
I.4 Originalité de la thèse .....	11
I.5 Choix du cas.....	12
I.6 Contributions associées à ce travail .....	15
<b>Chapitre II - Elaboration d'un modèle probabiliste.....</b>	<b>17</b>
II.1 Sélection des variables.....	18
II.1.1 Le transfert.....	19
II.1.2 La persistance .....	24
II.1.3 Le bruit de fond .....	25
II.1.4 La collecte et l'analyse de la trace en laboratoire.....	26
II.1.5 La pollution.....	27
II.1.6 Les erreurs de laboratoires.....	29
II.1.7 Discussion.....	30
II.2 Relation entre variables et insertion des données .....	31
II.2.1 Les propositions.....	31
II.2.2 Les observations .....	32
II.2.3 Le raisonnement de construction .....	35
II.2.4 États des variables.....	42
II.2.5 Données .....	43
II.3 Discussion.....	44
II.3.1 Les limites du réseau bayésien .....	44
II.3.2 Les avantages du réseau bayésien.....	45



---

<b>Chapitre III - Identification des variables d'intérêt .....</b>	<b>47</b>
III.1 Introduction .....	47
III.2 Identification des potentielles variables d'intérêt.....	50
III.2.1 Résultat général .....	53
III.2.2 Exemple de M. X.....	56
III.2.3 Effet des quantités d'ADN de POI et de Non POI sur le rapport de vraisemblance .....	58
III.3 Identification des variables nécessitant l'acquisition de données supplémentaires.....	60
III.3.1 Résultats généraux.....	61
III.3.2 Cas de M. X.....	63
III.4 Discussion .....	64
III.4.1 Prise en compte de la variation relative entre POI et Non POI .....	64
III.4.2 Généralisation des résultats .....	65
III.4.3 Effet des propositions sur le rapport de vraisemblance .....	66
III.4.4 Nombre d'expériences.....	67
<b>Chapitre IV- Etude du type de données à utiliser .....</b>	<b>69</b>
IV.1 Acquisition de données spécifiques .....	69
IV.1.1 Efficacité d'extraction et efficacité de prélèvement.....	69
IV.1.2 Quantité d'ADN sur les mains, Proportion de transfert et Bruit de fond .....	72
IV.2 Impacts du type de données utilisé sur le rapport de vraisemblance.....	78
IV.2.1 Introduction.....	78
IV.2.2 Méthode .....	78
IV.2.3 Résultats .....	80
IV.2.4 Discussion .....	82
<b>Chapitre V- Etude de l'efficacité de la méthodologie .....</b>	<b>85</b>
V.1 Introduction .....	85
V.2 Méthode.....	86
V.2.1 Obtention des rapports de vraisemblance dits « initiaux » .....	86
V.2.2 Obtention des rapports de vraisemblance dits « calibrés ».....	87
V.2.3 Obtention des rapports de vraisemblance dits « reportés » .....	88
V.3 Résultat.....	89
V.3.1 Rapports de vraisemblance initiaux.....	89
V.3.2 Rapports de vraisemblance calibrés .....	91
V.3.3 Rapports de vraisemblance reportés .....	93
V.4 Discussion .....	102
V.4.1 Valeur du rapport de vraisemblance .....	102
V.4.2 Statut du donneur de l'AO.....	105

---

---

<b>Chapitre VI - Vers une application pratique de la méthodologie.....</b>	<b>111</b>
VI.1 Les étapes de la méthodologie en pratique.....	111
VI.2 Généralisation de la méthodologie.....	113
VI.3 Plus-value de la démarche proposée .....	122
VI.3.1 Comparaison avec les résultats obtenus par Pun (2016).....	122
VI.3.2 Pré-évaluation .....	124
VI.4 Limitations pratiques de la méthode .....	126
VI.4.1 Besoin d'hypothèses bien définies .....	127
VI.4.2 Construction d'un réseau bayésien et effectuer son analyse de sensibilité .....	128
VI.4.3 Calibration du système.....	129
VI.4.4 Gestion d'impondérables : Exemple du port des gants .....	129
VI.4.5 Expériences : POI, type de surface et type de contact.....	130
VI.4.6 L'application à un nombre restreint de cas .....	131
VI.5 Perspectives.....	132
VI.5.1 Problématique de l'AO .....	132
VI.5.2 Besoin de formation .....	132
VI.5.3 Partage d'information.....	133
VI.5.4 Sensibilité des analyses ADN .....	134
VI.5.5 Investiguer le scénario où un contact légitime est allégué par la défense.....	135
VI.5.6 Intégrer les erreurs de laboratoire.....	135
<b>VII - Conclusion .....</b>	<b>137</b>
<b>Bibliographie.....</b>	<b>143</b>
<b>Annexes .....</b>	<b>- 1 -</b>
Annexe I: Articles publiés .....	- 1 -
I.1 Taylor et al., Helping to distinguish primary from secondary transfer events for trace DNA, <i>Forensic Science International: Genetics</i> . 28 (2017) 155–177. ....	- 1 -
I.2 Samie et al., The efficiency of DNA extraction kit and the efficiency of recovery techniques to release DNA using flow cytometry, <i>Science &amp; Justice</i> . 59 (2019) 405-410.....	- 52 -
Annexe II: Articles soumis à publication .....	- 71 -
II.1 Samie et al., The use of Bayesian Networks and simulation methods to identify the variables impacting the LR in stabbing cases evaluated given activity level propositions (2019) (Soumis pour publication) .....	- 71 -
II.2 Samie et al., Estimating the quantity of transferred DNA in primary and secondary transfers, <i>Science &amp; Justice</i> (2019) (Soumis pour publication, accepté, en cours de production) .....	- 102 -
Annexe III : Revue de la littérature centrée sur les traces ADN de contact .....	- 120 -

---

---

Annexe IV: Données issues de l'étude sur l'efficacité d'extraction et de prélèvement .....	- 140 -
IV.1 Efficacité d'extraction.....	- 140 -
IV.2 Efficacité de prélèvement et d'extraction.....	- 143 -
Annexe V: Données issues de l'étude sur la proportion de transfert : Transfert direct et prélèvement sur les mains .....	- 144 -
V.1 Participant 1.....	- 145 -
V.2 Participant 2.....	- 147 -
V.3 Participant 3.....	- 149 -
V.4 Participant 4.....	- 151 -
V.5 Participant 5.....	- 153 -
V.6 Participant 6.....	- 155 -
Annexe VI: Données utilisées pour informer les variables du BN.....	- 157 -
VI.1 Variable « Efficacité d'extraction » .....	- 157 -
VI.2 Variable « Efficacité de prélèvement » .....	- 158 -
VI.3 « Quantité d'ADN sur les mains .....	- 160 -
VI.4 « Proportion de transfert » : Propositions du procureur est vraie.....	- 162 -
VI.5 « Proportion de transfert » : Propositions de la défense Hd <sub>1</sub> est vraie .....	- 164 -
VI.6 « Proportion de transfert » : Propositions de la défense Hd <sub>3</sub> est vraie .....	- 166 -
VI.7 « Bruit de fond » : Scénario 1 .....	- 168 -
VI.8 « Bruit de fond » : Scénario 3 .....	- 171 -
Annexe VII: Résultats obtenus pour les cas reproduits.....	- 174 -
VII.1. Scénario 1 : Transfert secondaire allégué par la défense .....	- 174 -
VII.2. Scénario 3 : Utilisation légitime du couteau allégué par la défense.....	- 177 -

## Résumé

---

L'évaluation de très faibles quantités d'ADN est devenue de plus en plus complexe : les variables liées au transfert, à la persistance et au prélèvement deviennent de plus en plus importantes dans l'évaluation des résultats ADN, car la question pertinente qui se pose à la Cour n'est plus « qui est la source de cet ADN ? » mais plutôt « comment cet ADN est-il arrivé là ? ». Le guide de l'ENFSI (Willis et al, 2015) pour les rapports d'évaluation propose d'évaluer les résultats ADN en considérant des propositions de niveaux de l'activité et en utilisant l'approche de rapport de vraisemblance (RV), surtout si une très faible quantité d'ADN est récupérée.

Cependant, dans la pratique, les experts évaluent leurs observations principalement au niveau de la (sous-)source et rarement au niveau de l'activité car (1) les évaluations prenant en compte les propositions au niveau de l'activités sont rarement demandées. De plus, (2) il n'existe pas d'outil adapté au laboratoire pour faire l'évaluation en considérant les propositions au niveau de l'activité. L'objectif de cette recherche est de proposer une méthodologie pour réaliser ce type d'évaluation.

La méthodologie proposée comporte plusieurs étapes : (1) construire un réseau bayésien (BN), (2) identifier les variables ayant un impact sur le RV grâce à une analyse de sensibilité du BN effectuée par des simulations de ré échantillonnage, (3) définir, sur la base des résultats des simulations, le nombre de données supplémentaires à acquérir si nécessaire, (4) acquérir ces données supplémentaires et (5) déterminer si le RV est robuste et peut donc être reporté dans un rapport d'évaluation.

La méthodologie a été appliquée avec succès à 120 cas simulés d'attaques au couteau développés en laboratoire pour chacun des deux scénarios suivants: l'un impliquant un transfert secondaire potentiel, l'autre impliquant une utilisation récente du couteau avant l'attaque sous l'hypothèse de la défense. Cette méthodologie peut également être appliquée à d'autres cas. Elle apporte une solution opérationnelle. Toutefois, il existe des limites, qui résultent souvent des impondérables des circonstances de l'affaire.

## Abstract

---

Evaluation of very low amounts of DNA has become increasingly complex: variables related to transfer, persistence and recovery are becoming increasingly important in evidence evaluation, as the relevant question arising in Court is no longer “who is the source of this DNA?” but rather “how did this DNA get there?”. The ENFSI guideline (Willis et al., 2015) for evaluative reporting proposes to report trace DNA evidence considering activity level propositions adopting a likelihood ratio (LR) approach, especially if a very low amount of DNA is recovered.

In practice, however, experts evaluate their observations mainly at the (sub)-source level and rarely at the activity level because (1) assessments considering propositions at the activity level are rarely requested. Moreover, (2) there is no tool adapted to the laboratory to make the evaluation considering propositions at the activity level. The objective of this research is to propose a methodology to carry out such evaluation.

The proposed methodology consists of several steps: (1) to build a Bayesian Network (BN), (2) to identify variables having an impact on the LR based on a sensitivity analysis of the BN carried out by re-sampling simulations, (3) to define, based on the results of the simulations, the number of additional data to be acquired if necessary, (4) to acquire this additional data and (5) to determine if the LR is robust and can therefore be reported further.

The methodology was successfully applied to 120 simulated cases of knife attacks developed in the laboratory for each of two scenarios: one involving a potential secondary transfer, the other involving a recent use of the knife before the attack under the hypothesis of defence. This methodology can also be applied to other cases. It provides an operational solution. However, there are limitations, often resulting from the imponderables of the case circumstances.

## Liste des abréviations

---

ADN	Acide Désoxyribo-Nucléique
AO	Agresseur alternatif (Alternative Offender)
BN	Réseau bayésien (Bayesian Network)
CPT	Tables de probabilités conditionnelles (Conditionnal Probabilities Table)
KBS	Système de connaissance (Knowledge-Based System)
LR	Rapport de vraisemblance (Likelihood Ratio)
OBN	Object Oriented Bayesian Network
POI	Personne d'intérêt (Person of Interest)



## Chapitre I - Introduction

---

### I.1 Contexte

Affaire Anderson (Osagie, 2013), United States, San Jose. Suite à une agression, M. Kumra est retrouvé mort à son domicile. Un profil ADN correspondant à celui de M. Anderson est prélevé sous les ongles de la victime. Dans un premier temps, ce dernier est suspecté du meurtre. Mais observer un profil ADN correspondant au profil ADN d'un individu en fait-il l'agresseur ? Dans ce cas, les enquêteurs constatent que, si le profil ADN de M. Anderson correspond à celui de la trace, celui-ci ne peut être l'agresseur puisqu'il était hospitalisé au moment des faits (Hesman Saey, 2019).

Les procureurs se sont d'abord focalisés sur l'origine de la trace. Ce n'est qu'après avoir constaté qu'il était impossible que M. Anderson ait agressé la victime qu'ils se sont intéressés à la manière dont cet ADN avait pu se retrouver sous les ongles de la victime. La question pertinente est alors :

« Comment la trace ADN dont le profil ADN correspond au profil ADN de M. Anderson a-t-elle pu arriver là ? » et non plus « Qui est à l'origine de la trace d'ADN ? ». Cette distinction est importante car il existe un vrai risque de confusion dans la communication des résultats : le fait que les observations génétiques de la trace (i.e les caractéristiques propres de la trace ne sont pas différenciables de celles du profil du suspect) soutiennent que la trace ADN provient du suspect plutôt que d'une autre personne, n'implique pas de facto qu'elles soutiennent que le suspect ait agressé la victime.

En première intention, la cour mandate les experts afin de l'aider à se déterminer uniquement sur la source de la trace ADN. De plus, les connaissances et les données à disposition permettant d'aider la cour à discriminer, sur la base de données génétiques observées, entre les activités ayant pu entraîner le dépôt de la trace ADN sont limitées. Les experts se focalisent alors uniquement sur la source de la trace ADN.

Il y a dans cette pratique un risque de voir le système judiciaire donner plus de poids à un rapprochement que légitime. M. Anderson, malgré son alibi, a passé 5 mois en prison. Si M. Anderson n'avait pas eu un « alibi » aussi fort, le procureur aurait sans doute poursuivi son action, persuadé que l'ADN sous les ongles de la victime ne pouvait pas s'y trouver par hasard.



L'objectif de cette thèse est de proposer un outil et une méthodologie permettant aux experts d'évaluer les résultats ADN lorsque des activités sont alléguées par les différentes parties ou lorsque les circonstances du cas l'exigent.

## I.2 Problématique

Les résultats d'une expertise sont appréciés au regard d'au moins deux propositions alternatives, mutuellement exclusives (Evet, 1998), proposées par les différentes parties et qui dépendent des circonstances du cas (Buckleton et al., 2014).

Il existe une hiérarchie dans la formulation de ces propositions. Celles-ci peuvent être posées à trois niveaux possibles : la (sous-) source, l'activité et le délit (Cook et al., 1998 ; Evett et al., 2002 ; Jackson et al., 2006).

- Le niveau de la sous-source s'exprime sur l'origine de l'ADN détecté (la nature du fluide biologique de l'ADN étant inconnue). Les propositions au niveau de la source quant à elles, se positionnent sur l'origine de la trace dont la nature est caractérisée comme par exemple du sang/salive/sperme/poils.
- Le niveau de l'activité se concentre sur la ou les actions à l'origine de la trace afin d'aider à répondre à la question : « Comment la trace ADN est-elle arrivée là ? ».
- Enfin le niveau du délit se focalise sur le criminel potentiel, donc sur le fait que l'individu ait pu commettre le crime ou le délit.

Souvent, les experts analysent et évaluent leurs observations principalement au niveau de la (sous-) source et très peu au niveau de l'activité. Pourquoi ?

Différentes explications sont proposées par Biedermann et al. (2016), la principale étant le choix et l'utilisation des données. Au niveau (sous-) source, l'évaluateur doit notamment obtenir des données sur la rareté du profil génétique de la trace dans la population pertinente, sur les probabilités qu'un allèle ne soit pas observable ou qu'il soit observé alors qu'il ne devrait pas l'être ou qu'il s'agisse d'un artefact. Ces données sont souvent disponibles et publiées.

Les évaluations considérant des propositions au niveau de l'activité nécessitent quant à elles de nombreuses données sur les processus de transfert de persistance et de bruit de fond qui sont pour la plupart spécifiques au cas considéré.

Des données existent (cf. Annexe III) mais elles ne sont pas toutes publiées. Aussi, les experts hésitent quant au type de données à exploiter :

- Pour beaucoup, utiliser des données non publiées, telles que les rapports internes et celles basées sur leurs propres expériences et connaissances comportent, d'une part, le risque d'une moindre objectivité, et, d'autre part, le risque que la validité de ces éléments soit remise en cause par le tribunal. Beaucoup considèrent que seules les données publiées sont exploitables pour évaluer les résultats ADN considérant des propositions au niveau de l'activité (Taroni, 2013). Celles-ci sont considérées comme étant robustes car elles sont le résultat d'expériences scientifiques qui ont été évaluées et reconnues scientifiquement par les pairs.
- Pour d'autres experts (Goray, 2012), les données doivent être spécifiques à chaque cas considéré du fait du nombre important de facteurs entrant en jeu dans les processus de transfert et de persistance. Chaque cas nécessite donc de faire un « design expérimental » précis tenant compte de tous les facteurs connus, travail qui consiste à acquérir de nombreuses données et de pouvoir les analyser. A ce jour, peu de résultats portant sur le transfert et la persistance sur des cas spécifiques sont publiés.

Ceci explique que les experts considèrent souvent les résultats au regard uniquement de leur source potentielle sans les mettre dans le contexte des activités alléguées. Certains estiment que cette contextualisation devrait être laissée aux acteurs du système judiciaire (enquêteurs, magistrats, tribunaux). Or, il paraît risqué de laisser l'appréciation des résultats dans un contexte à des acteurs non spécialistes de l'évaluation de la trace, qui se contenteraient des observations faites et méconnaîtraient les processus mis en œuvre dans le dépôt de la trace. En effet, la complexité des phénomènes pouvant amener aux transferts d'ADN nécessite le regard d'un expert, seul apte à discuter des résultats de l'évaluation des traces tant au niveau de la source que de l'activité, de la solidité et des limites de ceux-ci et des interprétations adéquates.

Toutefois, de nombreux scientifiques font état du manque de données et de méthodologie permettant d'aborder ces questions avec le niveau de rigueur requis. Un des objectifs de cette thèse est de chercher à combler ces lacunes.

De plus, aujourd'hui, les experts ne peuvent plus limiter leur domaine d'intervention à la seule évaluation des résultats ADN au niveau de la (sous-) source. Du fait de l'amélioration

---

des techniques, de plus en plus de traces dites « ADN de contact » sont prélevées. Ces traces dites « ADN de contact » sont des traces qui ne peuvent pas être associées à des substances biologiques telles que le sang, le sperme, la salive, les poils (Meakin et Jamieson, 2013). Les traces de contact s'échangent dans des conditions multiples, impliquant autant des actions d'intérêt pour l'enquête que des actions légitimes de la vie de tous les jours. Elles représentaient en 2018 environ 80% de l'ensemble de traces analysées au Centre Universitaire Romand de Médecine Légale (CURML) de Lausanne (3645 traces de contact sur 4624 traces analysées d'après leur base de données). Ces traces, prélevées en quantité parfois infime, impliquent souvent des modes de transferts complexes tels que des transferts via plusieurs surfaces intermédiaires (transfert d'ADN d'un individu sur un objet puis de cet objet sur une autre surface sur laquelle l'ADN sera prélevé). Pour ces traces, il est essentiel que ces transferts soient pris en compte lors de l'évaluation, même si aucune hypothèse d'activité alternative n'est avancée par les parties. Si l'expert se prononce uniquement au niveau de la (sous-) source en ignorant le niveau de l'activité, l'information qui sera apportée ne sera pas toujours utile ou suffisante (Taroni, 2013) et il existera un risque d'incompréhension sur la réelle portée des résultats génétiques observés (Champod, 2013).

Il est nécessaire de s'interroger sur la manière dont la trace a été déposée sur la scène de crime ou d'infraction et à quel moment. Dans l'affaire Anderson évoquée précédemment, en l'absence d'alibi, peut être que seule la question de la source de la trace d'ADN aurait été étudiée et pas du tout celle de l'origine de l'action. La problématique liée aux processus de transfert n'aurait pas été soulevée. Le risque est alors de laisser condamner la personne dont le profil ADN correspond à celui de la trace alors qu'aucun mécanisme ayant mené au transfert d'ADN n'a été considéré, soit par l'expert soit par le tribunal. Il aurait été essentiel de pouvoir quantifier la force de l'observation (le profil ADN de la trace correspond au profil ADN de M. Anderson) sous au moins deux propositions au niveau de l'activité.

Dans son article, Cale (2015) met justement en avant la problématique spécifique liée à l'existence de transferts complexes et leurs risques en matière d'évaluation. Plusieurs affaires illustrant cette nécessité d'évaluer ces traces en considérant des propositions au niveau de l'activité sont décrites par Taylor et al. (2018).

Intuitivement, il est attendu que, lorsque qu'une personne saisit un objet à main nue, elle transfère de fait son ADN sur cet objet. Un profil peut ainsi être obtenu. Aussi, si le profil d'un suspect ne correspond pas au profil de la trace, ceci entraîne en général, l'exclusion du

---

suspect en tant que source potentielle de la trace mais également en tant que personne pouvant avoir manipulé l'objet.

Or, dans certains cas, lors de la manipulation d'objets, il est possible que l'ADN ne puisse pas être détecté (Ryan, 2012). Il est alors impossible d'observer de correspondance entre le profil de la trace et le profil de la personne ayant fait l'action d'intérêt. Evaluer l'absence de correspondance ou l'absence de détection d'ADN est possible seulement si les propositions sont formulées au niveau de l'activité et pas seulement au niveau de la (sous-) source (Taroni et al., 2019).

McKenna (2013) a étudié une telle situation où le profil de la personne d'intérêt ne correspond pas au profil retrouvé sur une arme à feu. En se basant sur les travaux de Polley et al. (2006), elle conclut que, dans le cas d'une absence de correspondance entre les profils, les observations soutiennent faiblement l'hypothèse qu'une autre personne a tiré avec l'arme en question plutôt que la personne d'intérêt. Partant de ces constatations, elle indique que la police ne devrait pas écarter un suspect possible sur la seule base d'une différence entre le profil ADN détecté sur l'objet d'intérêt et celui de la personne en cause.

A l'inverse, Victor Nealon, aujourd'hui libéré, a passé 17 ans en prison faussement accusé d'avoir violé une femme alors que le profil ADN retrouvé sur les habits de la victime ne correspondait pas à son profil. A cette époque, le Procureur avait estimé que le profil inconnu pouvait être du bruit de fond ou de la contamination et que Victor Nealon n'avait pas transféré son ADN au cours de l'agression (Tophan, The Guardian, 2013).

Ces affaires mettent en évidence la nécessité de considérer la possibilité que l'ADN de la personne qui a effectivement fait l'activité d'intérêt ne soit pas détecté. De fait, l'absence de profils ADN doit aussi être évalué au regard de propositions au niveau de l'activité. Cette évaluation n'est pas possible si seules des propositions au niveau de la source sont considérées.

Afin d'éviter le risque d'incompréhension sur la réelle portée des résultats génétiques observés, le guide de « l'European Network of Forensic Science Institutes » (ENFSI) sur la rédaction des rapports évaluatifs (Willis et al., 2015) préconise d'évaluer les résultats ADN considérant des propositions au niveau de l'activité notamment quand des facteurs, tels que le transfert, la persistance ou le bruit de fond peuvent avoir un impact sur l'interprétation des résultats relatifs aux activités alléguées. Le guide précise également que toutes les données,

---

publiées ou non, peuvent être utilisées à condition que la démarche soit transparente. Le guide préconise cependant d'avoir une hiérarchie des informations et des données à utiliser. Ainsi, il reste préférable d'exploiter, en premier lieu, l'ensemble des données publiées appropriées au cas. Si celles-ci ne sont pas disponibles, l'expert doit s'appuyer sur les données issues des rapports internes puis, en dernier lieu, sur celles issues de ses propres expériences et connaissances qu'il est en mesure de justifier.

Champod (2013) donne son point de vue sur la question et estime que la non-publication des données n'empêche pas une évaluation des résultats. L'expert peut se prononcer sur les probabilités d'un résultat, connaissant les divers mécanismes de transfert et de persistance, et en tenant compte des incertitudes, à partir du moment où son travail reste transparent sur les informations et la méthodologie utilisées.

Le « Body Fluid Forum » (BFF), d'après David et al. (2016), a indiqué que l'analyse critique de Meakin et Jamieson (2013) avait été utilisée au tribunal au Royaume-Uni et en Irlande pour soutenir que l'évaluation considérant des propositions posées au niveau de l'activité ne doit pas être faite car elle est discutable et peu sûre, du fait de la multitude de facteurs à prendre en compte. Suite à cet événement, David et al. (2016), ont fait remarquer dans leur commentaire qu'il était préférable de réaliser une évaluation avec des données mêmes limitées que pas d'évaluation du tout. Pour ce groupe, il est du devoir des experts d'évaluer les résultats. Pour ces auteurs, il est risqué de laisser cette appréciation à des acteurs qui ne sont pas des spécialistes de l'évaluation de la trace en question en raison de la complexité des phénomènes pouvant amener aux transferts d'ADN. Cette thèse soutient également cette position.

Dans l'affaire Weller en 2006 à Londres (Weller, R. v, 2010), les tribunaux ont admis qu'un expert puisse émettre une opinion en l'absence de données quantitatives ou statistiques à condition que les avis et leurs fondements soient communiqués en toute transparence. Dans cette affaire, M. Weller est accusé d'avoir agressé sexuellement sa victime en pratiquant une pénétration digitale vaginale. Des prélèvements ADN ont été réalisés sous ses ongles. Un profil ADN de mélange, dont la partie féminine est non différenciable du profil de la victime, est retrouvé sous deux ongles. M. Weller n'a pas contesté le fait que l'ADN retrouvé puisse être celui de la victime mais, selon lui, celui-ci a pu se déposer alors qu'il aidait cette dernière, fortement alcoolisée, à s'installer sur le lit en position de sécurité ou pendant qu'il ramassait ses vêtements abandonnés sur le sol. La question d'intérêt portait donc sur l'action

à l'origine du dépôt (soit une pénétration digitale vaginale, soit une aide apportée à la victime). Les experts ont cherché à évaluer les résultats ADN dans le cadre de ces activités alléguées. M. Weller a été reconnu coupable par le jury mais a fait appel en 2010, contestant la validité des données utilisées par les experts. Selon la défense, les experts n'avaient pas suffisamment de données publiées pour effectuer une évaluation considérant les activités alléguées. L'appel a été rejeté au motif que les données utilisées pour l'évaluation des résultats pouvaient reposer sur l'expérience des experts et les résultats de travaux non publiés. Depuis, des travaux ont été publiés en se basant sur le cas Weller ainsi que sur la problématique plus générale du transfert et de la persistance d'ADN sous les ongles (Flanagan et McAlister, 2011 ; Malsom et al., 2009 ; Cook et Dixon, 2007).

Des auteurs se sont également attachés à obtenir des données pour d'autres cas particuliers et à proposer des méthodes pour les utiliser dans une évaluation. Jones et al. (2016) se sont intéressés aux cas d'agressions sexuelles lorsque l'ADN de la victime est observé sur des prélèvements effectués sur le pénis ou le sous-vêtement du suspect. Fonnelop et al. (2017), eux, ont étudié le cas d'une agression où les prélèvements d'ADN sont effectués sur le tee-shirt de la victime.

Cependant, une étude de Goray (2012) montre que les taux d'ADN transféré, basés sur les quantités prélevées, qui avaient été obtenus par des expériences réalisées dans des conditions contrôlées lors d'une étude précédente (Goray, 2010) n'étaient plus valides lorsqu'ils étaient utilisés pour des cas réels. Selon les auteurs, plusieurs raisons peuvent expliquer les différences constatées entre les résultats observés lors d'expériences contrôlées et les résultats obtenus pour des cas réels :

- Lorsque l'expert est amené à évaluer un cas réel, certains paramètres ne sont pas connus, comme par exemple, le type de surface intermédiaire si la possibilité d'un transfert multiple est discutée ou encore la pression (type de contact, durée et intensité du contact) exercée sur les objets. Ces paramètres sont susceptibles d'être différents de ceux testés lors d'expériences contrôlées.
- Lors d'expériences contrôlées, le matériel utilisé est changé ou nettoyé après chaque expérience. Or dans les cas à évaluer, les objets sur lesquels les prélèvements sont effectués ont été manipulés et les utilisations successives peuvent y avoir ajouté un bruit de fond.

- Les types de surface de contact lors des études sont souvent différents de ceux que l'on peut rencontrer dans les cas réels.

Selon les auteurs donc, pour chaque cas étudié, les facteurs tels que les conditions environnementales, le type de substrat, la durée et le type de contact devraient être connus pour pouvoir réaliser des expériences spécifiques au cas. Cependant, l'influence de ces variables sur le résultat des expériences n'a pas été testée par les auteurs. D'après Fonnøløp et al. (2015), l'étude de l'occurrence des phénomènes de transfert indirect dans des conditions non contrôlées qui se rapprochent le plus des conditions du cas étudié est également nécessaire.

Biedermann et al. (2016), expliquent pourquoi les experts peuvent éprouver des difficultés avec ce type d'évaluation. En effet, les auteurs indiquent : "because each case has its own feature, the use of numerical values from experimental studies performed under controlled (laboratory) conditions cannot be used for evaluation in real life case", ce qui veut dire qu'il est difficile de concevoir un design expérimental qui permettrait de prendre en compte tous les facteurs pouvant intervenir dans un cas réel.

Pour résumer, le niveau de la source est souvent le seul niveau de proposition utilisé lors des évaluations des résultats ADN lors des expertises alors que de nombreux cas exigeraient que l'on s'intéresse également au niveau de l'activité. Cette situation s'explique par deux raisons principales. Tout d'abord, les experts sont généralement mandatés afin d'évaluer les résultats ADN considérant des propositions au niveau de la source. De plus, ils ne disposent pas actuellement de méthodologie reconnue et/ou n'ont pas connaissance des données pouvant être utilisées. Il en découle de leur part une attitude somme toute prudente qui consiste à considérer les résultats en fonction de leur source potentielle tout en laissant l'appréciation de la preuve au regard des activités alléguées aux acteurs judiciaires. Cette thèse se propose d'essayer de combler ce manque de méthodologie.

Ce travail fait suite à la thèse de Pun (2016) qui a étudié un scénario de vol de voiture pour lequel plusieurs propositions de la défense au niveau de l'activité sont posées. La pertinence des prélèvements ADN ainsi que l'impact des propositions sur le LR a été étudié. Ici, l'objectif n'est pas d'étudier un cas particulier et de montrer comment évaluer les résultats ADN en utilisant les résultats de l'étude mais de proposer une méthodologie générale pour l'évaluation des résultats ADN considérant des propositions au niveau de l'activité à partir

de l'étude d'un cas particulier. La méthodologie se base sur les résultats obtenus dans la thèse de Pun (2016) pour le cas du scénario de vol de voiture. L'originalité et l'apport de ce travail par rapport à la thèse de Pun (2016) sera décrit plus en détail par la suite.

### I.3 Objectifs de la thèse

**L'objectif est d'élaborer une méthodologie ainsi qu'un outil qui permettent aux experts d'évaluer des résultats ADN lorsque différentes activités d'intérêt sont alléguées.**

Cette méthodologie se veut applicable à de nombreux cas. Elle vise à proposer une solution opérationnelle et efficace. Elle repose sur l'étude de l'influence des différentes variables intervenant dans les processus de transfert, de persistance et de collecte sur le résultat de l'évaluation afin de cibler les variables d'intérêt. Les variables d'intérêt sont celles dont l'impact sur le résultat de l'évaluation, mesuré par une métrique probabiliste appelée rapport de vraisemblance (« Likelihood Ratio » ou LR), est tel que des données doivent être acquises. Cette recherche est structurée en 6 chapitres (comprenant cette introduction) :

1<sup>er</sup> chapitre : Introduction

Ce chapitre pose la problématique de ce travail dans un contexte précis afin d'en justifier la pertinence et formule les objectifs de la recherche.

2<sup>ème</sup> chapitre : Elaboration d'un modèle probabiliste

Un modèle graphique probabiliste (appelé réseau bayésien soit « Bayesian Networks » en anglais et noté BN) est élaboré. L'avantage de cet outil est de pouvoir d'une part, visualiser les influences et les interdépendances des variables entre elles en intégrant les probabilités des événements associés à chacune d'elles, et, d'autre part, effectuer des inférences probabilistes sur les variables d'intérêt.

Deux étapes permettent de construire un BN :

- Une première visant à l'identification des variables : un état de l'art a été effectué sur la base de la bibliographie disponible, afin d'identifier les différents processus intervenant au niveau de l'activité ainsi que les variables susceptibles d'avoir une incidence sur ces processus et au final sur les résultats ADN en termes de quantité et de qualité pour un cas spécifique.



- Une deuxième permettant de définir les relations entre les variables identifiées et insérer des données : le rôle de chacune des variables est étudié et leurs interdépendances mises en évidence. Une table de probabilités conditionnelles est associée à chaque variable. Les probabilités déterminent quantitativement la dépendance entre les variables.

### 3<sup>ème</sup> chapitre : Sélection des variables d'intérêts

Certaines variables ont une influence plus marquée que d'autres sur le résultat de l'évaluation. Ces dernières nécessitent des recherches plus approfondies. Ce chapitre étudie l'impact de chaque variable sur le résultat de l'évaluation afin de cibler les variables d'intérêt puis dans un second temps, se focalise uniquement sur ces dernières en déterminant si l'acquisition de nouvelles données est nécessaire.

### 4<sup>ème</sup> chapitre : Etude du type de données à utiliser

Plus les données sont spécifiques au cas (c'est-à-dire que les données correspondent aux circonstances du cas, aux personnes d'intérêt, aux activités alléguées, au moyen de prélèvement et au kit d'extraction d'ADN utilisé pour analyser la trace), plus la confiance dans le résultat de l'évaluation est élevée. Cependant, pour certaines observations, il est possible que l'impact de l'utilisation de données spécifiques au cas sur le résultat de l'évaluation par rapport à l'utilisation de données issues de la littérature soit négligeable. L'objectif de ce chapitre est de définir si les données à utiliser et/ou à ajouter doivent être spécifiques au cas. Ce chapitre est composé de deux sections :

- 1<sup>ère</sup> section : Acquisition des données spécifiques. L'objectif est de montrer comment acquérir les données spécifiques à un cas choisi, pour chaque variable identifiée.
- 2<sup>ème</sup> section : Impact du type de données utilisé sur le résultat de l'évaluation. Trois évaluations sont réalisées :
  - en utilisant uniquement les données issues de la littérature pour les variables testées,
  - en combinant les données issues de la littérature et les données spécifiques issues des expériences pour les variables testées,
  - ou encore en utilisant uniquement les données spécifiques.

Les résultats des évaluations sont ensuite comparés entre eux afin de définir s'il est préférable d'utiliser des données uniquement issues d'expériences spécifiques au cas et pour quelle variable.

---

5<sup>ème</sup> chapitre : Etude de l'efficacité de la méthodologie

Pour étudier l'efficacité de la méthodologie, différents cas ont été simulés sur la base de la proposition posée soit par le procureur, soit par la défense. La « bonne » proposition que les observations doivent soutenir par rapport à l'alternative est donc connue. Une évaluation est ensuite réalisée en appliquant la méthodologie proposée. Pour chaque cas simulé, le résultat de l'évaluation est étudié afin de savoir si les observations soutiennent la « bonne » proposition ou non et avec quelle force. Enfin, une calibration du système est effectuée sur la base des LR obtenus et le report ou non du LR dans un rapport d'expertise est discuté.

6<sup>ème</sup> chapitre : Vers une application pratique de la méthodologie

L'objectif principal de ce travail est d'utiliser la méthodologie proposée dans le contexte d'une application pratique afin de déterminer son intérêt et ses avantages mais aussi d'exposer ses limites et ses conditions d'application. Cette analyse permet de proposer des perspectives en termes d'utilisation pratique mais aussi d'amélioration de la méthodologie.

## I.4 Originalité de la thèse

L'originalité de cette thèse est de proposer une méthodologie opérationnelle, permettant d'effectuer des évaluations des résultats considérant des propositions au niveau de l'activité. La méthodologie se veut applicable à de nombreux cas pour lesquels des traces ADN sont prélevées et dont l'intérêt est une évaluation selon des propositions mettant en jeu des activités.

Cette méthodologie se base sur les résultats obtenus dans la thèse de Pun (2016) qui se composait de quatre étapes principales :

- dans un premier temps, un réseau bayésien, adapté au scénario d'un vol de voiture a été construit,
- dans un second temps, des données sur le transfert, la persistance et le bruit de fond d'ADN prélevé à différents endroits de la voiture, dans différentes conditions environnementales, selon différents jeux de propositions ont été acquises,
- dans un troisième temps, une analyse de sensibilité du BN a été effectuée afin d'identifier les variables ayant le plus d'impact sur le résultat de l'évaluation. Le transfert et le bruit de fond ont été les variables identifiées lorsque la défense allègue une présence légitime de l'ADN du POI,

- enfin, l'investissement en termes d'acquisition de données supplémentaires a été étudié. Seul le transfert nécessitait cet investissement pour le scénario considéré.

Les résultats de Pun (2016) montrent qu'il n'est pas nécessaire d'acquérir des données sur toutes les variables à prendre en compte, ce qui est, de toute façon, difficile de réaliser dans la pratique d'un laboratoire.

Partant de ce postulat, le premier objectif de ce travail sera d'identifier au préalable les variables importantes qui auront une influence sur le déroulement des processus et mécanismes de transfert et au final sur les résultats de l'évaluation plutôt que de réaliser des expériences afin d'obtenir des données sur l'ensemble des variables d'un cas particulier, comme cela a été fait dans la thèse de Pun (2016). L'approche est donc différente de la thèse de Pun (2016).

Dans ce travail, nous considèrerons également que le LR qui devrait être utilisé dérive de l'utilisation de données spécifiques au cas mais qu'il n'est pas nécessaire d'obtenir des données spécifiques au cas sur toutes les variables. Le deuxième objectif sera donc de déterminer **au préalable** si l'acquisition de données supplémentaires est indispensable et également de définir si ces données doivent être spécifiques au cas considéré ou si elles peuvent être issues de la littérature.

Le but est de limiter le nombre d'expériences sans pour autant dégrader la confiance à accorder aux résultats. Cette méthodologie pourra ainsi avoir une portée pratique facilitée.

## I.5 Choix du cas

Comme il est proposé d'étudier l'activité ayant entraîné le dépôt de la trace ADN si celle-ci est d'origine biologique inconnue ou si celle-ci contient d'infimes quantités d'ADN, un cas impliquant l'analyse de **traces ADN de contact** est donc préféré.

La méthodologie est élaborée sur la base d'un cas d'agression au couteau. L'exemple de l'affaire Anderson n'a pas été retenu car il ne fait pas partie des scénarios les plus courants en Suisse. Le cas du vol de voiture étudié dans la thèse de Pun (2016) n'est également pas étudié, mais un parallèle sera fait entre les résultats obtenus dans ce travail et ceux obtenus dans la thèse de Pun (2016).

Les supports sur lesquels des traces ADN sont recherchées peuvent être :

- Des objets en lien avec les personnes d'intérêt : ADN prélevé sur les vêtements du suspect et/ou de la victime, ADN sur des draps, un divan, une voiture, un téléphone portable...
- Des sachets pouvant contenir des produits d'intérêt : par exemple, dans le cadre d'un trafic de stupéfiants, ADN prélevé sur les supports de type « plastique » tels que les sachets, les rubans adhésifs...
- Des objets manipulés lors d'actions pour lesquelles des enquêtes sont menées : les supports « durs », tels que les rebords de fenêtres ou les meubles dans les cas de cambriolages, les différentes armes potentielles (couteau, arme à feu, batte de base ball, par exemple).

Or, selon le rapport annuel de la statistique policière de la criminalité de 2018 en Suisse, il y a eu 43'773 cas d'infractions avec violence relevés au cours de l'année. Parmi ceux-ci, des homicides et tentatives d'homicides (199) dont la majorité ont été commis à l'aide d'une arme coupante/tranchante (109). Concernant les cas de lésions corporelles graves (585), la majorité de ceux-ci résulte de la pure violence physique (308) et de l'utilisation d'une arme coupante/tranchante (107) ou de l'utilisation d'une arme autre que contondante ou arme à feu (119). De plus, selon l'Office des statistiques national (« Office for National Statistics ») en Angleterre et au Pays de Galles, le nombre d'agressions par arme coupante/tranchante ne cesse d'augmenter depuis plusieurs années. Environ 43'000 agressions par arme coupante/tranchante (sans compter Manchester) ont été répertoriées entre avril 2018 et mars 2019, soit une augmentation de 8% par rapport à l'année précédente (avril 2017-mars 2018) et d'environ 42% par rapport à 2011 (avril 2010-mars 2011).

Sur la base de ces statistiques, le cas d'application portera sur des prélèvements d'ADN effectués sur des objets coupants/tranchants. L'objet choisi est un couteau. Les résultats de ce travail pourront potentiellement mieux aider la Cour à mettre en cause ou pas des personnes d'intérêt lorsque différentes activités liées au couteau sont alléguées.

Le cas considéré dans le cadre de cette thèse, appelé « Cas M. X », est le suivant : « Les secours sont appelés pour intervenir dans un immeuble à la suite d'un appel de voisins. Les secouristes trouvent le corps de M. X poignardé de deux coups de couteau. Un couteau à manche lisse, dont la lame est ensanglantée, est retrouvé près de la victime. Celui-ci est conservé dans un sachet papier. Le couteau est ensuite transmis au laboratoire qui effectue

---

un prélèvement d'ADN sur le manche au moyen d'un écouvillon afin de l'analyser. La mort se situe environ deux jours avant la découverte du corps. Durant l'enquête, M. A est mis en cause. Il s'agit donc de la personne d'intérêt appelée POI. D'après les proches de la victime, le couteau appartient à M. X. Aucun élément d'enquête ne permet de mettre en cause une personne apparentée au suspect. ».

La mission confiée par le procureur à l'expert est la suivante :

« Veuillez procéder aux analyses ADN de l'arme afin de déterminer si l'échantillon remis comporte du matériel ADN. Le cas échéant, nous vous demandons de comparer le profil obtenu avec le profil ADN de M. A. Veuillez finalement indiquer dans quelle mesure vos résultats appuient l'une ou l'autre des propositions décrites ci-après. ».

Le contexte de l'affaire et les propositions formulées jouent un rôle fondamental dans l'évaluation des résultats. Pour ce cas, deux scénarios possibles seront étudiés.

Dans le premier scénario, M. A indique qu'étant caissier dans un commerce situé près du domicile de la victime, il a pu serrer la main de l'agresseur appelé par la suite AO (« Alternative Offender »). Le couple de propositions indiqué dans la mission est donc :

- Hypothèse  $H_p$  : « Le POI a poignardé la victime »
- Hypothèse  $H_{d1}$  : « Le POI a serré la main de l'AO qui a poignardé la victime avec le couteau retrouvé sur les lieux ».

Où  $H_p$  désigne la proposition du procureur et  $H_d$  la proposition alléguée par la défense.

Ce travail s'appuie principalement sur ce premier scénario qui sera étudié tout au long de cet écrit.

Dans le deuxième scénario, M. A dit ne rien à voir avec l'affaire et ne suggère aucun mécanisme alternatif à la présence de son ADN. Il ne connaît pas la victime et n'a jamais vu le couteau retrouvé sur les lieux. Le couple de propositions est donc :

- Hypothèse  $H_p$  : « Le POI a poignardé la victime »
- Hypothèse  $H_{d2}$  : « Le POI n'a jamais touché le couteau retrouvé sur les lieux, ce n'est pas lui mais une autre personne inconnue, appelée AO, qui a poignardé la victime ».

Ce deuxième scénario est étudié seulement dans les Chapitres II et III de ce travail afin de mettre en évidence et discuter de l'impact de la formulation des propositions sur le résultat de l'évaluation, impact qui a également été mis en évidence dans la thèse de Pun (2016).

## I.6 Contributions associées à ce travail

Ce travail fait suite à un travail de master dont les résultats ont été publiés (Samie et al., 2016). Il a fait l'objet de plusieurs articles publiés (Annexe I) ainsi que d'articles en cours de soumission (Annexe II) cités ci-dessous. Ceux-ci font partie intégrante de ce travail mais ne sont pas tous détaillés dans le corps de ce manuscrit. Les différents résultats obtenus au cours de ce travail ont également été présentés dans le cadre de plusieurs conférences (voir ci-après) grâce au soutien financier de l'Ecole des Sciences Criminelles, de l'Association des diplômés en Sciences Criminelles et de la Fondation pour l'Université de Lausanne. Un article (Taylor et al., 2019) a été publié en lien avec la problématique de ce travail et auquel une contribution a été apportée. Celui-ci est marginal par rapport au présent travail.

### Articles publiés (Annexe I)

Taylor D., Biedermann A., Samie L., Pun K.-M., Hicks T., Champod C., *Helping to distinguish primary from secondary transfer events for trace DNA*, Forensic Science International: Genetics. 28 (2017) 155–177.

Samie L., Champod C., Glutz V., Garcia M., Castella V., Taroni F. *The efficiency of DNA extraction kit and the efficiency of recovery techniques to release DNA using flow cytometry*, Science & Justice 59. (2019) 405-410. Il est noté Samie et al. (2019<sub>(a)</sub>) dans la suite de ce travail.

### Articles soumis à publication (Annexe II)

Samie L., Champod C., Taylor D., Taroni F., *The use of Bayesian Networks and simulation methods to identify the variables impacting the LR in stabbing cases evaluated given activity level propositions*, Forensic Science International: Genetics. (2019). Il est noté Samie et al. (2019<sub>(b)</sub>) dans la suite de ce travail. Accepté avec révision mineure.

Samie L., Champod C., Taroni F., *Estimating the quantity of transferred DNA in primary and secondary transfers*. Science & Justice. (2019). Il est noté Samie et al. (2019<sub>(c)</sub>) dans la suite de ce travail. Accepté, en production.

### Présentation poster

Samie L., Champod C., Taroni F., *Application of a methodology to evaluate DNA results with propositions at the activity level using Bayesian network*, 28<sup>th</sup> Congress of the International Society for Forensic Genetics (ISFG), Prague, 2019.

### Présentations orales

Samie L., Champod C., Taroni F., *Evaluation of DNA Results considering propositions at the activity level: how to identify variables impacting the LR?*, 70th Annual Scientific Meeting of American Academy of Forensic Sciences (AAFS), Seattle, 2018.

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

---

Samie L., Evaluation des résultats ADN considérant des propositions au niveau de l'activité, Concours MT180, 2017. <https://www.youtube.com/watch?v=hnsxlnf9jCk&t=72s>.

Samie L., Champod C., Taroni F., *Stabbing and DNA transfer*, 26<sup>th</sup> Congress of the International Society for Forensic Genetics (ISFG), Cracovie, 2015.

## Chapitre II - Elaboration d'un modèle probabiliste

---

L'objectif de ce chapitre est de décomposer le scénario à travers la construction d'un BN qui est un modèle graphique probabiliste permettant de visualiser les variables et les dépendances probabilistes entre elles. Ce type d'approche possède des avantages certains, comme souligné dans la littérature (Coquoz et al., 2013 ; Taroni et al., 2014; Hu et al., 2014, Evett et al., 2002, Biedermann et Taroni, 2012). En effet, ils permettent de modéliser graphiquement une problématique simple ou complexe, en mettant en évidence les influences et les interdépendances des variables entre elles, et d'effectuer des inférences probabilistes en combinant la théorie des probabilités et la théorie des graphes.

Les BNs sont devenus des outils essentiels pour l'évaluation des observations au niveau de l'activité comme signalé par Taylor et al (2018). Ils ont également été utilisés dans les thèses de Pun (2016) et Palmer (2016).

Taylor et al. (2018) ont défini 7 étapes, déjà présentées par Aitken et al. (1996), permettant de construire un BN utilisable pour les cas de biologie forensique comportant des propositions au niveau de l'activité :

- définir les propositions,
- définir les activités,
- définir les groupes de variables dont les observations sont similaires,
- définir les observations,
- définir les variables du transfert et de la persistance,
- définir les relations entre les nœuds,
- informer le réseau et vérifier la cohérence de celui-ci.

La construction du BN utilisé dans ce travail respecte ces sept points et sont présentés dans cette partie, les propositions et les activités ayant été définies précédemment (Voir Chapitre I.5).

Dans un premier temps, les variables utilisées pour construire le BN sont sélectionnées. Ces variables sont celles qui, d'après l'état de l'art, ont une influence sur les résultats ADN.

Dans un deuxième temps, les relations entre les variables sont définies et les données permettant d'informer ces variables sont présentées.



## II.1 Sélection des variables

Entre l'action d'intérêt et l'obtention des résultats ADN, différents mécanismes interviennent qui peuvent mener à une perte ou un gain plus ou moins important de quantité d'ADN. Ces mécanismes sont :

- Le transfert d'ADN ou déposition de la trace.
- La persistance ou capacité de la trace à se maintenir (ou à subsister) sur le support.
- La contamination, plus communément appelée « bruit de fond », qui est de l'ADN présent sur la surface prélevée autre que celui des personnes impliquées. Nous parlerons de bruit de fond dans la suite de cette thèse.
- La collecte et l'analyse de la trace en laboratoire.
- La pollution, qui est de l'ADN qui n'aurait pas dû se retrouver dans la trace. Elle est communément appelée « contamination », qui est à différencier du « bruit de fond ».
- Les erreurs de laboratoire.

De nombreuses variables peuvent influencer sur ces mécanismes et avoir un impact sur la quantité d'ADN et la qualité des profils ADN observés et in fine sur le résultat de l'évaluation.

Ces variables ont fait l'objet de nombreux travaux précédents. Citons :

- Pun (2016) a réalisé une revue bibliographique de ces variables dans le cadre de sa thèse.
- Meakin et Jamieson (2013) ont écrit un article bibliographique présentant l'ensemble des travaux relatifs au transfert et à la persistance d'ADN publiés jusqu'en 2013.
- Van Oorshot et al. (2018) ont par la suite publié, à la fin de l'année 2018, une revue de la littérature sur ces variables.

La littérature sur les variables intervenant dans les phénomènes de transfert, persistance, bruit de fond, collecte en laboratoire et pollution, est détaillée en Annexe III. Elle reprend les articles présentés dans les trois travaux précédemment cités, ainsi que les articles nouvellement publiés depuis 2018. La revue publiée en Annexe III offre une structure permettant de faire un bilan, pour chaque variable, des données acquises et sous quelle forme (données quantitatives, qualitatives) et de leur influence potentielle sur le résultat de l'évaluation. Le choix des variables utilisées pour construire le BN, présentées ci-après, est basé sur les éléments qui se dégagent de cette revue.

### II.1.1 Le transfert

Le premier mécanisme influençant les résultats ADN est le transfert qui est la déposition d'ADN d'une surface sur une autre.

Il existe différents types de transfert. Le transfert d'ADN peut être direct (dit « primaire ») ou indirect (dit « multiple »).

Le transfert est direct lorsque les cellules du donneur A sont transférées directement du donneur sur la surface d'intérêt B, il n'y a pas de surface intermédiaire.



Il est indirect lorsqu'il y a une ou plusieurs surfaces entre le donneur et la surface réceptrice d'intérêt. Le transfert est secondaire s'il y a une seule surface intermédiaire C : du donneur vers la surface intermédiaire puis de celle-ci vers la surface d'intérêt, soit deux transferts successifs.



Sur ce même principe, le transfert est tertiaire s'il y a deux intermédiaires, quaternaire s'il y a trois intermédiaires et ainsi de suite.

Selon les propositions qui sont alléguées par les différentes parties, les types de transfert mis en jeu peuvent être différents.

Les variables qui impactent le processus de transfert les plus étudiées à ce jour sont au nombre de quatre : le statut du donneur, le type de contact, le temps de contact et le type de surface.

#### **1. Statut du donneur**

Le statut décrit la capacité d'un individu à déposer des traces contenant son ADN sur une surface, par contact, suite à un transfert direct, sans l'intermédiaire d'un fluide biologique comme le sang, la salive ou le sperme.

Selon plusieurs auteurs (Lowe et al., 2002; Farmen et al., 2008), il est possible de distinguer un bon donneur d'un mauvais. La littérature fournit plusieurs définitions sur ce qu'est un « bon » ou « mauvais » donneur, dont la rédaction et le contenu varient selon les auteurs, introduisant une ambiguïté sur le statut de « bon » ou « mauvais » donneur. Ainsi, selon la définition considérée, les unités de « bons donneurs » observables peuvent être différentes.

Daly et al. (2012) indique que le pourcentage de bons donateurs, « bon donneur » étant défini comme un individu pour qui un profil complet a pu être obtenu, est seulement de 13%. Ce pourcentage est de 22% si un « bon donneur » est défini comme un individu pour lequel un profil exploitable de 6 loci ou plus est obtenu. Pun (2016) considère un « bon donneur » lorsque la quantité d'ADN transférée est d'au moins 5625 pg, seuil de quantité d'ADN nécessaire pour obtenir un profil complet avec les techniques définies dans sa thèse. Le pourcentage de bons donateurs est alors de 2% selon cette définition, qui est dépendante de la méthode utilisée pour amplifier l'ADN. Kanokwongnuwut et al. (2018) définissent, sur la base des observations faites dans leur étude, 3 catégories de donateurs : les donateurs forts, les donateurs intermédiaires et les donateurs faibles.

Cependant, Phipps et Petricevic en 2007 indiquent, qu'il est plus difficile de qualifier des personnes comme « bons » ou « mauvais » donateurs que ne le laisse penser l'étude de Lowe et al., 2002. En effet, la qualité du donneur varie de façon significative selon les individus (variabilité inter individus) mais aussi pour une même personne (variabilité intra individu). Cela a été observé par plusieurs groupes de recherche (Pesaresi et al., 2001; Van Oorschot et al., 2003; Bright et Petricevic, 2004; Goray et al., 2016; Taylor et al., 2016; Pun, 2016). Une personne peut être bon donneur à certains moments et mauvais donneur à d'autres. Cependant, une personne peut être, en moyenne, meilleur donneur qu'une autre personne (Pun, 2016).

Différents facteurs peuvent expliquer la variabilité inter individus : le sexe, l'âge ou encore les habitudes du donneur. Ces facteurs sont plus particulièrement étudiés car ils ont un impact potentiel sur le statut du donneur :

- **Sexe du donneur** : Lacerenza et al. (2016) ont effectué des prélèvements ADN sur chacune des deux mains (palmes et doigts) de 60 individus (30 hommes et 30 femmes). Ils ont observé que la quantité d'ADN présente sur les mains des hommes était plus importante que chez les femmes. Ils ont également constaté que les femmes avaient plus d'ADN étranger sur les mains que les hommes.

Cependant, Bontadelli (2009) contredit ces résultats. Elle n'a observé aucune influence du sexe sur la quantité d'ADN présente sur les mains en étudiant les prélèvements effectués sur 30 femmes et 10 hommes entre 21 et 24 ans.

- **Age du donneur** : Une étude de Poetsch (2013), montre qu'il existe une corrélation entre la quantité d'ADN transféré, la qualité des profils provenant de traces digitales

déposées et l'âge des individus. Cette étude constate également des différences dans les résultats selon qu'il s'agit d'un enfant ou d'un adulte. Ainsi, sur une population de 209 individus composée de 47 enfants de moins de 11ans, 32 adolescents entre 11 et 20ans, 81 adultes entre 20 et 60 ans et 49 personnes âgées de plus de 60 ans qui ont pressé leurs mains sur un support en plastique pendant 10 secondes, les résultats montrent qu'un profil complet est obtenu dans 75% des cas chez les enfants, 9% chez les adolescents, 25% chez les adultes et 8% chez les personnes âgées. Les résultats montrent que la probabilité d'obtenir un profil ADN complet à partir de traces digitales décroît avec l'âge. Toutefois le résultat obtenu pour la tranche d'âge 11-20 ans pose question mais aucune explication n'est avancée dans l'article.

A contrario, selon Raymond et al. (2008), l'âge n'influencerait pas la capacité à transférer son ADN. L'âge des volontaires n'est cependant pas précisé dans l'article.

- **Les habitudes du donneur** telles que le lavage des mains, la fréquence de toucher du visage ou des cheveux ou encore la pratique d'un sport sont des actions susceptibles d'avoir une influence sur le transfert d'ADN.

Le lavage des mains est certainement le facteur le plus étudié à ce jour. Celui-ci augmente la probabilité de retrouver un profil simple car il enlève l'ADN étranger présent sur les mains (Pesaresi et al., 2003). Cependant, une étude de Lowe et al. (2002), montre que plus le lavage est récent plus le profil obtenu tend à être incomplet.

Certaines études émettent l'hypothèse que la sueur faciliterait le transfert des cellules épithéliales (Wickenheiser, 2002 ; Bright et Petricevic, 2004 ; Van den Berge et al., 2016). Pun (2016) conclue que, d'après ses résultats, la sueur joue effectivement un rôle dans le transfert de l'ADN, l'écoulement menant à un chargement en ADN de la sueur elle-même, une fois sortie des orifices sudoripares.

Les expériences menées par Lacerenza et al. (2016) n'ont pas montré d'influence des habitudes personnelles et hygiéniques sur la quantité d'ADN présente sur les mains des donneurs.

Seule l'habitude de se toucher fréquemment les cheveux semble augmenter la quantité d'ADN récupérée.

Goray et al. (2016) quant à eux, ont observé que le moment de la journée et la dominance de la main ne semblaient pas être des facteurs ayant un effet significatif sur les quantités d'ADN.

Outre le statut du donneur lui-même, ce qui influence le résultat ADN est **le statut relatif entre les personnes impliquées**. Par exemple, dans une situation 1, un suspect dit « bon donneur » serre la main d'un agresseur dit « mauvais donneur ». Dans une deuxième situation, un suspect dit « mauvais donneur » serre la main d'un agresseur dit « bon donneur ». La situation 1 et la situation 2 ne donneront pas un type de profil similaire pour ce qui est des proportions dans le mélange d'ADN, comme en témoignent Fonnøløp et al. (2017).

Un autre aspect à mettre en relation avec le statut de donneur, est **la présence sur les mains d'ADN étranger** à celui-ci. La capacité du donneur à transférer cet ADN dépend alors de son statut (Szkuta et al., 2017).

**En conclusion**, le statut du donneur est une variable importante à prendre en compte. Elle ne peut cependant pas être définie suivant des catégories telles que « Bon/Mauvais donneur ». Ce statut dépend des caractéristiques propres à l'individu (âge, sexe, habitudes) ainsi que d'évènements ponctuels tels que le lavage des mains ou la transpiration qui influencent la capacité de transférer de l'ADN à un moment précis. De plus, le statut relatif des donneurs d'intérêt, dans notre cas le POI et l'AO, semble impacter le résultat.

Le statut des donneurs sera donc pris en compte dans la construction du BN en créant les variables « Quantité d'ADN sur les mains » et « Proportion de transfert » pour chaque contributeur d'intérêt, le POI et l'AO. Une distribution de la quantité d'ADN sur les mains et de la proportion de transfert doit être obtenue (Voir Chapitre IV.1.2) afin de conserver l'information quant à la variabilité intra-individuelle. Il sera également possible d'utiliser des quantités d'ADN différentes pour l'AO du POI, permettant de conserver l'information quant au statut relatif des donneurs.

## 2. Type de contact

Il existe trois types de contact :

- Contact par pression (toucher en appuyant),
- Contact passif (simple toucher),
- Contact par friction (toucher avec un mouvement de frottement).

Goray et al., 2010 ont montré en 2010 que la probabilité de transfert varie en fonction du type de contact. Le type de contact est une variable qui sera retenue.

---

### **3. Temps de contact**

Selon Van Oorschot (1997), le temps de contact n'aurait pas d'influence sur la quantité d'ADN déposée. L'expérience portait sur des temps de contact de 5 secondes, 30 secondes, 3 minutes et 10 minutes et les objets utilisés étaient des tubes en polypropylène. Une autre étude utilisant des câbles métalliques montre qu'un contact même bref avec l'objet est suffisant pour transférer de l'ADN (Saravo et al., 2004).

Au vu des résultats obtenus par Van Oorschot (1997), ce facteur n'est pas retenu.

### **4. Type de surface**

Les surfaces réceptrices sont de nature très diverses. En général, celles-ci sont classées en deux catégories, les surfaces poreuses (rugueuses) et les surfaces non poreuses (lisses). Mais tous les objets ont des propriétés de surface spécifiques qui dépendent principalement des matériaux qui les composent et de leur forme.

Pour de nombreux auteurs (Daly et al., 2012; Goray et al., 2010; Fonnøløp A. et al., 2015), le type de surface influence la quantité d'ADN « de contact » transférée et la qualité des profils obtenus.

Cependant, d'autres auteurs, comme Alessandri et al. (2003), jugent ce facteur comme non significatif. Cale et al. (2015) qui ont étudié l'influence de ce facteur sur la possibilité de transfert secondaire arrivent à la même conclusion. Dans leur étude, les surfaces étudiées étaient des manches de couteaux, soit en plastique lisse, soit en plastique rugueux. Aucune différence significative dans les quantités prélevées n'a été observée entre les deux surfaces. L'incidence des deux types de surface sur la qualité des profils n'a pas été étudiée.

Selon les études, il semblerait donc que le type de surface peut avoir une influence sur les observations ADN, en terme de quantité d'ADN et de qualité des profils ADN, en particulier si les surfaces sont faites de matériaux radicalement différents, par exemple un tissu *vs* une surface en verre. Cependant cette influence serait moindre entre deux types de surfaces faites d'un matériel similaire telles que deux surfaces en plastique, l'une plutôt lisse et l'autre plutôt rugueuse. Le « Type de surface » est retenu pour la construction du BN.

**Les variables retenues concernant le phénomène de transfert sont donc la quantité d'ADN disponible sur les mains des individus et la proportion de transfert qui reflètent le statut des donneurs, le type des surfaces de transfert ainsi que le type de contact.**

### II.1.2 La persistance

La persistance est la capacité de la trace à se maintenir sur un support. Elle intervient entre le moment du transfert et celui de la collecte de la trace. Les principaux facteurs étudiés pouvant influencer la persistance de l'ADN « de contact » sur un support sont : le temps écoulé entre le dépôt et le prélèvement, l'environnement et le type de la surface.

#### **1. Temps**

Plus le temps écoulé entre le dépôt de la trace et le prélèvement est long, plus la persistance des traces ADN diminue. L'ADN se dégrade et la qualité des profils obtenus est moins bonne (Raymond et al., 2009). La durée entre deux phénomènes de transfert ou entre un phénomène de transfert et le prélèvement est donc prise en compte dans le BN.

#### **2. Environnement**

Les résultats d'une expérience menée par Raymond et al. (2009) montrent qu'aucun profil ADN n'a pu être retrouvé deux semaines après le dépôt dans un milieu ouvert (en extérieur). A l'inverse, un profil ADN peut être observé jusqu'à six semaines si la trace est déposée en milieu fermé ou protégé (en intérieur). Les conditions environnementales comme par exemple l'humidité, la température, l'exposition aux UV par exemple, favorables ou défavorables à la conservation de l'ADN, ont un impact sur la persistance. L'environnement est une variable identifiée pour la construction du BN.

#### **3. Type de surface**

De nombreuses études (Goray et al., 2010; Sczech, 2011; Van Oorschot et al., 2014; Pfeiffer et Wiegand, 2017; Oldoni et al., 2015, Oldoni et al., 2016; Fonnelop et al., 2015; Meakin et al., 2015; Michelle Breathnach, 2016; Matte et al., 2012) se sont intéressées à la persistance de l'ADN sur différents types de support montrant ainsi que la persistance de l'ADN dépend du type de surface.

**En conclusion pour le phénomène de persistance**, le type de surface, le temps écoulé et les conditions environnementales vont avoir une influence sur la qualité et la quantité d'ADN transféré. **Les variables retenues sont le temps écoulé et les conditions dites « environnementales »**. Le type de surface est un facteur pris en compte dans les conditions dites « environnementales » afin de déterminer si les conditions étaient « favorables » ou « défavorables » à la conservation de l'ADN. En effet, si une surface est susceptible d'avoir été touchée ou manipulée entre les faits et la saisie de l'objet, cela se traduira par des conditions « défavorables ». A l'inverse, une surface qui n'a potentiellement pas été touchée ou manipulée, cela se traduira par des conditions dites « favorables ».

### II.1.3 Le bruit de fond

Le bruit de fond est défini comme de l'ADN autre que celui du suspect et qui n'est pas en relation avec l'activité précise en investigation. Quelle est la proportion du bruit de fond par rapport au profil de la personne ayant fait l'action d'intérêt ? Généralement le bruit de fond est associé à la fraction minoritaire de la trace, constituée d'un ou plusieurs contributeurs. Mais si le profil du bruit de fond correspondant à une tierce personne (personne étrangère au cas) est majoritaire et couvre celui de la personne ayant fait l'action d'intérêt (qui se retrouve dans la fraction minoritaire), le risque est de considérer cette tierce personne comme impliquée dans les actions en enquête.

Des études sur cette thématique ont été publiées récemment par Fonnelløp et al. (2015), Van den Berge et al. (2016) et Lehmann et al. (2015). Ces derniers ont notamment étudié l'influence du bruit de fond sur la capacité à détecter le profil d'une personne cible. Ils ont observé que si le bruit de fond est constitué de l'ADN d'une seule personne, aucune influence significative de celui-ci sur la capacité à détecter le profil cible n'est observé, quel que soit l'objet. En revanche, plus le nombre de contributeurs au bruit de fond augmente, plus le profil cible est difficile à détecter.

Plusieurs études ont été réalisées afin d'avoir des données sur l'exploitabilité des traces ADN en bruit de fond présent sous les ongles (Cook et Dixon, 2007; Malsom et al., 2009 ; Matte et al., 2012), sur les habits (Breathnach et al., 2016 et Stouder et al., 2001), sur la peau au niveau du cou (Graham et Rutty, 2008; et Rutty, 2002) et sur des couteaux (Meakin et al., 2017).



L'étude de la quantité/qualité relative entre les différents contributeurs (contributeur majeur/contributeur minoritaire) peut permettre dans certains cas de différencier le bruit de fond des profils d'intérêt (Goray et Van Oorschot, 2013 et Goray et Van Oorschot, 2015).

Le bruit de fond a donc une influence sur la quantité d'ADN prélevée : plus il y a de bruit de fond, plus la quantité d'ADN prélevée sera grande. Il intervient aussi sur la qualité du profil de la trace : plus le bruit de fond est important, plus le nombre de contributeurs peut être élevé. De plus, plus la quantité de bruit de fond est élevée, plus celui-ci aura tendance à masquer le profil d'intérêt.

**Cette variable est sélectionnée comme pouvant avoir un impact sur le résultat de l'évaluation.**

#### II.1.4 La collecte et l'analyse de la trace en laboratoire

Les variables liées à la méthode de prélèvement et à l'analyse en laboratoire n'ont pas été étudiées dans la thèse de Pun (2016). Néanmoins, pour une même trace ADN, la qualité des profils dépend des outils et des protocoles utilisés : outils et techniques de prélèvements, taille de la surface prélevée, kits et protocoles d'extraction, kits d'amplification, systèmes d'électrophorèses capillaires ainsi que des protocoles de stockage.

##### **1. Outils et techniques de prélèvement**

Deux études (Cook et Dixon, 2007 et Malsom et al., 2009) sur le bruit de fond présent sous les ongles d'individus montrent des résultats différents qui s'expliquent par l'utilisation d'outils de prélèvements différents (Meakin et Jamieson, 2013).

Dans l'étude de Lacerenza et al. (2016) concernant le prélèvement d'ADN sur les mains, les auteurs ont observé des différences significatives de la quantité d'ADN prélevé selon la technique de prélèvement utilisée. Ainsi, pour des prélèvements effectués sur des mains, l'utilisation d'un écouvillon en coton humide permet de recueillir une plus grande quantité d'ADN que l'utilisation d'un écouvillon en coton sec ou d'une bande collante (« tape lift »). Dans l'étude de Noël et al. (2016), les auteurs ont remarqué que lorsqu'ils prélevaient l'ADN sur des sous-vêtements en coton avec un écouvillon, aucun profil n'était observable, contrairement au prélèvement d'une partie du sous vêtement en le coupant.

## **2. Kits et protocoles d'extraction**

De même, certains kits et protocoles d'extraction sont plus performants que d'autres et permettent d'obtenir de plus grandes quantités d'ADN et de limiter les pertes pouvant se produire lors de l'extraction. Cet effet ne doit pas être ignoré (Verdon, 2014).

## **3. Kits d'amplification et système d'électrophorèse capillaires**

Dans leur étude, Fonnøløp et al. (2015), obtiennent, pour des traces ayant de faibles quantités d'ADN, un pourcentage de profil ADN plus élevé que celui habituellement donné dans la littérature. D'après les auteurs, cette divergence pourrait s'expliquer par l'utilisation d'un kit d'amplification différent et d'un système d'électrophorèse capillaire plus sensible.

## **4. Persistance d'ADN : Conditions de stockage**

Breathnach (2016) montre que les conditions de stockage mises en place par son laboratoire sont optimales et qu'elles permettent de limiter la perte et la dégradation de l'ADN au cours du temps. L'ADN peut rester stable pendant plusieurs heures.

La perte d'ADN pouvant dépendre des outils et des méthodes de prélèvements, des kits et protocoles d'extraction d'ADN utilisées, les variables concernant l'efficacité des outils et des méthodes de prélèvements et l'efficacité des kits d'extraction sont retenues pour la construction du BN. Même si l'efficacité de la méthode de prélèvement était de 100%, une perte d'ADN peut avoir lieu si seulement une partie de la surface est prélevée, cette variable est donc également retenue. En revanche, l'efficacité du kit d'amplification et de l'électrophorèse capillaire ne sont pas retenues car les observations sont faites à partir des quantités d'ADN, étape ayant lieu avant l'amplification et la sortie des profils (cf. Partie II.1.2, « Observations »). La persistance d'ADN selon les conditions de stockage ne sera, par contre, pas une variable qui interviendra une fois la trace prélevée.

### II.1.5 La pollution

La pollution est constituée d'ADN qui n'aurait pas dû se retrouver dans la trace. Les pollutions sont possibles à tout moment entre l'instant où l'action d'intérêt a eu lieu et les traces pertinentes déposées jusqu'à l'obtention du profil.

Fonnøløp et al. (2016) ont rapporté, entre 2009 et 2015, 16 cas de pollution par la police sur 3545 cas étudiés. Au département de biologie forensique du « Norwegian Institute of Public

Health » (NIPH), « Department of forensic biology », 9 cas ont été reportés sur les 10590 traités. Basset et Castella (2018) ont, quant à eux, recensés 709 cas de pollution entre 2011 et 2015, ce qui représente une moyenne de 11,5 cas par an pour 1000 profils entrés dans la base de données ADN suisse. Sur tous ces profils ADN, 85% résultent de pollution par la police, 11% des employés du laboratoire et 3% d'autres sources. De nouvelles procédures au laboratoire ont ensuite été mises en place en 2015 et entre 2015 et 2017. 110 cas de pollution sur 11 852 profils envoyés dans la base de données ADN suisse ont été détectés dont seulement 3% provenaient des employés du laboratoire (Basset et Castella, 2019). Lapointe et al. (2015) ont observé 58 cas de pollution par la police du Québec lors de la comparaison entre 327 profils ADN de policiers et la base de données ADN canadienne en juillet 2014 qui contenait alors 31 071 profils ADN.

De nombreuses causes peuvent être à l'origine de pollutions (Gill, 1997; Gill et Kirkham, 2004; Coquoz et al., 2013) telles que :

- Le non-respect des consignes de précaution par les intervenants sur la scène de crime et en laboratoire (Lee et al., 1998; Lee et Ladd, 2001; Ruddy et al., 2000; Scherzinger et al., 1999). Par exemple si le masque couvrant le nez et la bouche n'est pas mis, il a été démontré que des profils ADN complets de personnes ayant parlé ou éternué à proximité d'objets pouvaient être obtenus à partir des prélèvements effectués sur ces objets (Port et al., 2006). Concernant les gants, si ceux-ci ne sont pas changés régulièrement, le risque d'un transfert d'ADN au travers de ces derniers et de contamination des pièces à conviction, est augmenté (Bussard, 2008).
- La mauvaise décontamination des emballages ou des instruments servant à la manipulation des pièces à conviction (Sullivan et al., 2004; Tamariz et al., 2006; Shaw et al., 2008). Ainsi, les pinceaux, non décontaminés ou mal décontaminés, utilisés pour la détection de traces digitales, peuvent être pollués et contaminer les pièces à conviction par transfert secondaire ou multiple (Proff et al., 2006). De même pour les gants et les instruments d'examen des pièces (Poy et Van Oorschot, 2006). Une étude a montré qu'un transfert d'ADN peut s'effectuer entre des pièces à conviction quand celles-ci sont rangées dans un même emballage. De même, un transfert d'ADN entre les différentes parties d'une même pièce est possible. L'ADN se trouvant sur la pièce à conviction peut se transférer sur l'emballage, qui, à son tour, peut transférer cet ADN sur une autre partie de la pièce (Goray et al., 2012). Ces

expériences ont été réalisées avec des traces riches déposées sur les pièces mais il est possible que ce processus se produise également pour des traces de contact. Taylor et al. 2016 ont également observé le cas d'un transfert d'ADN d'un scientifique sur le dossier que celui-ci tenait, puis de ce dossier sur les gants de la personne en charge de l'examen de la pièce à conviction et pour finir de ces gants à la pièce examinée.

Des études ont aussi montré la présence de pollution via les instruments d'autopsies dans les morgues (Rutty et al., 2000; Schwark et al., 2012; Cemper-Kiesslich et al., 2013).

- La pollution croisée : pollution ayant lieu par transfert d'ADN d'un échantillon à un autre lorsque ceux-ci sont analysés en même temps (Scherczinger et al., 1999).

Les pollutions peuvent être minimisées lorsque les règles et les normes de protection et de sécurité sont respectées telles que :

- le port de masques pour le nez et la bouche, de lunettes, de combinaisons et de gants qui doivent être changés régulièrement,
- la décontamination des instruments entre chaque analyse de chaque trace,
- la disposition des laboratoires évitant le croisement entre des pièces avec des traces riches en ADN et des pièces avec des traces pauvres en ADN.

La pollution est inévitable, par exemple lorsqu'il y a l'intervention des secours mais, dans ce cas elle devrait être documentée.

**La pollution, au sens défini précédemment, ne peut être dissociée du bruit de fond que très difficilement ; essentiellement pour des cas particuliers où un profil ADN obtenu est comparé au profil ADN des intervenants sur la scène de crime qui ne sont pas des personnes d'intérêt. Dans la suite de cette thèse, la pollution sera prise en compte dans la variable « Bruit de fond ».**

#### II.1.6 Les erreurs de laboratoires

Les erreurs de laboratoire sont à différencier de la pollution. Elles sont dues à des erreurs humaines. Elles résultent par exemple d'un mauvais étiquetage ou de tout événement qui induirait une inversion entre les traces. D'après Kloosterman et al. (2014), le NFI a relaté, entre 2008 et 2012, en moyenne 123 cas d'erreur par an dont la plupart n'ont pas eu de conséquences car elles ont pu être corrigées à temps.

**La possibilité d'une erreur de laboratoire peut** soit constituer la proposition de la défense, soit être prise en compte, par défaut, quelle que soit les propositions du procureur ou de la défense.

Les erreurs de laboratoire constituent des évènements très particuliers, susceptibles d'arriver mais qui, dans ce cas, peuvent généralement être mis en évidence par les différents éléments de l'enquête et/ou résultats de laboratoire et par les incohérences qui résultent de cette erreur.

**Pour ce travail, cette variable ne sera pas prise en compte dans le BN afin de se concentrer essentiellement sur la problématique des propositions au niveau de l'activité. Les personnes intéressées par cette question peuvent se référer aux travaux de Taroni et al. (2014).**

#### II.1.7 Discussion

Dans les revues publiées, les variables utilisées dans la construction d'un BN lié à la problématique d'un transfert secondaire d'ADN de contact allégué par la défense sont les suivantes :

- le transfert d'ADN et le statut du donneur (Fonneløp et al., 2017),
- le transfert d'ADN et le bruit de fond (Kokshroon et al.2017),
- le transfert d'ADN, la persistance comprenant le temps écoulé et le bruit de fond (Szkuta et al., 2018).

Les BNs proposés étaient adaptés pour un cas particulier, ce qui peut expliquer l'absence de certaines variables comme la persistance si les actions alléguées ont été effectuées peu de temps après les faits, tout comme les prélèvements ADN. Le bruit de fond, en fonction de la surface prélevée peut avoir été considéré comme nul.

Pun (2016) a proposé un BN qui peut être adapté à un plus grand nombre de cas. Les variables retenues pour la construction de son BN sont : le statut du donneur, la persistance, le bruit de fond, la pollution, le transfert et la localisation du prélèvement.

Il est noté que dans ces quatre BN, la perte d'ADN liée aux analyses ADN n'est pas prise en compte. Dans ce travail, toutes les variables citées sont retenues pour la construction du BN, sauf la localisation du prélèvement. En effet, dans le scénario proposé, un seul prélèvement

est effectué sur la surface d'intérêt, ici le couteau. Il est cependant possible de prendre en compte cette variable si cela est pertinent pour le cas étudié.

### **Conclusion**

Suite à la revue de littérature, les processus et variables associées retenus pour la construction du BN sont donc :

- le transfert : les variables à considérer sont le statut du donneur (comprenant la quantité d'ADN présent sur les mains du donneur ainsi que la proportion d'ADN transféré), le type de contact, le type de surface,
- la persistance : les variables sont le temps écoulé, l'environnement (comprenant le type de la surface) ainsi que la dégradation,
- le bruit de fond, qui comprend également la pollution,
- les outils d'analyse : efficacité des outils de prélèvement des traces, taille de la surface prélevée et efficacité des kits d'extraction.

La prise en compte des outils d'analyse comme variable est nouveau et l'impact de cette variable sur le LR n'a, à notre connaissance, pas été étudié.

## II.2 Relation entre variables et insertion des données

Les variables étant sélectionnées, il faut préciser :

- les propositions (dont nous allons faire un rappel),
- les observations finales,
- les relations entre les variables, les propositions et les observations,
- les états des variables,
- les données informant le BN.

### II.2.1 Les propositions

Les propositions sont celles déterminées par les différentes parties. Dans notre scénario il s'agit des positions du procureur ( $H_p$ ) et de la défense ( $H_d$ ).

Pour rappel, dans le cas de M. X (*Voir partie I.5*) retrouvé poignardé chez lui, les propositions sont les suivantes :

- Hypothèse  $H_p$  : « M. A a poignardé la victime ».

- Hypothèse H<sub>d1</sub> (dans le premier scénario) : « M. A a serré la main de l'agresseur qui a poignardé la victime avec le couteau retrouvé sur les lieux ».
- Hypothèse H<sub>d2</sub> (dans le deuxième scénario) : « M. A n'a jamais touché le couteau retrouvé sur les lieux, ce n'est pas lui mais une personne inconnue, non apparentée à M. A, qui a poignardé la victime ».

## II.2.2 Les observations

Deux types d'observations peuvent être faites suite à l'analyse d'une trace ADN :

- **Des observations qualitatives**

Elles peuvent être obtenues à l'aide d'un système d'électrophorèses capillaires sous forme d'électrophorégrammes, appelés profils ADN, sur lequel des pics sont observés.

La qualité d'un profil est déterminée par sa capacité à pouvoir discriminer les profils des personnes à l'origine de la trace et à fournir un profil permettant d'alimenter des bases de données. Elle dépend de la hauteur des pics mesurée en Unité de Fluorescence Relative (RFU), du nombre de loci, du nombre de contributeurs et de la présence possible d'artéfacts tels que des « pulls up », « stutters », « drop in ». Une trace est dite de bonne qualité d'après l'électrophorégramme obtenu, défini avec un seuil de détection limitant l'observation d'artéfacts, si elle satisfait les critères suivants :

- des pics élevés,
- de nombreux loci,
- la contribution d'une à trois personnes,
- la reproductibilité des profils entre différents réplicas,
- l'absence d'artéfact est dite de bonne qualité.

Concernant les traces de mélanges, plus les contributeurs ont des proportions différentes de contribution au mélange (qui se manifeste par des hauteurs de pics différentes), plus la trace est facile à analyser. Lorsque les proportions de contribution au mélange des contributeurs sont différentes, il peut être identifié une fraction dite majoritaire, partie du profil ADN constituée des pics les plus élevés, et une fraction dite minoritaire, partie du profil ADN constituée des pics les plus faibles. Les fractions majoritaires et minoritaires peuvent être constituées des profils ADN d'un ou plusieurs contributeurs, appelés respectivement majoritaires et minoritaires.

- **Des observations quantitatives**

La quantité d'ADN peut être déterminée d'après la hauteur des pics, une fois l'électrophoregramme obtenu. Cependant, celle-ci peut également être obtenue lors de l'étape de quantification de l'ADN dans la trace, étape qui intervient avant l'obtention du profil ADN. L'étape de quantification de l'ADN peut servir à déterminer un seuil en dessous duquel l'exploitation de la trace ADN n'aura plus aucune utilité car la possibilité d'obtenir un profil ADN exploitable sera jugée trop faible.

Certains experts (Daly, 2012) recommandent d'utiliser un seuil de quantification de 0.03 ng/mL, seuil en dessous duquel il ne faudrait pas faire d'analyse principalement parce que les probabilités de n'observer aucun profil ADN ou d'observer le résultat d'artefact sont élevées. Dans ce travail, les observations quantitatives sont celles obtenues lors de l'étape de la quantification d'ADN et non celles obtenues sur la base de la hauteur des pics et tous les profils ADN des traces extraites seront analysés quel que soit le résultat de la quantification. En effet, l'absence de profil ADN est une information que l'on souhaite exploiter (comme vu dans le Chapitre I) et que le logiciel STRmix™ (Moretti et al., 2017), qui prend en compte ses artefacts lors de l'évaluation, sera utilisé.

De manière générale, la quantité d'ADN ne peut pas être utilisée pour en déterminer si l'ADN provient d'un transfert direct ou indirect (Goray et al., 2010; Farmen et al., 2008). La perte totale d'ADN lors d'un transfert indirect, qui met en jeu une ou plusieurs surfaces intermédiaires entre le donneur et la surface d'intérêt, est supérieure à celle d'un transfert direct. Il est donc attendu qu'après un transfert indirect, la quantité récoltée soit plus faible que celle qui serait prélevée si le transfert avait été direct. Toutefois, en situation non contrôlée, la quantité initiale déposée sur la 1<sup>ère</sup> surface intermédiaire lors d'un premier transfert n'est pas connue. Il est possible que celle-ci soit très importante et qu'elle entraîne, selon les conditions de transferts suivants, une quantité finale déposée sur la surface prélevée plus élevée que celle qui proviendrait d'un transfert direct. La quantité récoltée peut donc être très faible sans pour autant signifier que la trace provienne obligatoirement d'un transfert indirect et inversement, la quantité récoltée peut être élevée sans pour autant signifier que la trace soit obligatoirement la conséquence d'un transfert direct.

De même, pour les traces de contact pauvres en ADN, il n'y a pas de relation directe entre la qualité du profil ADN obtenu post électrophorèse capillaire et la quantité d'ADN initialement déposée, contrairement aux traces riches en ADN (Goray et al., 2010). Cela signifie qu'une



trace de contact de très faible quantité d'ADN peut donner un profil de meilleure qualité qu'une autre trace de contact avec une quantité supérieure.

Cependant, si la probabilité d'obtenir une quantité d'ADN et une qualité du profil sachant que la proposition du procureur est vraie (allégation d'un transfert direct), est différente de la probabilité d'obtenir cette même quantité d'ADN et cette même qualité de profil si la proposition de la défense est vraie (allégation d'un transfert secondaire) alors cela signifie que le résultat ADN permet de soutenir préférentiellement une proposition plutôt qu'une autre. Ce rapport de probabilités est appelé rapport de vraisemblance (Aitken et Taroni, 2004).

Jones et al. (2016) ont publié une étude sur des cas d'agressions sexuelles lorsque de l'ADN de la victime est observé sur des prélèvements effectués sur le pénis ou le sous-vêtement du suspect. L'hypothèse du procureur est que « l'ADN prélevé provient d'un rapport sexuel entre la victime et le suspect » tandis que celle de la défense est que « l'ADN de la victime résulte d'un transfert indirect. L'ADN s'est d'abord déposé sur les mains du suspect suite à un contact non intime avec la victime puis a été transféré de ses mains au sous vêtement et/ou à son pénis ». Les auteurs ont reproduit par des expériences les deux scénarios, celui de la défense et celui du procureur. Ils ont observé que si le profil ADN de la partenaire est observé en majoritaire sur un prélèvement effectué sur le pénis, cette observation est plus probable si la proposition du procureur est vraie plutôt que si la proposition de la défense est vraie. Cette observation soutient donc la proposition du procureur par rapport à celle de la défense.

Il faudrait donc que les observations définies dans le BN tiennent compte de l'aspect quantitatif et qualitatif des résultats ADN.

Dans les études menées par Fonnelløp et al. (2017) et Kokshoorn et al. (2017), les observations dans le BN sont définies uniquement d'après la détection de l'ADN du POI. Dans ces études, utilisant chacune un BN différent, seul le fait d'observer ou non le profil ADN correspondant au POI est étudié. Le fait que dans une trace de mélange, le profil du POI soit en majoritaire ou en minoritaire n'est pas exploité. L'information quant à la proportion relative des différents contributeurs dans le cas d'un mélange est ainsi perdue.

A l'inverse, Szkuta et al. (2018) ont construit un BN qui prend en compte le fait que l'ADN peut provenir du POI mais également d'autres personnes telles que la victime ou des personnes inconnues, en fonction des propositions alléguées par le procureur et la défense.

---

Cependant, les quantités d'ADN ainsi que l'information quant à la proportion relative des différents contributeurs dans le cas d'un mélange ne sont pas utilisées.

**En conclusion**, la qualité d'un profil ADN dépend du nombre de contributeurs, de la proportion dans le mélange des différents contributeurs et de la hauteur des pics, donc de la quantité d'ADN.

Afin d'utiliser les informations relatives à la quantité totale d'ADN de la trace et celles concernant les proportions des différents contributeurs, il est proposé d'utiliser deux variables complémentaires comme observation :

- $E_1$  : « Résultat ADN POI », la quantité d'ADN correspondant au profil de la personne d'intérêt (POI). Cette variable est constituée, si la proposition du procureur est vraie, de la quantité d'ADN provenant du POI ainsi que du bruit de fond dont le profil ADN correspond de manière fortuite au profil ADN du POI et, si la proposition de la défense est vraie, de la quantité d'ADN provenant de l'agresseur (l'AO) ou du bruit de fond dont les profils ADN correspondent de manière fortuite au profil ADN du POI.
- $E_2$  : « Résultat ADN Non POI », la quantité d'ADN différent du profil de la personne d'intérêt (POI). Cette variable est constituée, si la proposition du procureur est vraie, de la quantité d'ADN provenant du bruit de fond dont le profil ADN est différent du profil ADN du POI et, si la proposition de la défense est vraie, de la quantité d'ADN provenant du bruit de fond et de l'agresseur (l'AO) dont les profils sont différents du profil ADN du POI.

De cette manière l'observation d'un profil ADN composé d'une fraction majeure correspondant au profil ADN du POI et d'une fraction mineure de mélange peut être prise en compte dans le BN.

L'ajout de ces variables, avec des états constitués d'intervalle de quantité d'ADN, constitue une plus-value par rapport aux autres BN proposés dans la littérature.

### II.2.3 Le raisonnement de construction

Une fois que les propositions, les observations, et toutes les variables ont été définies, les relations entre ces variables doivent être posées.

Pour cela, le raisonnement est de partir des propositions. Tout d'abord, pour chaque proposition, les différentes sources possibles d'ADN doivent être déterminées. Pour rappel, la possibilité de pollution n'est pas prise en compte.

Prenons l'exemple du cas de M. X. Le BN développé est présenté en Figure 1 sous forme de « Object Oriented Bayesian Network » ou OOBN (Taroni et al., 2014). Celui-ci est ici dans sa version française et s'appuie sur l'OOBN présenté par Taylor et al. (2017) (Annexe I.1). L'OOBN est un ensemble de plusieurs blocs, chaque bloc constituant en lui-même un BN.

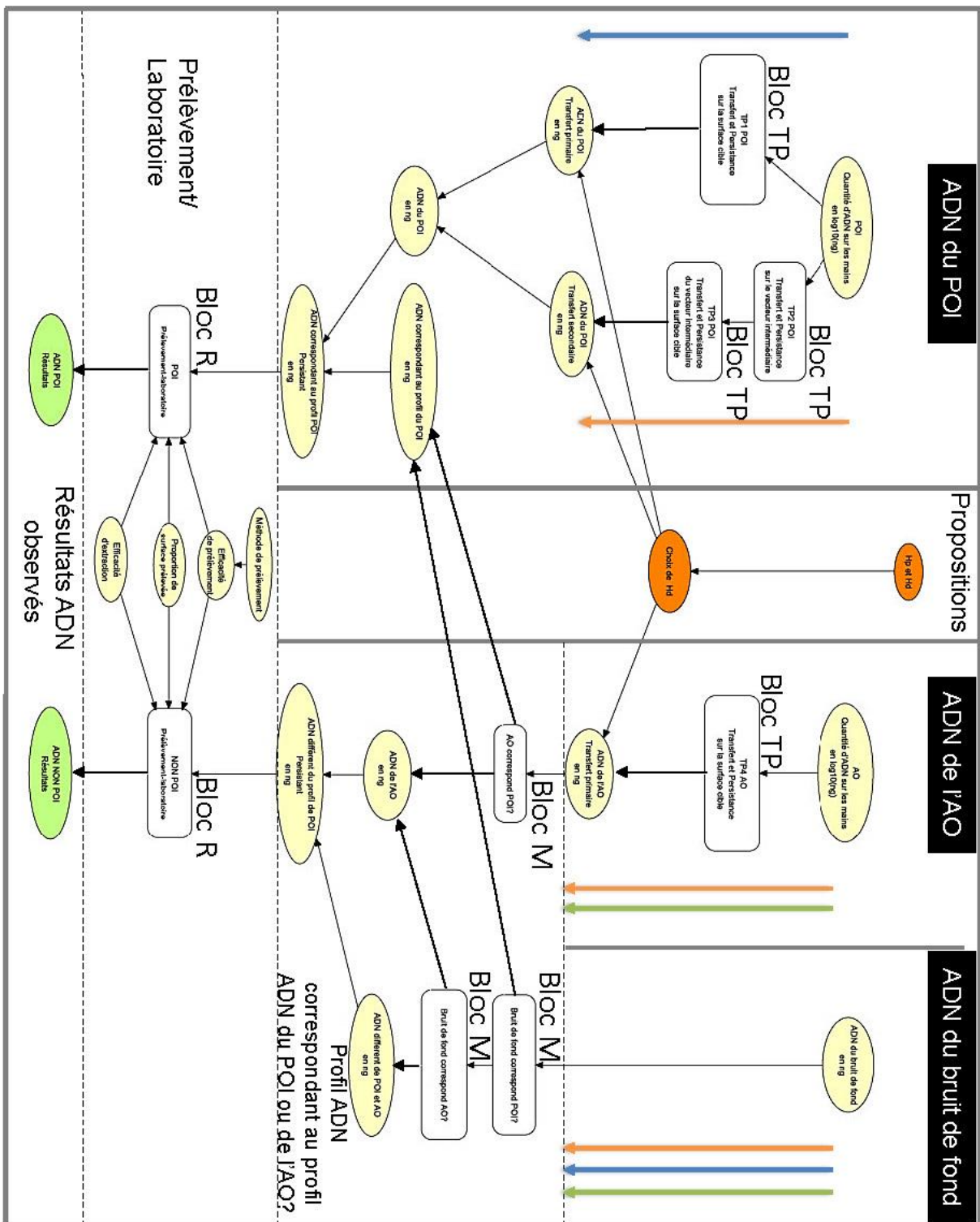


Figure 1 : OOBN utilisé pour évaluer les observations ADN considérant des propositions, H<sub>p</sub>, H<sub>d1</sub> et H<sub>d2</sub>, au niveau de l'activité.

Les différentes sources possibles d'ADN sont indiquées dans un encadré noir dans la Figure 1. Elles dépendent de la proposition alléguée.

Si la proposition du procureur  $H_p$  « M. A a poignardé la victime » est vraie, les quantités d'ADN obtenues peuvent provenir :

- **Du POI**, qui est la personne qui a utilisé l'objet d'intérêt au moment de l'action délictueuse. Il y a alors un phénomène de transfert d'ADN du POI directement sur le couteau via l'agression et un phénomène de persistance de l'ADN sur le manche du couteau entre l'action d'intérêt et la mise sous scellés soit un bloc « Transfert et Persistance » dit « TP ».
- **Du bruit de fond**, qui provient des personnes ayant pu laisser de manière directe ou non de l'ADN sur l'objet d'intérêt avant l'action. Le bruit de fond peut être différent ou correspondre fortuitement au profil ADN du POI soit un bloc « Coïncidence fortuite » dit « M ».

Ces voies sont indiquées par des flèches bleues sur la Figure 1.

Si la proposition de la défense  $H_{d1}$  : « M. A a serré la main de l'agresseur qui a poignardé la victime avec le couteau retrouvé sur les lieux » est vraie, les quantités d'ADN obtenues peuvent provenir :

- **Du POI**, qui est la personne qui a serré la main de l'agresseur. Dans ce cas, il y a un premier phénomène de transfert et de persistance de l'ADN du POI sur la main de l'agresseur via la poignée de main puis par un deuxième phénomène de transfert et de persistance de l'ADN du POI de la main de l'agresseur sur le couteau via l'agression soit deux blocs TP.
- **De l'AO**, qui est la personne qui a utilisé l'objet d'intérêt au moment de l'action délictueuse, par un phénomène de transfert et de persistance de l'ADN de l'agresseur sur le couteau via l'agression soit un bloc TP.
- **Du bruit de fond**, provenant des personnes ayant pu laisser de manière directe ou non de l'ADN sur l'objet d'intérêt avant l'action. Le bruit de fond peut être différent ou correspondre fortuitement au profil ADN du POI et/ou de l'AO soit un bloc M.

Ces voies sont indiquées par des flèches orange sur la Figure 1.

Si la proposition de la défense  $H_{d2}$  : « M. A n'a jamais touché le couteau retrouvé sur les lieux, ce n'est pas lui mais une autre personne inconnue, non apparentée à M. A, qui a poignardé la victime » est vraie, les quantités d'ADN obtenues peuvent provenir :

- **De l'AO**, par un phénomène de transfert et de persistance de l'ADN de l'agresseur sur le couteau suite à l'agression, soit un bloc TP.
- **Du bruit de fond**, provenant des personnes ayant pu laisser de manière directe ou non de l'ADN sur l'objet d'intérêt avant l'action. Le bruit de fond peut être différent ou correspondre fortuitement au profil ADN du POI et/ou de l'AO.

Ces voies sont indiquées par des flèches vertes sur la Figure 1.

**En conclusion**, les quantités d'ADN observées peuvent provenir :

- Du POI via un transfert direct (un bloc TP) si  $H_p$  est vraie ou via un transfert secondaire (deux blocs TP) si  $H_d$  est vraie.
- De l'AO via un transfert direct (un bloc TP) si  $H_{d1}$  est vraie.
- Du bruit de fond qui peut être différent ou correspondre fortuitement au profil de l'AO et/ou du POI (un bloc M) quel que soit l'hypothèse.

Chaque bloc TP comprend les variables sélectionnées de la partie précédente qui interviennent dans les phénomènes de transfert et de persistance de l'ADN considéré, ainsi que le paramètre  $\alpha$  de la loi de décroissance exponentielle qui permet de déterminer la perte d'ADN au cours du temps.

Une perte d'ADN peut ensuite avoir lieu lors du prélèvement et/ou de l'analyse en laboratoire de l'ADN. Les variables impliquées dans cette perte d'ADN liée à la collecte et au laboratoire sont prises en compte dans le bloc « Prélèvement et laboratoire » dit « R » pour « Recovery ». Il y a un bloc R pour l'ensemble de l'ADN provenant du POI et un bloc R pour l'ensemble de l'ADN provenant d'une personne autre que le POI appelé « Non POI » indiqué au niveau du Prélèvement/Laboratoire sur la Figure 1.

Les Figure 2, Figure 3 et Figure 4 illustrent les relations entre les variables pour les blocs TP, R et M.

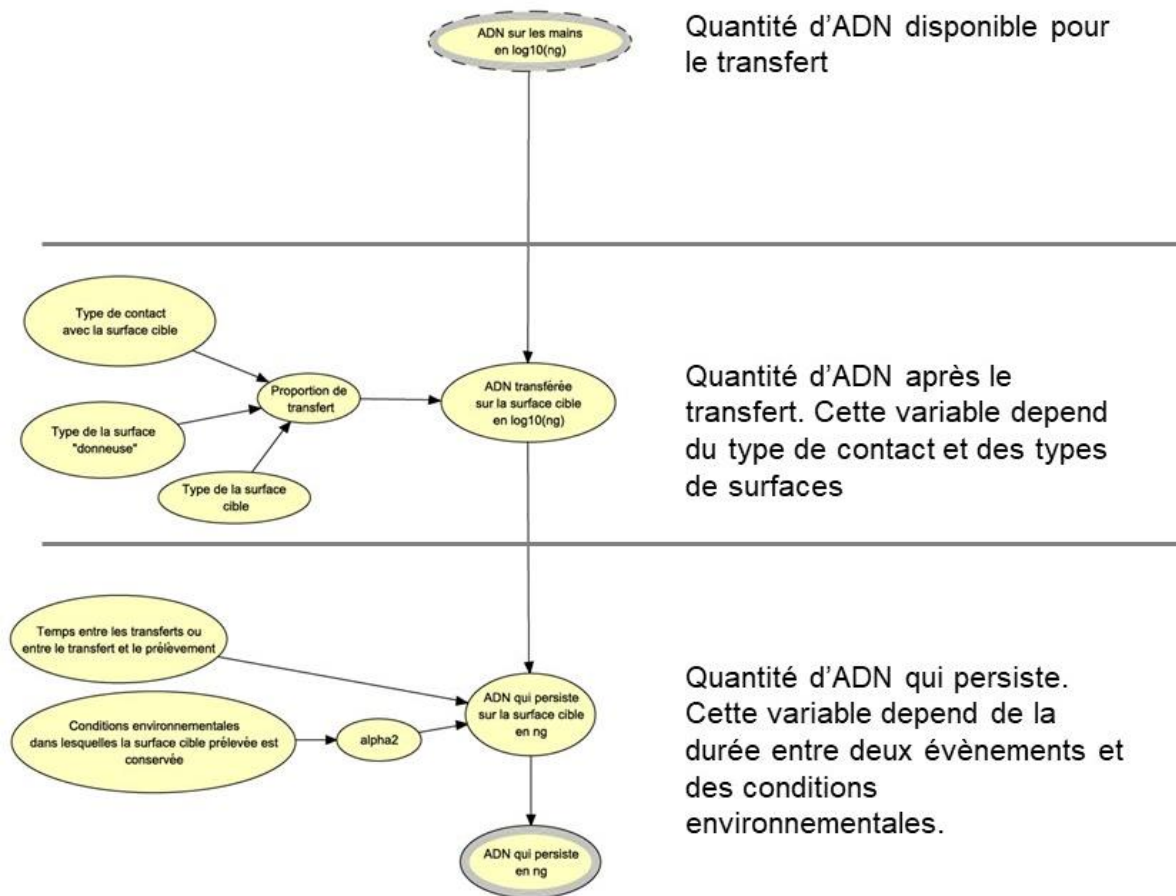


Figure 2: Réseau Bayésien du bloc TP.

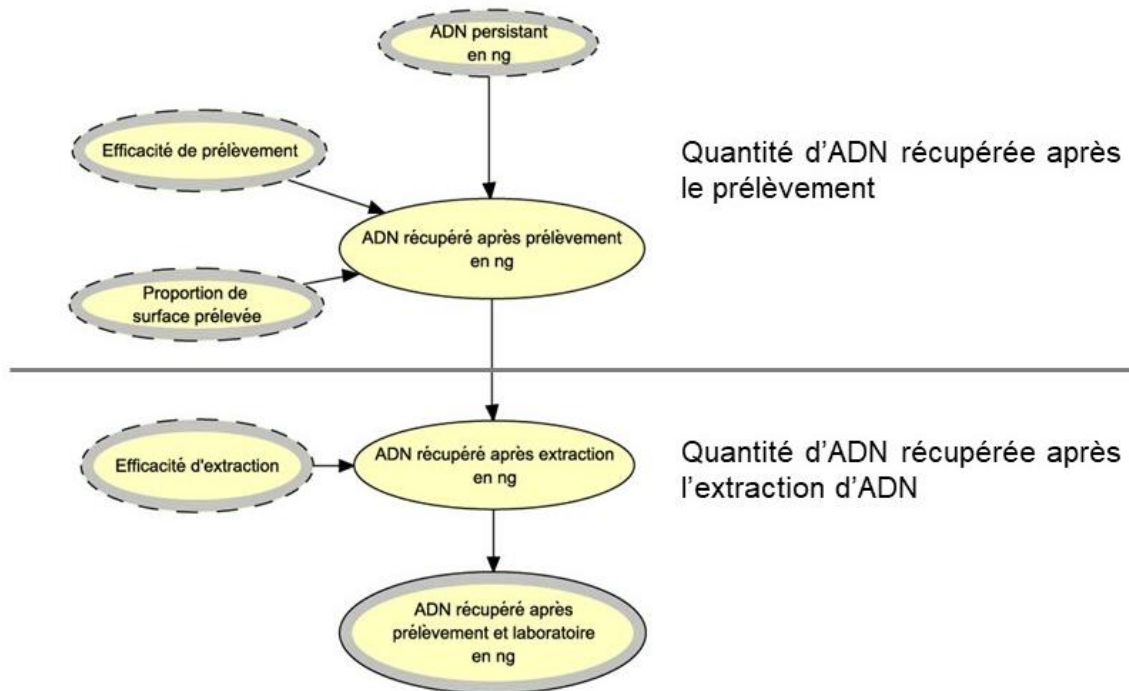


Figure 3: Réseau Bayésien du bloc R.

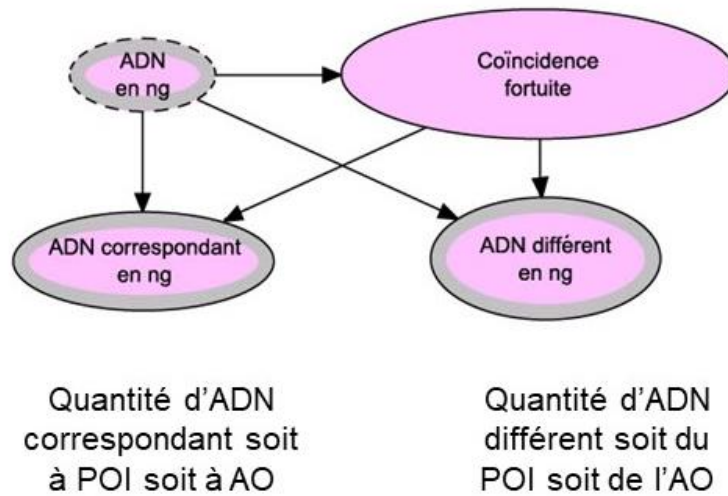


Figure 4: Réseau Bayésien du bloc M.

La probabilité de coïncidence fortuite est utilisée afin de rendre compte du fait que le bruit de fond peut correspondre fortuitement au profil ADN du POI et/ou de l'AO.



II.2.4 Etats des variables

Les différents états des nœuds (variables) sont décrits dans le Tableau 1 ci-dessous :

**Tableau 1 : Présentation des variables ainsi que leurs états associés.**

<b>Variable</b>	<b>Etat</b>
Hp et Hd	H <sub>p</sub> H <sub>d</sub>
Choix d'Hd	H <sub>d1</sub> H <sub>d2</sub>
Résultats DNA POI (en ng) E <sub>1</sub>	Intervalle de quantité; 0 à 0.1 par pallier de 0.01 0.1 à 1 par pallier de 0.1 1 à 5 par pallier de 0.5 5 à 10 par pallier de 1 10 à 25 par pallier de 5 25 à 1000 1000 à inf.
Résultats DNA Non POI (en ng) E <sub>2</sub>	Intervalle de quantité; 0 à 0.1 par pallier de 0.01 0.1 à 1 par pallier de 0.1 1 à 5 par pallier de 0.5 5 à 10 par pallier de 1 10 à 25 par pallier de 5 25 à 1000 1000 à inf.
Type de surface	Main Lisse Rugueux
Type de contact	Passif Pression Friction
Conditions environnementales	Favorable Défavorable
Méthode de prélèvement	Tapelift Ecouvillon
Temps (en jour)	intervalle de quantité; 0, 0.5 puis 1 à 31 par pallier de 1
Alpha2	-0.052 -0.022
Proportion de la surface prélevée (en %)	intervalle de quantité; valeurs de 0 à 1 par pallier de 0.1
Proportion de transfert (en %)	intervalle de quantité; de 0 à 1 par pallier de 0.05
Efficacité du prélèvement (en %)	intervalle de quantité; de 0 à 1 par pallier de 0.05

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

Efficacité d'extraction (en %)	intervalle de quantité; de 0 à 1 par pallier de 0.05
Quantité d'ADN sur les mains en log10 (ng)	Intervalle de log10 (quantité); -inf. à -1.5 -1.5 à 3 par pallier de 0.1 3 à 3.5 3.5 à inf.
Bruit de fond (en ng)	Intervalle de quantité; 0 à 0.1 par pallier de 0.01 0.1 à 1 par pallier de 0.1 1 à 5 par pallier de 0.5 5 à 10 par pallier de 1 10 à 25 par pallier de 5 25 à 1000 1000 à inf.
Coïncidence fortuite	Oui (coïncidence fortuite) Non (différence)

La manière dont les états sont définis concernant les quantités d'ADN, permet, dans une certaine mesure, de prendre en compte l'incertitude du résultat de la quantification d'ADN. Par exemple si 0.02 ng/μL d'ADN est observé soit 1ng, il est possible qu'il y ait en réalité 0.04 ng/μL d'ADN soit 2 ng. Qu'une quantité de 1 ou 2 ng d'ADN soit obtenue n'a pas d'impact sur l'observation qui sera la suivante dans le BN: entre 1 et 5 ng d'ADN.

#### II.2.5 Données

Une fois le réseau construit, les tables de probabilités conditionnelles (CPT) de chaque variable sont à compléter.

Les variables « Type de surface », « Type de contact », « Conditions environnementales » et « Méthode de prélèvement » ont des états instanciés qui sont déterminés en fonction des circonstances du cas. Il s'agit donc s'informations connues au moment de l'analyse du cas.

Les données permettant de remplir les CPT des variables dont les états ne sont pas instanciés, « Bruit de fond », « Quantité d'ADN sur les mains », « Proportion de transfert », « Efficacité des méthodes de prélèvements », « Efficacité d'extraction ADN », « Temps », « Proportion de surface prélevée » et « Coïncidence fortuite », sont fournies par les circonstances du cas, les conditions de laboratoire, les connaissances et les études publiées.

Les données utilisées sont celles qui ont été publiées et qui sont détaillées dans l'article de Samie et al. (2019<sub>(b)</sub>) (Annexe II.1). Elles résultent d'expériences faites par différents laboratoires. Les conditions sont contrôlées et connues : l'instant du lavage des mains avant

de toucher une surface, l'intervalle de temps entre le serrage de mains et le transfert sur la surface d'intérêt, le statut des donneurs ainsi que le type des surfaces intermédiaires. Ces expériences contrôlées sont utiles afin de comprendre les processus de transfert, persistance et de bruit de fond et surtout l'impact des variables qui interviennent dans ceux-ci. Cependant, dans une situation réelle, tous les paramètres ne sont pas connus ou peuvent être différents de ceux considérés dans les expériences. Les résultats en termes de quantité d'ADN prélevé et/ou de qualité peuvent ainsi différer de ceux donnés par la littérature et obtenus en conditions contrôlées.

L'OBN utilisé dans cette thèse diffère légèrement de celui développé par Taylor et al. (2017). Contrairement à l'OBN présenté dans l'article de Taylor et al. (2017), le log10 de la quantité d'ADN, et non la quantité d'ADN, est spécifié quantitativement pour la variable « Quantité d'ADN sur les mains » pour deux raisons :

- les données de cette variable en log 10 suivent une distribution dites de type Normale; la modélisation en est alors facilitée,
- l'utilisation de la distribution Normale permet de déterminer plus facilement les paramètres de la distribution que suivent les données de la variable « Proportion de transfert ». Afin de calculer les paramètres de la distribution que suivent les données de cette variable, les données obtenues pour la « Quantité d'ADN sur les mains » sont nécessaire et l'utilisation d'une distribution de type Normale facilite ce calcul (Cf. partie IV.1.2).

De plus, la variable « Proportion de contact » présente dans l'OBN de Taylor et al. (2017) n'est pas utilisée. En effet, les données attachées à cette variable sont déjà prises en compte dans la variable « Proportion de transfert ».

## II.3 Discussion

### II.3.1 Les limites du réseau bayésien

Le réseau bayésien développé présente certaines limites. La quantité 0 ng ne peut pas être retenue comme un état, car les états sont définis comme des intervalles de quantités. Si la quantité 0 ng était disponible, l'observation « Profil ADN d'une seule personne correspondant au profil du POI » devrait être représentée par :

- $E_1$  : quantité d'ADN observée
- $E_2$  : 0

Ceci n'est pas possible dans la présente construction. Le plus proche est :

- $E_1$  : quantité d'ADN observée
- $E_2$  : 0-0.01 (ng)

Dans l'étude de Samie et al. (2016), sur 64 profils analysés, lorsqu'une quantité plus petite que 0.01 ng a été obtenue, aucun allèle ou 1 allèle maximum n'a été observé. Ceci signifie qu'une absence de profil est obtenue pour des quantités comprises entre 0 et 0.01 ng. L'observation « Profil ADN d'une seule personne correspondant au profil du POI » correspond donc également à cette représentation du résultat même si la quantité diffère de 0.

Ce BN simplifie également certains phénomènes. Lors du transfert par exemple, l'ADN se transfère de la surface donneuse vers la surface cible mais il y a en réalité, également un transfert d'ADN de la surface cible vers la surface donneuse. Ce transfert n'est pas représenté dans ce BN. Cette problématique a fait l'objet de l'étude de Taylor et al. (2019).

### II.3.2 Les avantages du réseau bayésien

En revanche, le BN décrit ci-dessus présente de multiples avantages :

- Il permet de gérer les observations faites par rapport au profil ADN du POI mais également celles faites par rapport aux profils ADN différents du profil ADN du POI.
- Les observations permettent, par rapport à ce qui a été publié jusqu'à maintenant, de conserver l'information concernant les quantités relatives d'ADN entre les différents contributeurs. La notion de fraction « majoritaire » et fraction minoritaire » est conservée.
- Le raisonnement sur lequel repose cette construction est généralisable à de multiples scénarios impliquant différentes activités comme le montre Taylor et al. (2017) qui utilise un BN similaire.
- Pour certaines variables comme la « Quantité d'ADN sur les mains » ou « l'Efficacité d'extraction », l'information concernant la distribution que suivent les données est préservée, ce qui présente un avantage par rapport à des variables instanciées comme les « Conditions environnementales ». Ceci permet, pour les variables informant sur la quantité d'ADN sur les mains et la proportion de transfert, de prendre en compte la variabilité intra individuelle et donc la probabilité que l'individu donne ou transfère une quantité définie d'ADN.

Concernant les variables informant sur l'efficacité d'extraction et l'efficacité du prélèvement ADN, l'information concernant la distribution que suivent ces données

---

permet de prendre en compte la variation entre plusieurs kits d'extractions ou méthode de prélèvement.

Pour les variables concernant le temps écoulé et la proportion de la surface prélevée, ne pas instancier un état permet de gérer l'incertitude qu'il peut y avoir autour de ces variables et ainsi de refléter des circonstances particulières comme : l'objet a été prélevé environ deux jours après l'action d'intérêt, plus ou moins une demi-journée.

Dans les réseaux bayésiens proposés jusqu'à présent, la perte possible d'ADN liée aux techniques de détection des traces digitales et donc l'influence de ces techniques sur la quantité et la qualité des profils ADN n'est pas prise en compte. Kumar et al. (2015) montrent pourtant que les méthodes utilisées pour détecter des traces digitales influent sur le résultat ADN en termes de qualité de profil et de quantité d'ADN. Cette influence dépend aussi du temps écoulé entre l'application de la méthode et l'extraction d'ADN. Partant de cet article, si des techniques n'ayant pas d'influence sur les observations ADN sont appliquées avant le prélèvement ADN, le BN présenté dans cette thèse peut être utilisé. En revanche, si une technique utilisée a une influence sur les observations ADN, une variable supplémentaire prenant en compte la dégradation d'ADN lors, par exemple, de l'application de la méthode de détection de traces digitales doit être ajoutée au BN. L'effet de cette variable sur le résultat de l'évaluation devra être déterminé.

**En conclusion**, le réseau présenté dans cette partie va permettre de gérer nos observations en tenant compte de la quantité d'ADN liée au POI, des quantités d'ADN issues d'autres individus ainsi que de leurs proportions relatives. Ce BN s'appuie principalement sur des données issues de la littérature. La paramétrisation est présentée en détail dans Samie et al. (2019<sub>(b)</sub>) dans l'Annexe II.1.

Une fois le BN construit, l'objectif est de déterminer les variables pour lesquelles l'acquisition de données supplémentaires est nécessaire.

## Chapitre III - Identification des variables d'intérêt

L'objectif de ce chapitre est d'étudier l'impact des variables et de l'acquisition de nouvelles données sur le résultat de l'évaluation pour le cas de M. X.

### III.1 Introduction

Pour rappel, le cas de M. X est le suivant : les secours sont appelés pour intervenir dans un immeuble suite à un appel de voisins. Les secouristes trouvent le corps de M. X poignardé de deux coups de couteau. Un couteau à manche lisse, dont la lame est ensanglantée, est retrouvé près de la victime. Celui-ci est conservé dans un sachet papier. Le couteau est ensuite transmis au laboratoire qui effectue un prélèvement d'ADN sur le manche au moyen d'un écouvillon afin de l'analyser et d'obtenir, si possible, un profil ADN. La mort se situe environ deux jours avant la découverte du corps. Durant l'enquête, M. A est mis en cause. Il s'agit donc de la personne d'intérêt appelé POI. D'après les proches de la victime, le couteau appartient à M. X. Aucun élément d'enquête ne permet de mettre en cause une personne apparentée au suspect.

L'état des variables « Type de surface », « Type de contact », « Environnement » et « Méthode de prélèvement » est connue, donc, dans le BN, ces états sont instanciés (activés) selon les circonstances du cas décrites ci-dessus et précisées dans le Tableau 2.

**Tableau 2 : Présentation des variables ainsi que de l'état instancié qui leur est associé accompagné d'une explication sur cet état instancié.**

Variable	Etat instancié	Explication
Type de surface cible	Lisse	Le couteau est la surface sur laquelle l'ADN peut se transférer
	Main	La main est la surface cible lors de la poignée de main sous $H_{d2}$
Type de surface donneuse	Main	La main est la source ADN d'origine
Type de contact	Friction	Sous $H_p$ , et $H_{d1}$ , l'action est des coups de couteau
	Pression	Sous $H_{d2}$ , l'action est une poignée de main
Environnement	Favorable	Le couteau se trouve chez M. X, protégé de l'extérieur
	Défavorable	Sous $H_{d2}$ , lorsque l'ADN du POI est transféré sur la main de l'AO, cette main est considérée comme un milieu défavorable car le risque de contact, et donc de perte d'ADN, entre la poignée de main et l'agression est grand
Méthode de prélèvement	Ecouvillon	L'écouvillon est la méthode de prélèvement ADN utilisé par le laboratoire

Une trace de mélange composé d'une fraction majeure correspondant au profil du POI et d'une fraction mineure pas interprétable (moins de 8 allèles composant le profil ADN de la fraction mineure différenciable du profil ADN de la fraction majeure) a été obtenue à partir d'un prélèvement effectué sur le manche du couteau. Une quantité totale d'ADN de 0.12 ng a été récoltée. L'analyse du profil ADN de la trace avec le logiciel STRmix™ v2.5 montre que 87.5 % de la quantité totale d'ADN, soit 0,105 ng, correspond au profil du POI et 22.5%, soit 0.015 ng, est différent. Le logiciel STRmix™ est basé sur un modèle continu (Moretti et al., 2017). Il permet d'analyser un profil ADN en le décomposant en différents contributeurs potentiels. La proportion dans le mélange des différents contributeurs est également analysée.

Les observations sont alors :

$E_1$  : 0.1-0.2

$E_2$  : 0.01-0.02

Nous rappelons que dans le premier scénario, les propositions alléguées sont :

- Hypothèse  $H_p$  : « Le POI a poignardé la victime »
- Hypothèse  $H_{d1}$  : « Le POI a serré la main de l'AO qui a poignardé la victime avec le couteau retrouvé sur les lieux ».

Dans le deuxième scénario, les propositions alléguées sont :

- Hypothèse  $H_p$  : « Le POI a poignardé la victime »
- Hypothèse  $H_{d2}$  : « Le POI n'a jamais touché le couteau retrouvé sur les lieux, ce n'est pas lui mais une autre personne inconnue appelée AO, qui a poignardé la victime ».

Les observations génétiques (Profil ADN détecté ou non, correspondant ou non au profil ADN du POI) sont évaluées au regard de ces propositions.

Le résultat de l'évaluation est présenté sous forme d'un rapport de vraisemblance (Likelihood Ratio, LR). Le rapport de vraisemblance constitue une méthode logique pour évaluer les observations et la valeur de l'apport de nouvelles informations (Aiken et Taroni, 2004 ; Jackson et al., 2006 ; Taroni et al., 2014).

Le LR est défini comme suit :

$$LR = \frac{P(E_i|H_p,I)}{P(E_i|H_d,I)}$$

---

Où  $P$  représente les probabilités,  $H$  les propositions au niveau de l'activité respectivement du procureur ( $H_p$ ) et de la défense ( $H_d$ ),  $E_i$  les observations et  $I$  les informations circonstanciées. Les observations ADN sont donc dépendantes des propositions posées par les différentes parties.

A noter ici que nous avons deux observations :  $E_i = \{E_1, E_2\}$  où  $E_1$  est défini par la quantité d'ADN correspondant au profil du POI (« Résultat ADN POI ») et  $E_2$  la quantité d'ADN différent du profil ADN du POI (« Résultat ADN Non POI »).

Le résultat de l'évaluation peut également être donné sous forme d'un équivalent verbal, selon une échelle verbale. Un exemple d'échelle verbale est donné par Willis et al. (2015) à la page 17.

Le numérateur de l'équation précédente,  $P(E_i|H_p, I)$ , représente la probabilité de faire les observations  $E_i = \{E_1, E_2\}$  si l'hypothèse du procureur est vraie et le dénominateur,  $P(E_i|H_d, I)$ , la probabilité de faire ces mêmes observations si l'hypothèse de la défense est vraie.

Les valeurs de LR sont comprises entre 0 et  $\infty$ . La valeur de 1 est dite « neutre » car elle souligne que  $P(E_i|H_p, I) = P(E_i|H_d, I)$ . Les observations  $E_i = \{E_1, E_2\}$  soutiennent autant l'hypothèse du procureur que celle de la défense. Toutes valeurs inférieures à 1 signifient que les observations  $E_i = \{E_1, E_2\}$  soutiennent plus l'hypothèse de la défense que celle du procureur. A l'inverse, un LR supérieur à 1 signifie que les observations  $E_i = \{E_1, E_2\}$  soutiennent plus l'hypothèse du procureur que celle de la défense.

L'objectif de cette partie est d'étudier l'impact des variables intégrées au BN sur le résultat de l'évaluation donc sur le LR afin de sélectionner les variables d'intérêt. Les variables d'intérêt sont celles dont l'impact sur le LR est important et pour lesquelles l'acquisition de données supplémentaires est nécessaire à l'évaluation afin de diminuer la variation des valeurs des LR. L'impact sur le LR est dit important si l'étendue des LR, qui résultent de la modification des données (décrite ci-après) de cette variable, est supérieure à un ordre de 1. Un ordre de 1 correspond à l'écart de LR pour lequel il y a un changement d'équivalent verbal selon l'échelle verbale.

Deux étapes composent cette partie :

- Identification des potentielles variables d'intérêt, *i.e.* celles dont une modification des données associées induit un impact important (une variation significative) sur le LR.



- Identification des variables d'intérêt, *i.e.* celles dont l'ajout de données supplémentaires permet de diminuer la variation des LR.

### III.2 Identification des potentielles variables d'intérêt.

Tout d'abord l'impact des variables sur le LR est étudié. Pour cela, une analyse de sensibilité du BN est effectuée en utilisant la plateforme Rstudio (RStudio Team, 2016; R Core Team, 2018) ainsi que le package RHugin (Kjell Konis, 2017). Des analyses de sensibilité utilisant des méthodes similaires ont été réalisées par Pun (2016) et Palmer (2016).

L'analyse de sensibilité consiste à mener plusieurs simulations en faisant varier, à chaque simulation, les données d'une ou plusieurs variables informant le BN, les autres étant inchangées. L'analyse de sensibilité a pour objectif de mettre en évidence l'impact de l'échantillonnage, lié à la quantité de données, sur le LR. Elle ne permet pas de se prononcer sur la pertinence des données, c'est-à-dire l'adéquation des données utilisées par rapport au cas. La pertinence des données est étudiée au chapitre IV. Dans cette partie, nous disposons de données dites « initiales », en partant du présupposé que celles-ci sont adéquates, pour informer le BN. Les données varient, à chaque simulation, seulement dans une certaine fourchette de valeurs. En effet, dans ce travail, nous faisons varier les données par ré-échantillonnage autour d'une valeur moyenne définie par la paramétrisation initiale. En revanche, si nous n'avions eu aucune information sur la variable, nous aurions fait varier les données entre des valeurs extrêmes et non dans une fourchette de valeurs. Par exemple, nous aurions fait varier l'efficacité du prélèvement entre 0 et 100%. La manière dont les données sont ré-échantillonnées à chaque simulation pour chaque variable est décrit dans la publication de Samie et al. (2019<sub>(b)</sub>) (Annexe II.1). A chaque simulation, un LR est obtenu pour chaque observation fixe  $E_i = \{E_1, E_2\}$ . Au final, la variation que ce changement dans les données de/des variables a induit sur le rapport de vraisemblance est analysée.

Cent simulations ont été effectuées pour chacune des variables. Cent autres simulations ont aussi été réalisées en ré-échantillonnant simultanément les données des variables non instanciées, excepté les variables « Temps » et « Proportion de surface prélevée » dont les données sont fixes comme illustré dans le Tableau 3 .

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

**Tableau 3 : Présentation des variables dont les données sont ré-échantillonnées à chaque simulation ainsi que le nombre de simulation et le nom donnée à cet ensemble de simulation.**

Nom des simulations utilisé pour observer les résultats	Variable(s) non instanciée(s) dont les données sont ré-échantillonnées	Nombre de simulation
Etat Initial	Aucune	1
Toutes les variables	« Coïncidence fortuite », « Quantité d'ADN sur les mains » du POI et de l'AO, « Proportion de transfert » du POI et de l'AO, « Efficacité du prélèvement », « Bruit de fond », « Efficacité d'extraction »	100
Proportion de surface prélevée	Proportion de surface prélevée	100
Coïncidence fortuite	« Coïncidence fortuite »	100
Quantité d'ADN sur les mains	« Quantité d'ADN sur les mains » du POI et de l'AO	100
Temps	« Temps »	100
Proportion de transfert	« Proportion de transfert » du POI et de l'AO	100
Efficacité du prélèvement	« Efficacité du prélèvement »	100
Bruit de fond	« Bruit de fond »	100
Efficacité d'extraction	« Efficacité d'extraction »	100

En effet, le temps écoulé entre les différentes actions et le prélèvement est connu, de même que la proportion de surface prélevée. Cependant ces données peuvent ne pas être aussi précises par rapport aux états proposés pour les états des variables « Temps » et « Proportion de surface prélevée ». Par exemple, nous savons que l'entièreté du manche du couteau a été prélevée et nous pensons que cela représente environ 90% de la surface de la main en contact avec le manche du couteau. Dans cet exemple, doit-on fixer l'état « 0.8-0.9 » ou « 0.9-1 » ? Afin de rendre compte de l'impact sur le LR de l'incertitude sur ces données, une analyse de sensibilité a été réalisée pour les variables « Temps » et « Proportion de surface prélevée ». En revanche, leurs données ne sont pas modifiées lorsque l'ensemble des données des autres variables non instanciées sont modifiées simultanément.

Les simulations ont été réalisées selon les circonstances du cas de M. X mais également en prenant des circonstances drastiquement différentes : un cas dit « transfert élevé » dont les

circonstances favorisent le transfert d'ADN et un cas dit « transfert faible » dont les circonstances sont défavorables au transfert d'ADN (Tableau 4) afin d'évaluer l'importance des conditions circonstancielle dans la sélection des variables ayant le plus d'impact avec cette paramétrisation.

**Tableau 4 : Présentation des variables ainsi que leur état instancié associé pour chacun des cas « Cas de M. X », « Cas Transfert élevé », « Cas Transfert faible ».**

Variable	Cas de M. X	Cas Transfert élevé	Cas Transfert faible
Type de surface cible	TP1-TP3-TP4: Lisse TP2: Main	TP1-TP3-TP4: Rugueux TP2: Main	TP1-TP3-TP4: Rugueux TP2: Main
Type de surface donneuse	Main	Main	Main
Type de contact	TP1-TP3-TP4: Friction TP2: Pression	TP1-TP3-TP4: Friction TP2: Pression	TP1-TP3-TP4: Friction TP2: Pression
Temps (en jours)	TP1-TP3-TP4: 2 TP2: 0.5	TP1-TP3-TP4: 0.5 TP2: 0.5	TP1-TP3-TP4: 20 TP2: 0.5
Conditions environnementales	TP1-TP3-TP4: Favorable TP2: Défavorable	TP1-TP3-TP4: Favorable TP2: Défavorable	TP1-TP3-TP4: Unfavorable TP2: Défavorable
Méthode de prélèvement	Ecouvillon	Ecouvillon	Ecouvillon

Afin d'explorer quelles variables ont un impact sur le LR, l'étendue de la variation des  $\log_{10}(\text{LR})$ , qui est la différence entre le  $\log_{10}(\text{LR})$  maximale et le  $\log_{10}(\text{LR})$  minimal, obtenue pour chaque simulation et pour chaque observation, est analysée. L'étendue est un critère choisi car il permet de rendre compte du plus grand écart pouvant être obtenu suite à ces 100 simulations. Cependant, ce paramètre est dépendant du nombre de simulations. Suite à ces résultats, les potentielles variables d'intérêt sont retenues. Les variables considérées comme variables d'intérêt potentielles sont celles dont l'étendue est supérieure à un ordre de 1 en  $\log_{10}$ . Un ordre de 1 en  $\log_{10}$  correspond à l'écart de LR pour lequel il y a un changement d'équivalent verbal selon l'échelle verbale. A noter que chaque LR résulte d'une simulation, donc chaque LR est associé à un jeu de données utilisé. Il n'est pas question ici d'intervalle de LR, vu qu'il n'y a pas plusieurs LR obtenus pour un même jeu de données mais plusieurs LR obtenus pour plusieurs sets de données.

Dans un premier temps, les résultats généraux, c'est-à-dire les potentielles variables d'intérêt qui sont identifiées pour la plupart des observations  $E_i = \{E_1, E_2\}$  sont présentées. Dans un

second temps, le cas de M. X est pris comme exemple. Enfin, l'effet de l'augmentation de la quantité d'ADN du POI et du Non POI sur le LR et est étudié.

### III.2.1 Résultat général

Pour chacun des scénarios, la variation des  $\log_{10}(\text{LR})$  pour chaque observation  $E_i = \{E_1, E_2\}$  et chaque variable étudiée peut être visualisée en sélectionnant « Part 1 : Sensitivity analysis » sur le site : [https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/). Ce site présente de manière visuelle l'ensemble des résultats des simulations effectuées dans ce travail. Une synthèse de ces résultats est proposée ci-dessous.

#### 1. Scénario 1

Pour rappel, les propositions d'intérêt pour le scénario 1 sont les suivantes :

- $H_p$  : Le POI a agressé la victime
- $H_{d1}$  : Le POI a serré la main de l'AO qui a poignardé la victime avec le couteau retrouvé sur les lieux (Un potentiel transfert secondaire est donc allégué)

La Figure 5 illustre pour chacune des variables, la variation des étendues des  $\log_{10}(\text{LRs})$  obtenues pour l'ensemble des observations  $E_i = \{E_1, E_2\}$  suite aux simulations, considérant le scénario 1. Elle permet d'identifier les potentielles variables d'intérêt pour la plupart des observations.

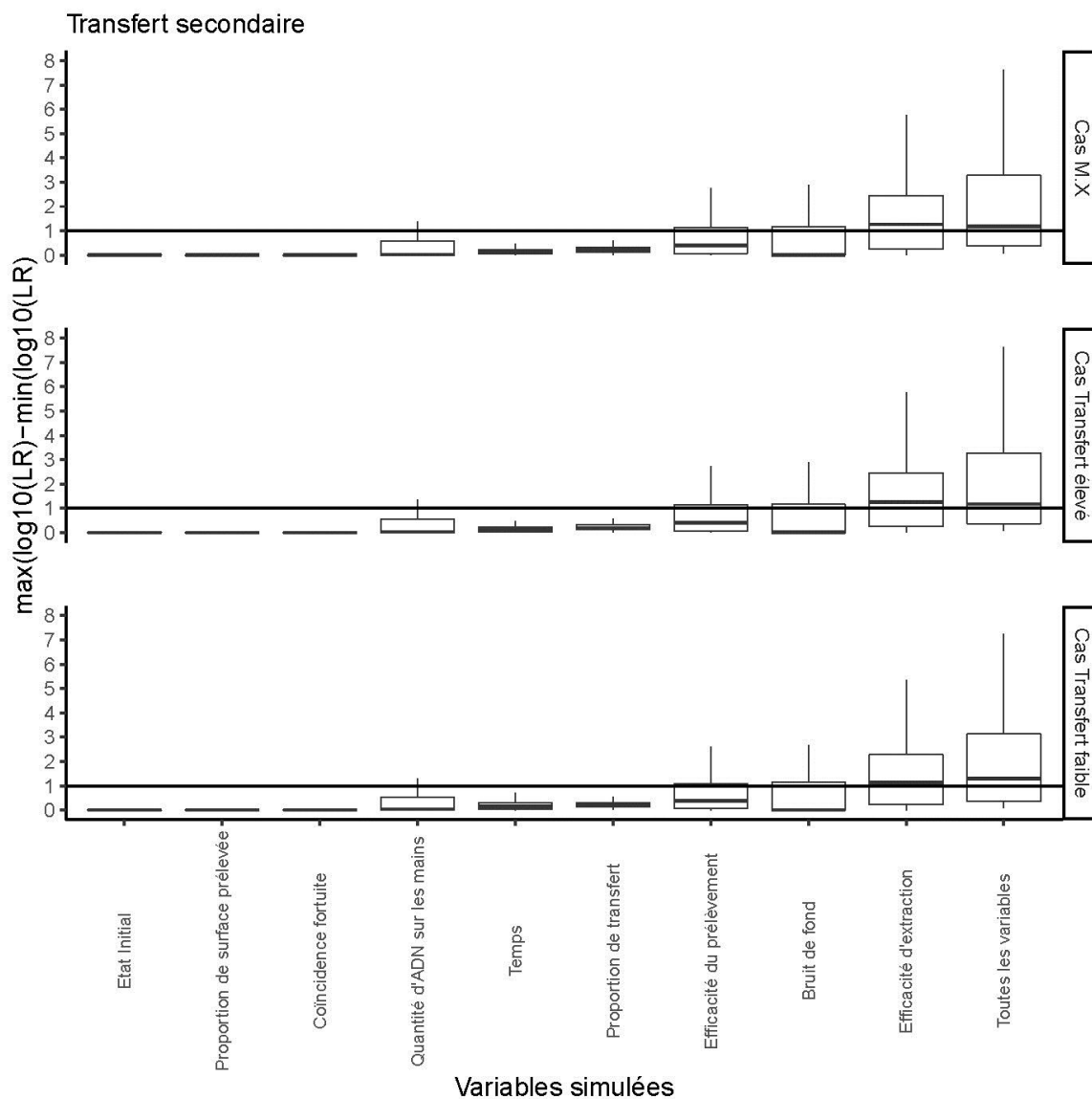


Figure 5 : Variation pour chaque variable des étendues des log<sub>10</sub>(LRs) obtenues pour chaque observation dans le cas du scénario 1 pour lequel un transfert secondaire est allégué par la défense.

L'analyse de sensibilité montre, pour la plupart des observations, qu'un ré-échantillonnage à partir des données de la paramétrisation initiale des variables « Quantités d'ADN sur les mains », « Efficacité du prélèvement », « Bruit de fond » et « Efficacité d'extraction » contribue à une variation des log<sub>10</sub>(LRs) significative (variation des log<sub>10</sub>(LRs) de plus d'un ordre de 1).

Les variables « Proportion de surface prélevée », « Coïncidence fortuite », « Temps » et « Proportion de transfert » n'ont pas d'impact significatif (au sens de la définition adoptée) sur le LR.

L'impact sur le LR est similaire pour les 3 cas, cas de M. X, cas Transfert élevé et cas Transfert faible.

## 2. Scénario 2

Pour rappel, les propositions d'intérêt pour le scénario 2 sont les suivantes :

- $H_p$  : POI a agressé la victime
- $H_{d2}$  : Le POI n'a jamais touché le couteau retrouvé sur les lieux, ce n'est pas lui mais une autre personne inconnue, appelée AO, qui a poignardé la victime

La Figure 6 illustre pour chacune des variables, la variation des étendues des  $\log_{10}(\text{LRs})$  obtenues pour chaque observation suite aux simulations, considérant le scénario 2.

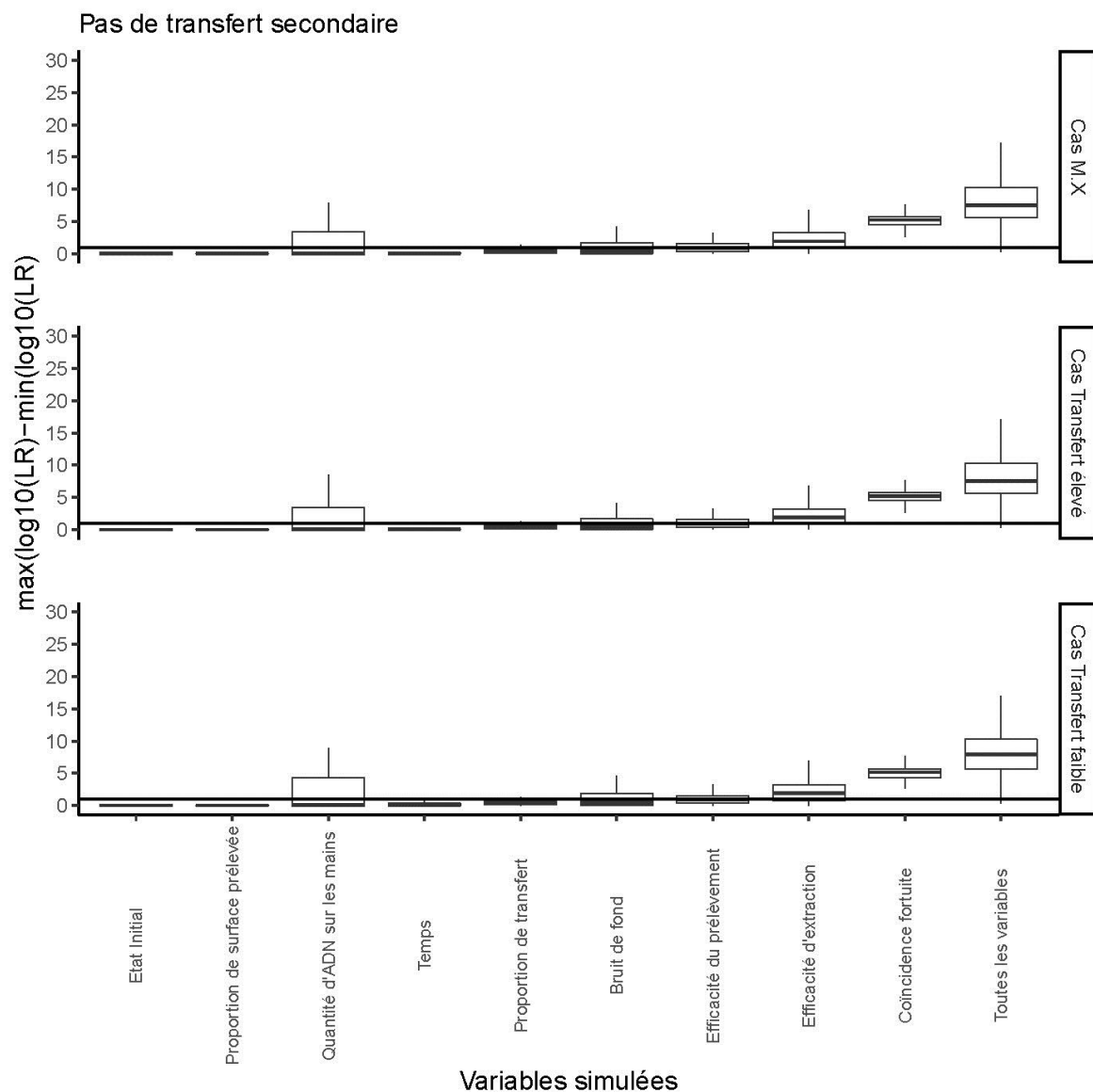


Figure 6 : Variation pour chaque variable des étendues des  $\log_{10}(\text{LRs})$  obtenues pour chaque observation dans le cas du scénario 2 pour lequel aucun transfert secondaire n'est allégué par la défense.

Dans le cas du scénario 2, l'analyse de sensibilité montre, pour la plupart des observations, qu'un ré-échantillonnage à partir des données de la paramétrisation initiale des variables « Quantités d'ADN sur les mains », « Bruit de fond », « Efficacité de prélèvement », « Efficacité d'extraction » et « Coïncidence fortuite » a un impact sur le LR. Nous observons également que la variable « Coïncidence fortuite » contribue le plus à la variation des  $\log_{10}(\text{LRs})$ .

Les variables « Proportion de surface prélevée » et « Temps » n'ont pas d'impact significatif sur le LR.

L'impact sur le LR des modifications des données des variables est similaire pour les 3 cas.

### III.2.2 Exemple de M. X

La Figure 7 illustre la variation des  $\log_{10}(\text{LRs})$  obtenue dans les deux scénarios pour l'observation  $E_1$  (0.1-0.2 ng) et  $E_2$  (0.01-0.02 ng) spécifiques au cas de M. X. Chaque combinaisons de résultats  $E_i=\{E_1, E_2\}$  peut être directement investigué sélectionnant « Part 1 : Sensitivity analysis » dans l'application :

[https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/).

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

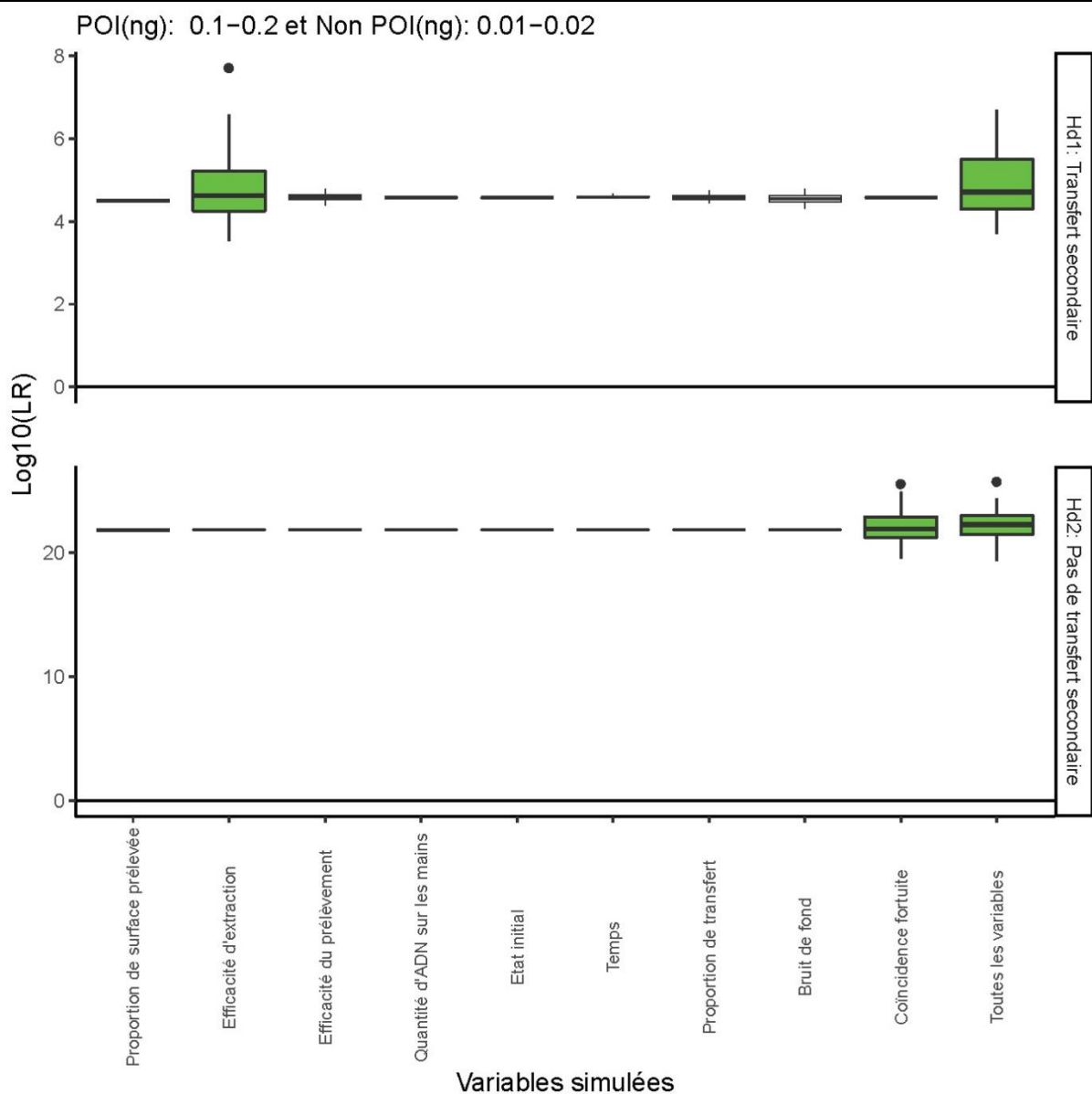


Figure 7 : Variation pour chaque variable des  $\log_{10}(\text{LRs})$  lorsque des quantités entre 0.1-0.2 ng d'ADN correspondant au profil du POI et 0.01-0.02 ng d'ADN ne correspondant pas au profil du POI sont obtenues considérant les scénarios 1 (en haut) et 2 (en bas). Les variables impactant significativement (plus d'un ordre de grandeur) les  $\log_{10}(\text{LRs})$  sont indiquées en vert.

Dans le scénario 1, pour lequel la proposition de la défense allègue que le POI a serré la main de l'AO et pour lequel les observations sont  $E_1$  (0.1-0.2 ng d'ADN correspondant au profil du POI) et  $E_2$  (0.01-0.02 ng d'ADN différent du profil ADN du POI), la variable contribuant le plus à la variation des valeurs de LR est la variable « Efficacité d'extraction ».

Dans le scénario 2, pour lequel la proposition de la défense allègue que le POI n'a jamais touché le couteau retrouvé sur les lieux mais que c'est l'AO, qui a poignardé la victime, l'analyse de sensibilité montre que la « Coïncidence fortuite » est la variable qui impacte significativement le LR.



III.2.3 Effet des quantités d'ADN de POI et de Non POI sur le rapport de vraisemblance

La Figure 8 illustre les  $\log_{10}(\text{LR})$  obtenus en fonction de la quantité d'ADN de Non POI observé pour une quantité fixe d'ADN correspondant au POI et inversement. Pour rappel, la quantité d'ADN de Non POI est la quantité d'ADN différent du profil ADN du POI.

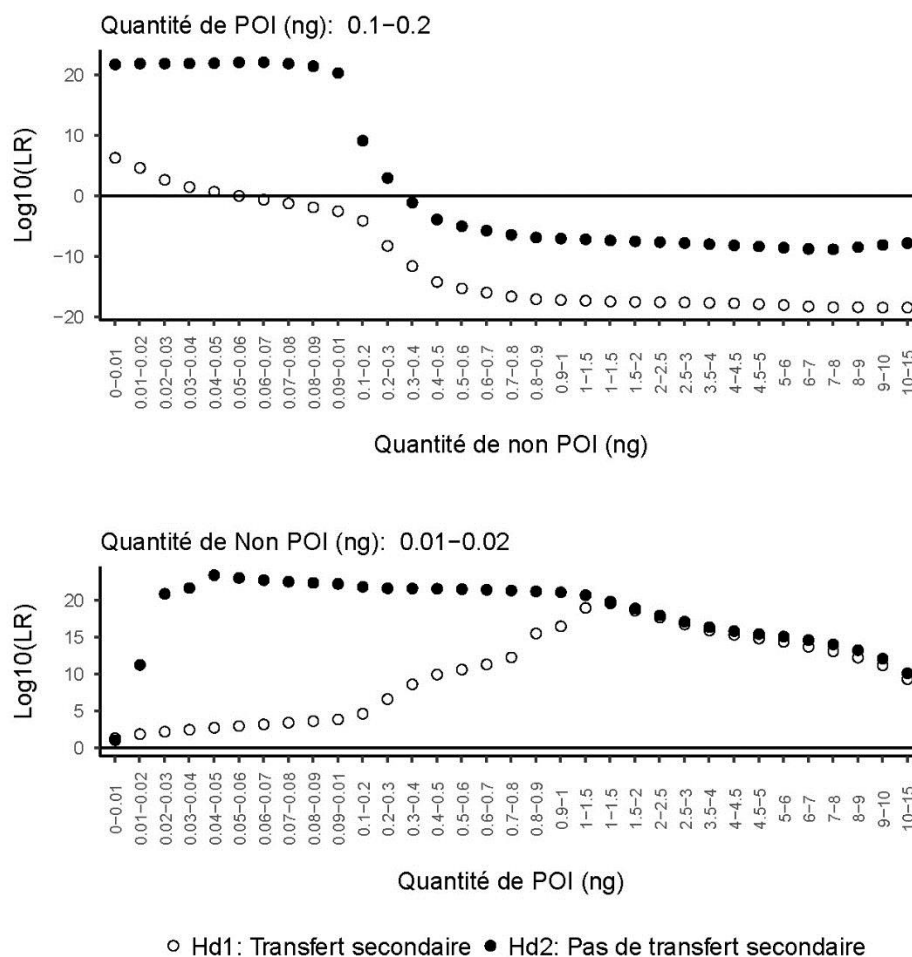


Figure 8 :  $\log_{10}(\text{LR})$  en fonction de la quantité d'ADN de Non POI pour une quantité fixe d'ADN de POI (en haut) et en fonction de la quantité d'ADN de POI pour une quantité fixe d'ADN de non POI (en bas) pour les deux scénarios.

Pour cette quantité fixe d'ADN de POI ( $E_1$  : 0.1-0.2), plus la quantité de Non POI augmente, plus le  $\log_{10}(\text{LR})$  diminue jusqu'à un obtenir une valeur seuil pour les deux scénarios.

Plus spécifiquement, dans le scénario 1 (pour lequel la proposition de la défense  $H_{d1}$ , allègue que le POI a serré la main de l'AO), deux phases peuvent être décrites :

- Une phase de diminution : le  $\log_{10}(\text{LR})$  diminue lorsque la quantité d'ADN de Non POI augmente. Cette diminution s'explique principalement par le fait que sous  $H_p$ , la quantité de POI par rapport à celle de du Non POI s'explique de moins en moins bien tandis que sous  $H_{d1}$ , ces quantités de POI et de Non POI s'explique de mieux en mieux.
- Une phase de plateau : le  $\log_{10}(\text{LR})$  ne varie pas alors que la quantité d'ADN de non POI augmente. Ce plateau s'explique par le fait que la quantité de Non POI est tellement grande et le  $\log_{10}(\text{LR})$  tellement bas que l'augmentation de la quantité de Non POI n'a plus d'influence sur l'ordre de grandeur du LR.

Dans le scénario 2 (pour lequel la proposition de la défense  $H_{d2}$ , allègue que le POI n'a jamais touché le couteau retrouvé sur les lieux, ce n'est pas lui mais l'AO, qui a poignardé la victime), 3 phases peuvent être décrites :

- Une phase de plateau : le  $\log_{10}(\text{LR})$  ne varie pas alors que la quantité d'ADN de Non POI augmente. Ce plateau peut s'expliquer par le fait que le  $\log_{10}(\text{LR})$  est tellement haut que l'augmentation de la quantité de Non POI n'a pas d'influence sur l'ordre de grandeur du LR.
- Une phase de diminution : le  $\log_{10}(\text{LR})$  diminue brutalement lorsque la quantité d'ADN de Non POI augmente. Cette forte diminution peut s'expliquer principalement par le fait que sous  $H_p$ , la quantité de Non POI est telle que celle-ci s'explique de moins en moins par le bruit de fond et ainsi les observations s'expliquent alors mieux sous  $H_{d2}$  que sous  $H_p$ .
- Une phase de plateau : le  $\log_{10}(\text{LR})$  ne varie pas alors que la quantité d'ADN de Non POI augmente. Ce plateau peut s'expliquer par le fait que la quantité de Non POI est tellement grande et le  $\log_{10}(\text{LR})$  tellement bas que l'augmentation de la quantité de non POI n'a pas d'influence sur l'ordre de grandeur du LR.

De manière générale, pour la quantité fixe d'ADN de Non POI ( $E_2 : 0.01-0.02$ ), plus la quantité de POI augmente, plus le  $\log_{10}(\text{LR})$  augmente jusqu'à un obtenir une valeur seuil, puis diminuer pour les deux scénarios.

Plus spécifiquement, dans les deux scénarios, deux phases peuvent être décrites :

- Une phase d'augmentation : le  $\log_{10}(\text{LR})$  augmente lorsque la quantité de POI augmente. Cette augmentation s'explique principalement par le fait que sous  $H_p$ , la quantité de POI par rapport à celle du Non POI s'explique de mieux en mieux tandis que sous  $H_d$ , ces quantités de POI et de Non POI s'expliquent de moins en moins jusqu'à atteindre une valeur maximale pour laquelle les observations s'expliquent le mieux sous  $H_p$ .
- Une phase de diminution : à partir de cette quantité de POI pour laquelle le LR maximal est atteint, le  $\log_{10}(\text{LR})$  diminue lorsque la quantité d'ADN de POI augmente. Cette diminution s'explique par le fait que la quantité de POI est tellement grande qu'elle s'explique de moins en moins sous  $H_p$  (par le fait que c'est une trace de contact d'une certaine quantité qui est déposée et que plus cette quantité est grande, moins elle s'explique comme une trace n'étant pas du fluide biologique).

L'influence de la quantité de Non POI pour chaque quantité fixe de POI ainsi que l'influence de la quantité de POI pour chaque quantité fixe de Non POI peuvent être visualisées en sélectionnant « Part 1 : Sensitivity analysis » sur le site :

[https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/)

De manière générale, lorsque le scénario 1 est considéré, les LR sont plus bas que si le scénario 2 est considéré, ce qui était attendu.

### III.3 Identification des variables nécessitant l'acquisition de données supplémentaires

Une fois les potentielles variables d'intérêt identifiées, l'objectif est de sélectionner celles pour lesquelles il est utile d'acquérir des données supplémentaires, c'est-à-dire que pour ces variables, l'ajout de données supplémentaires contribue à une diminution de la variation sur le  $\log_{10}(\text{LRs})$ .

Si nous observons les résultats généraux prenant en compte l'ensemble des observations possibles  $E_i = \{E_1, E_2\}$ , l'objectif à atteindre est une variation de l'étendue des  $\log_{10}(\text{LRs})$  inférieure à 1, lorsque les données des variables non instanciées (sauf « Temps et « Proportion de surface prélevée ») sont simultanément modifiées.

Si un cas avec une observation spécifique  $E_1$ - $E_2$  est étudié, l'objectif à atteindre est une variation des  $\log_{10}(\text{LRs})$  inférieure à 1, lorsque les données des variables non instanciées (sauf « Temps et « Proportion de surface prélevée ») sont simultanément modifiées.

Cent simulations ont été effectuées pour chacune des variables. Cent autres simulations ont aussi été réalisées en ré-échantillonnant simultanément les données des variables non instanciées, excepté les variables « Temps » et « Proportion de surface prélevée » dont les données sont fixes.

Les simulations sont effectuées de manière identique aux précédentes simulations mais en ré-échantillonnant, dans les données initiales, un nombre de données trois fois plus élevé que le nombre de données à disposition pour la paramétrisation initiale, comme décrit dans l'article Samie et al. (2019<sub>(b)</sub>) (Annexe II.1). Pour rappel, ces simulations permettent ainsi d'étudier l'impact de la quantité de données à disposition sur le LR et non la pertinence de ces données.

### III.3.1 Résultats généraux

La Figure 9 illustre les variations des étendues des  $\log_{10}(\text{LRs})$  obtenues suite aux simulations, pour chacune des variables, quel que ce soit les observations  $E_i = \{E_1, E_2\}$ , lorsque 3 fois plus de données sont ajoutées pour les variables d'intérêt sélectionnées. Elles permettent d'identifier les variables pour lesquelles il est, en général, utile d'acquérir des données supplémentaires.

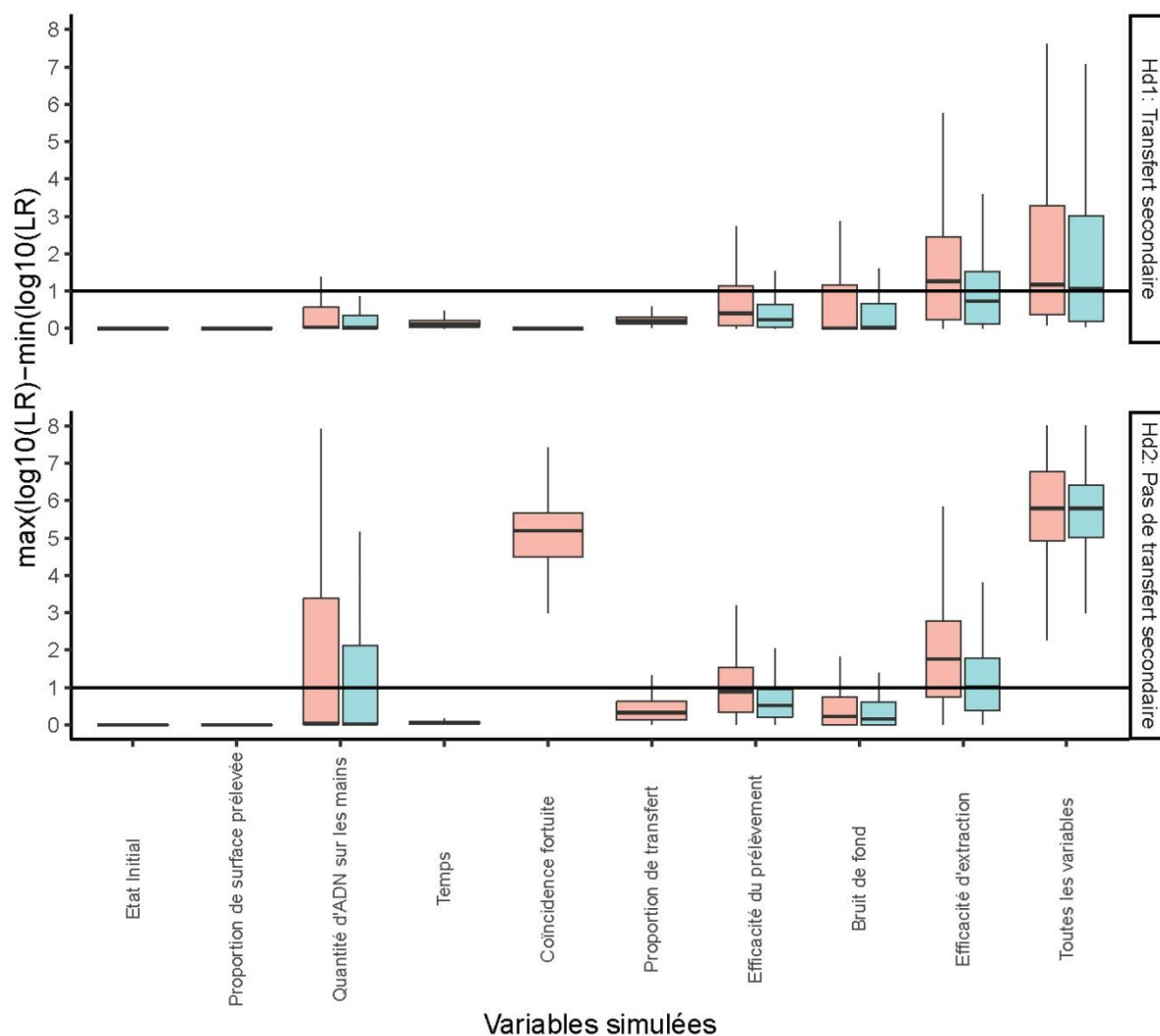


Figure 9 : Variation des étendues des  $\log_{10}(\text{LRs})$  obtenues pour l'ensemble des observations  $E_1$  et  $E_2$  pour chacune des variables pour chacun des scénarios lorsque le nombre de données est trois fois plus grand (en bleu) que le nombre de données informant initialement le BN (en rose).

Une diminution de la variation des étendues des  $\log_{10}(\text{LRs})$  est observée pour chacune des variables pour chacun des scénarios lorsque le nombre de données est augmenté, ce qui était attendu.

La Figure 9 montre que, pour le scénario 1, en acquérant trois fois plus de données sur la « Quantités d'ADN sur les mains », celle-ci peut alors ne plus être une variable d'intérêt, c'est-à-dire que la variation est en dessous de l'ordre de grandeur. Cependant la variation des  $\log_{10}(\text{LRs})$  lorsque toutes les variables sont simulées ne diminue pas en dessous de 1.

La Figure 9 montre que, pour le scénario 2, malgré l'acquisition de données supplémentaires, la variation des étendues des  $\log_{10}(\text{LRs})$  ne diminue pas en dessous de 1, ce qui était attendu. En effet, la variable « Coïncidence fortuite » est celle qui induit une variation de LR

importante par rapport aux autres. Or cette variable n'est pas simulée à partir d'un nombre de données. Il est donc attendu que la variation des étendues des  $\log_{10}(\text{LRs})$ , lorsque toutes les variables sont simulées simultanément, reste supérieure à 1, quel que soit le nombre de données supplémentaires informant les variables. L'ensemble est conditionné par la probabilité de coïncidence fortuite.

Il n'est donc, généralement, pas nécessaire de mener des expériences supplémentaires pour le scénario 2, si la pertinence des données initiales est jugée adéquate.

### III.3.2 Cas de M. X

La Figure 10 illustre la variation sur les  $\log_{10}(\text{LR})$  obtenus pour le cas de M. X pour chacune des variables, pour chacun des scénarios, lorsque l'ensemble des données est exploité et lorsque le nombre de cet ensemble de données est multiplié par trois.

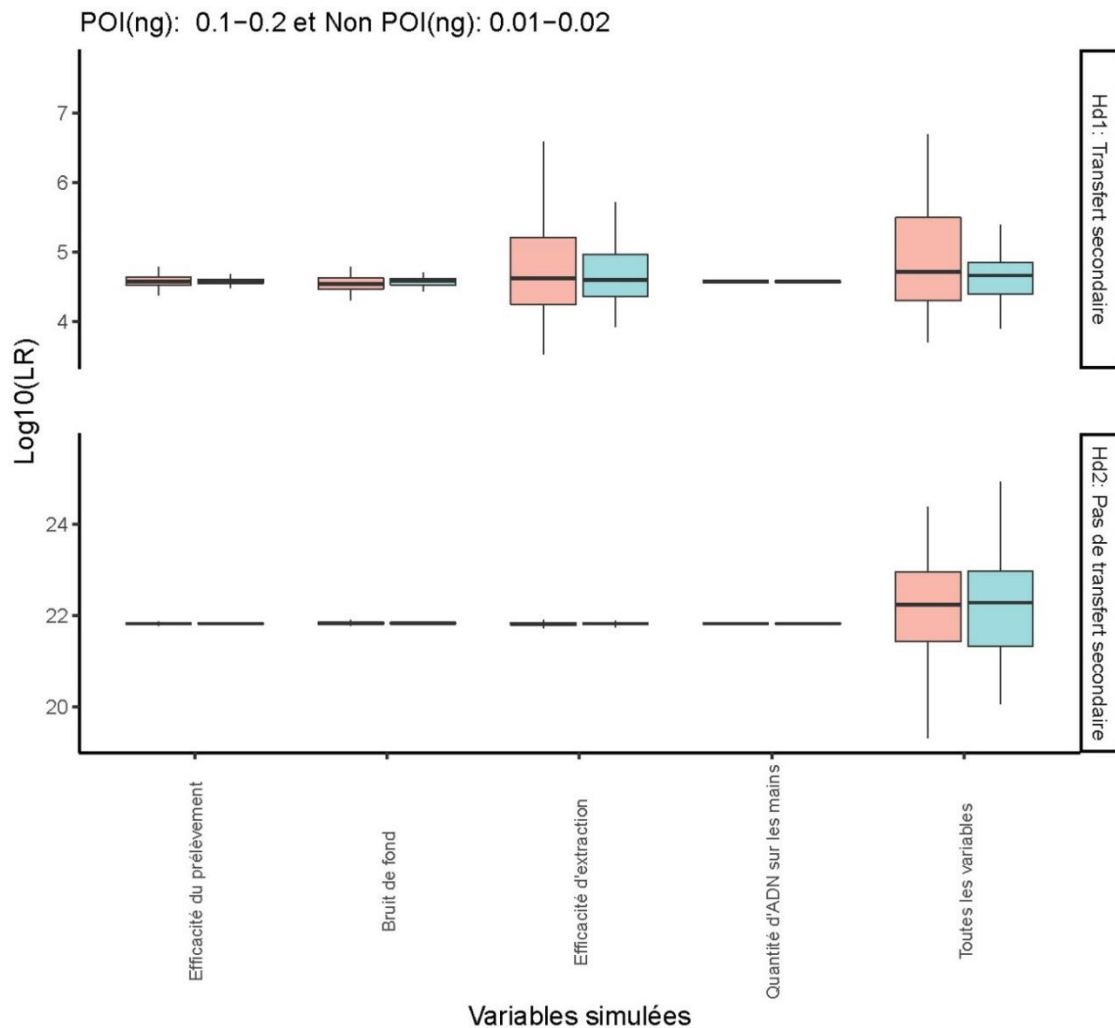


Figure 10 : Variation des  $\log_{10}(\text{LRs})$  obtenus lorsque des quantités entre 0.1-0.2 ng d'ADN correspondant au profil du POI et 0.01-0.02 ng d'ADN ne correspondant pas au profil du POI sont obtenues considérant chacun des scénarios lorsque le nombre de données est trois fois plus grand (en bleu) que le nombre de données informant initialement le BN (en rose).

Pour le scénario 1, l'augmentation par un facteur 3 du nombre de données permet de diminuer l'étendue des  $\log_{10}(\text{LRs})$  mais pas en dessous de 1, lorsque les données de toutes les variables non instanciées (sauf « Temps et « Proportion de surface prélevée ») sont modifiées simultanément. Cela signifie qu'il faudrait au moins 45 données informant l'efficacité d'extraction pour atteindre une variation inférieure à l'ordre de grandeur et donc en acquérir au moins 30 de plus (15 données étaient initialement disponibles).

En conclusion, la variable « Efficacité d'extraction » est donc une variable d'intérêt dans laquelle il serait utile d'investir. Cela nécessiterait l'acquisition de plus de 30 données supplémentaires pour le scénario 1 afin d'obtenir une variation inférieure à 1. Le choix d'acquérir ou non ces données supplémentaires sera discuté dans la partie Discussion (Voir III.4.4).

Pour le scénario 2, les résultats des simulations montrent qu'une augmentation du nombre de données pour les variables d'intérêt ne permet pas de réduire la variation des  $\log_{10}(\text{LRs})$  lorsque les données de toutes les variables sont ré-échantillonnées simultanément. La variable « Coïncidence fortuite » est celle qui induit la variation de LR importante. Or cette variable n'est pas simulée à partir d'un nombre de données comme détaillé par Samie et al. (2019<sub>(b)</sub>) (Annexe II.1). Aucune donnée supplémentaire ne peut donc diminuer la variation des LR résultant des simulations effectuées sur cette variable. En revanche, lors de l'évaluation d'un cas d'espèce, le profil ADN est connu. Si le profil ADN est connu, la probabilité de coïncidence fortuite est connue et ainsi l'état de cette variable pourrait être instancié.

En conclusion, pour le cas de M. X, dans le scénario 2, il n'est pas utile d'acquérir des données supplémentaires.

## III.4 Discussion

### III.4.1 Prise en compte de la variation relative entre POI et Non POI

Il ressort de la littérature qu'il existe une grande variabilité inter-individu sur les quantités d'ADN retrouvées sur les mains et transférées. Les données sur les « Quantités d'ADN sur les mains » et « Proportion de transfert » se référant au POI peuvent être différentes des données se référant à l'AO.

Or, dans le cas initial, les quantités d'ADN obtenues pour différents individus n'étant pas disponibles, aussi les mêmes données sont utilisées pour informer les tables de probabilités

conditionnelles des variables « Quantité d'ADN sur les mains » et « Proportion de transfert » tant pour le POI et que pour l'AO.

Lors des simulations, les données utilisées pour informer les tables de probabilités conditionnelles sont ré-échantillonnées et induisent des différences entre les données associées aux variables « Quantité d'ADN sur les mains » et « Proportion de transfert » se rapportant au POI ou à l'AO. Cependant, ces différences ne reflètent pas les différences réelles qui peuvent exister entre les quantités d'ADN sur les mains ou les quantités d'ADN transférées par différentes personnes.

Au regard de cette limitation et sur la base de la littérature précédemment citée dans le chapitre 1, nous restons sur l'hypothèse que les variations induites par les données utilisées pour les variables « Quantités d'ADN sur les mains » et « Proportion de transfert » sont sous-estimées dans ces simulations.

#### III.4.2 Généralisation des résultats

L'impact sur le LR du ré-échantillonnage des données associées aux variables est similaire pour les 3 cas, cas de M. X, cas Transfert élevé et cas Transfert faible, ceci pour les deux scénarios.

Cette observation suggère que l'impact des variables sur le LR, pour la plupart des observations  $E_1$ - $E_2$ , ne dépend pas des circonstances du cas mais du scénario c'est-à-dire des propositions posées. Les variables mises en évidence comme ayant généralement un impact élevé sur le LR restent valides pour tous les scénarios dont les propositions sont  $H_{d1}$  ou  $H_{d2}$  et pour lesquels le BN présenté dans cette thèse est applicable et cela quelles que soient les circonstances du cas.

Pour évaluer un cas, il est essentiel de s'intéresser, de préférence, aux variables identifiées correspondant aux observations  $E_i=\{E_1, E_2\}$  propres au cas en question, plutôt qu'à celles identifiées de façon générique considérant l'ensemble des observations  $E_i=\{E_1, E_2\}$  présentées dans ce manuscrit (Parties « Résultat général » de ce chapitre). Pour rappel, chaque combinaisons d'observations  $E_i=\{E_1, E_2\}$  peut être directement investigué sélectionnant « Part 1 : Sensitivity analysis » dans l'application :

[https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/).



### III.4.3 Effet des propositions sur le rapport de vraisemblance

Nous observons pour la plupart des observations  $E_i = \{E_1, E_2\}$  que si la proposition de la défense est « Quelqu'un d'autre l'a agressé », le LR obtenu sera plus haut que si la défense allègue un possible transfert secondaire, ce qui est attendu. Cette observation a également été faite par Pun (2016).

Il est général attendu que la partie et non l'expert soit en mesure de poser sa proposition. Or les parties peuvent ne pas avoir connaissance qu'un transfert secondaire est possible. Dans ce cas, ils ne sont donc pas en mesure de pouvoir proposer une activité alléguée impliquant un transfert secondaire. Cependant, l'expert, lui, sait qu'un transfert secondaire est possible. La question doit donc être posée : doit-on toujours considéré la possibilité d'un transfert secondaire même si la partie ne l'allègue pas et indique seulement que « quelqu'un d'autre a agressé la victime » ?

Plusieurs options sont envisageables :

- toujours considéré la possibilité d'un transfert secondaire même si la partie ne l'invoque pas explicitement,
- ne considéré la possibilité d'un transfert secondaire que si celle-ci est mise en avant par la partie,
- déterminer des probabilités « a priori » pour chacune des propositions, en fonction du contexte de l'affaire. Par exemple, dans ce cas, trois propositions sont alléguées, une par le procureur ( $H_p$  : le POI a agressé la victime) et une par la défense ( $H_d$  : quelqu'un d'autre a agressé la victime) qui peut être décomposée en deux propositions  $H_{d1}$  (Transfert secondaire allégué) et  $H_{d2}$  (Aucun transfert secondaire allégué). Supposons que, d'après les circonstances de l'affaire,  $H_p$  soit autant probable que  $H_d$ .

Deux exemples sont proposés :

- Dans un premier exemple, la zone de vie de la personne d'intérêt (lieu de travail, lieu d'habitation, lieu d'intérêt, trajet en transport) est entièrement différente de celle de la victime. Il est considéré que la probabilité que  $H_{d1}$  soit vraie soit plus petite que celle que  $H_{d2}$  soit vraie. Nous poserions alors par exemple que la probabilité a priori que  $H_p$  soit vraie est de 0.5, que la probabilité a priori que  $H_{d1}$  soit vraie est de 0.1 et que la probabilité a priori que  $H_{d2}$  soit vraie est de 0.4.

- Dans un deuxième exemple, la zone de vie de la personne d'intérêt recoupe toute ou partie celle de la victime. Il est considéré que la probabilité que  $H_{d1}$  soit vraie soit plus grande que celle que  $H_{d2}$  soit vraie. Nous poserions alors par exemple que la probabilité a priori que  $H_p$  soit vraie est de 0.5, que la probabilité a priori que  $H_{d1}$  soit vraie est de 0.4 et que la probabilité a priori que  $H_{d2}$  soit vraie est de 0.1.

La deuxième option paraît arbitraire car elle dépend essentiellement des connaissances des parties, qui ne sont pas des spécialistes dans le domaine ADN. La troisième option permet de tenir compte des deux possibilités (Transfert secondaire ou non) mais nécessite d'assigner des probabilités a priori sur les propositions, ce qui devrait incomber à la Cour et non à l'expert. Au vu de ces différents éléments, la première option est la plus adéquate. Cependant, il faut être averti que, si la proposition du procureur est effectivement vraie, les LR obtenus seront plus bas si  $H_{d2}$  est alléguée par la défense que si  $H_{d1}$  est alléguée.

Le scénario 2 ( $H_{d2}$  : Pas de transfert allégué) permettait de rendre compte de l'impact du choix des propositions sur le LR, discussion traitée dans cette partie même. Il ne sera plus étudié dans la suite de ce travail.

#### III.4.4 Nombre d'expériences

Au vu des résultats observés à partir de la Figure 9, nous savons qu'afin d'étudier le scénario 1 ( $H_{d1}$  : transfert secondaire), il serait utile d'acquérir au moins 100 données supplémentaires sur la quantité d'ADN sur les mains (50 données initialement disponibles), 42 données supplémentaires sur l'efficacité du moyen de prélèvement (21 données initialement disponibles), 124 données supplémentaires sur le bruit de fond (62 données initialement disponibles) et 30 données supplémentaires sur l'efficacité d'extraction (15 données initialement disponibles).

Le choix d'acquérir ou non ces données supplémentaires dépend de la confiance que l'on veut accorder au LR obtenu à partir des données initiales mais aussi des moyens à disposition pour aider la Cour. Pour ce travail, nous jugeons que le nombre de données à acquérir n'est pas raisonnable du point de vue pratique pour un laboratoire opérationnel à cause du temps et des coûts engendrés. Certaines données pourront être réutilisées pour tous types de cas traités si les méthodes restent inchangées (données d'efficacité de prélèvement et d'extraction par exemple), ce qui justifierait d'acquérir un très grand nombre de données. En revanche, nous supposons qu'il serait difficile d'acquérir de si grands nombres de données pour les variables

---

pour lesquelles les données seraient à acquérir pour chaque cas, (en fonction des résultats du chapitre suivant de cette thèse).

Le choix a été fait d'acquérir 30 données pour chacune des variables d'intérêt dans la suite de ce travail. Ce chiffre semble raisonnable dans le contexte de la pratique. Cependant, le nombre de données à acquérir peut être réévalué par la suite en fonction des résultats obtenus.

**En conclusion**, la méthodologie est composée de plusieurs étapes :

- Construire un BN représentant la problématique du cas.
- Identifier les variables d'intérêt sur la base d'une analyse de sensibilité du BN.
- Identifier les variables pour lesquelles l'acquisition de données supplémentaires peut être utile.

Les variables d'intérêt pertinentes pour le scénario 1 (Un transfert secondaire de l'ADN du POI est allégué sous  $H_{d1}$ ) sont en général :

- « Efficacité de prélèvement »,
- « Efficacité d'extraction »,
- « Quantité d'ADN sur les mains » et « Proportion de transfert »,
- « Bruit de fond ».

Des données supplémentaires sur ces cinq variables doivent être acquises. Une étude sur chacune de ces variables est donc présentée dans le chapitre suivant.

Pour le scénario 2 (Aucun transfert secondaire de l'ADN du POI n'est allégué sous  $H_{d2}$ ), aucune expérience ne permet de diminuer la variation induite sur les LR, dû à la méthode de simulation de la variable « Coïncidence fortuite ». Aucune variable d'intérêt pertinente n'est donc à retenir, dans la mesure où les données utilisées pour paramétrer initialement le réseau sont jugées adaptées au cas.

## Chapitre IV- Etude du type de données à utiliser

---

Ce chapitre comprend deux parties.

La première s'intéresse aux données « spécifiques ». Les données sont dites « spécifiques » au cas de M. X, si elles proviennent d'expériences reproduisant les circonstances du cas effectuées, avec le POI et appliquant les moyens utilisés (kit d'extraction et moyen de prélèvement) pour analyser la trace ADN. Dans cette partie, les données spécifiques au cas de M. X sont acquises pour chacune des variables d'intérêt identifiées au chapitre précédent.

La deuxième partie étudie les impacts possibles du type de données utilisé pour évaluer les observations relatives au cas de M. X (données spécifiques au cas de M. X et/ou données issues de la littérature) sur le LR. Cette deuxième partie permet de définir dans quelle mesure les données informant le BN doivent être spécifiques ou non au cas en question.

### IV.1 Acquisition de données spécifiques

L'objectif de cette première partie est d'acquérir des données spécifiques au cas de M. X pour chacune des variables identifiées :

- « Efficacité d'extraction » et « Efficacité de prélèvement »,
- « Quantité d'ADN sur les mains », « Proportion de transfert » et « Bruit de fond ».

#### IV.1.1 Efficacité d'extraction et efficacité de prélèvement

L'analyse de sensibilité a montré que, dans la plupart des observations, un ré-échantillonnage des données d'efficacité d'extraction ou d'efficacité de méthode de prélèvement contribue significativement à la variation des LR associés à chaque situation de simulation.

De cette observation découle la campagne d'acquisition de données décrites ci-après.

Les données acquises pour les variables, « Efficacité d'extraction » et « Efficacité de prélèvement », sont détaillées dans l'étude de Samie et al. (2019<sub>(a)</sub>) (Annexe I.2). Cette étude sera brièvement décrite avant de spécifier les données utilisées pour paramétrer le BN.

#### **Etude de l'efficacité d'extraction**

Samie et al. (2019<sub>(a)</sub>) (Annexe I.2) ont étudié l'efficacité d'extraction ADN manuelle de deux kits, le kit QIAshredder-QIAamp Mini DNA en utilisant les colonnes Microcon® 30, et le kit Investigator® Lyse&Spin Basket-Mini DNA QIAamp, en utilisant la cytométrie en flux. La cytométrie en flux permet de sélectionner un nombre précis de cellules et donc de connaître

la quantité initiale utilisée pour les expériences. Pour chaque kit d'extraction ADN, 30 ou 22 expériences ont été réalisées. Pour chaque expérience, un nombre connu de cellules a été déposé directement dans un tube Ependorf servant à l'extraction ADN. L'ADN a ensuite été extrait en suivant les consignes données par le fournisseur du kit testé. Enfin, l'ADN a été quantifié en utilisant le kit Investigator® Quantiplex de Qiagen sur le Rotor-Gene® Q. La quantité initiale d'ADN des échantillons et la quantité obtenue après extraction sont donc connues. L'efficacité d'extraction est définie comme étant la quantité d'ADN obtenue après extraction en pourcentage de la quantité initiale.

En utilisant le kit QIAshredder-QIAamp DNA Mini seulement 23% en moyenne de la quantité initiale d'ADN est récupérée. La quantité d'ADN récupérée avec le kit Investigator® Lyse&Spin Basket- QIAamp DNA Mini est, lui, de 63%.

Nous avons également étudié l'effet d'un changement de laboratoire sur cette efficacité en faisant extraire 30 échantillons par un autre laboratoire utilisant le kit Investigator® Lyse&Spin Basket- QIAamp DNA Mini. Ce changement de structure n'a pas eu d'impact sur les résultats. Cependant, seulement 2 laboratoires ont été étudiés et il faut donc prendre cette observation avec précaution.

Dans une partie non publiée, les auteurs ont également testé l'efficacité du kit d'extraction ADN PrepFiler Express™ de ThermoFisher Scientific, l'extraction étant effectuée par un robot TECAN. Trente expériences ont également été réalisées suivant un protocole identique à celui décrit ci-dessus. La quantité d'ADN initiale récupérée après l'extraction est alors de 30%. Toutefois, ce kit ne permet pas de déposer les cellules directement dans le tube, de dimension trop grande pour le cytomètre en flux. Un transfert des cellules du tube utilisé pour le cytomètre en flux dans celui utilisé par le robot d'extraction est nécessaire. Une perte d'ADN est possible lors de ce transfert. Nous attendons donc que plus de 30% de la quantité initiale d'ADN soit récupéré en utilisant ce kit associé à l'utilisation du robot.

L'ensemble des résultats de quantification et d'efficacité d'extraction de chaque kit sont présentés en Annexe IV.1.

Pour le cas de M. X, la trace ADN est extraite avec le kit QIAshredder-QIAamp DNA Mini. Les données utilisées pour informer le BN pour le cas de M. X sont donc **les données d'efficacité d'extraction du kit QIAshredder-QIAamp DNA Mini** présentées dans l'Annexe VI.1.

## Effacité de prélèvement

Samie et al. (2019<sub>(a)</sub>) (Annexe I.2) ont également étudié l'efficacité de l'écouvillon (FLOQSwab™ de COPAN) pour libérer de l'ADN. Vingt mille cellules introduites dans 1.1 mL de PBS, qui permet de conserver les membranes cellulaires, ont été sélectionnées pour cette étude. Trente expériences ont été réalisées. Pour chaque expérience, 35 µL du mélange de cellules et de PBS, soit 636 cellules, est déposé directement sur un écouvillon puis l'ADN est extrait en suivant le protocole du kit utilisé, le QIAshredder-QIAamp DNA Mini avec colonnes Microcon® 30. Les extraits d'ADN sont ensuite quantifiés en utilisant le kit Investigator® Quantiplex de Qiagen sur le Rotor-Gene® Q. Les résultats de la quantification sont présentés en Annexe IV.2.

La quantité initiale d'ADN ainsi que la quantité d'ADN prélevée et extraite sont connues donc l'efficacité combinée du prélèvement et de l'extraction peut être calculée. L'efficacité du kit d'extraction est également connue comme vu précédemment. En revanche, la quantité d'ADN qui a été prélevée avant d'être extraite est inconnue donc l'efficacité de prélèvement, qui est le pourcentage de la quantité d'ADN prélevée par rapport à la quantité initiale, est inconnue. Les paramètres de la distribution représentant l'efficacité de prélèvement peuvent être déterminés d'après les données d'efficacité d'extraction et les données d'efficacité combinée de prélèvement et d'extraction, comme l'illustre la Figure 11.

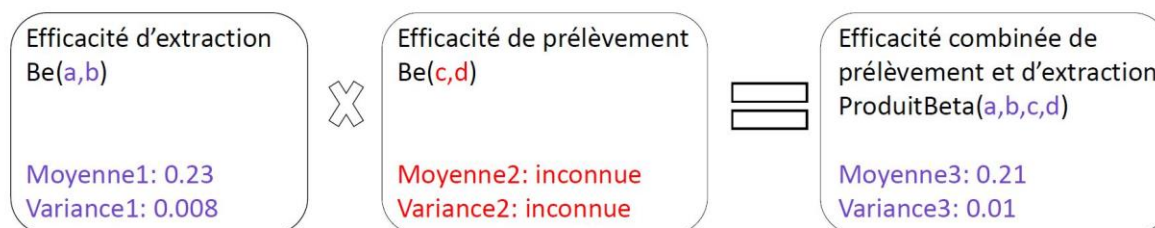


Figure 11 : Illustration des efficacités d'extraction, de prélèvement ainsi que de l'efficacité combinée de prélèvement et d'extraction d'ADN avec les paramètres connus (en violet) et inconnus (en rouge) associés à chaque distribution.

Les données d'efficacité d'extraction et d'efficacité de prélèvement suivent respectivement des distributions Be(a, b) et Be(c, d).

Les paramètres c et d sont calculés à partir des équations définies par Dufresne (2010) :

$$c = \frac{X^2 - XY}{XY + Y}$$

$$d = \frac{X - Y}{XY + Y}$$

Avec:

$$X = \frac{\text{moyenne3/moyenne1}}{1 - (\text{moyenne3/moyenne1})}$$

Et:

$$Y = \frac{\text{variance3}}{\text{moyenne3} * \text{moyenne1}} * \frac{(a + b + 1)}{a + 1}$$

Ainsi les données d'efficacité de prélèvement suivent une distribution Be(32.26, 0.98).

En échantillonnant aléatoirement 1000 valeurs de cette distribution Be(32.26, 0.98), les résultats montrent que les FLOQSwab™ relâchent environ 97% des cellules préalablement déposées sur celui-ci.

Pour le cas de M. X, la trace ADN a été prélevée avec un FLOQSwab™. Les paramètres  $c$  et  $d$  de la distribution Beta représentant l'efficacité de prélèvement sont déterminés d'après **les données d'efficacité d'extraction du kit QIAshredder-QIAamp DNA Mini et d'efficacité combinées de prélèvement du FLOQSwab™ et d'extraction du kit QIAshredder-QIAamp DNA Mini** (Annexe VI.2).

La capacité du moyen de prélèvement à récupérer l'ADN sur un type de surface est un facteur pris en compte dans la proportion de transfert, dont l'acquisition de données est détaillée ci-après.

#### IV.1.2 Quantité d'ADN sur les mains, Proportion de transfert et Bruit de fond

Basé sur l'état de l'art et sur le fait que les données informant le BN pour les variables « Quantités d'ADN sur les mains » et « Proportion de transfert » ne reflètent pas la différence qu'il peut exister entre les différents contributeurs, ces variables ont été considérées comme pertinentes. Sur la base des résultats du chapitre III, le bruit de fond est également pris comme une variable d'intérêt.

De cette observation découle la campagne d'acquisition de données décrites ci-après.

Les données obtenues pour les variables « Quantités d'ADN sur les mains », « Proportion de transfert » et « Bruit de fond » sont issues des expériences présentées dans l'article Samie et al. (2019<sup>(c)</sup>) (Annexe II.2).

Les auteurs ont étudié la capacité de six volontaires à « produire » et à transférer directement leur ADN sur un couteau. Trente expériences par volontaires ont été réalisées. A chaque expérience, les volontaires se frottent les mains puis poignent un savon balistique avec leur main dominante. Le savon balistique a des propriétés physiques similaires à celles des parties molles du corps humain. La force et la pression exercées sur le couteau sont donc proches des conditions réelles d'une agression au couteau. Le couteau est préalablement décontaminé de tout ADN. Le bruit de fond sera constitué de l'ADN présent sur les mains des participant et dont le profil ADN est différent du leur.

Deux prélèvements ADN avec le FLOQSwab™ sont ensuite effectués, un sur le manche du couteau et l'autre sur la main libre n'ayant pas tenu le couteau. L'ADN est ensuite extrait avec le kit QIAshredder-QIAamp Mini DNA en utilisant les colonnes Microcon® 30 puis quantifié en utilisant le kit Investigator® Quantiplex de Qiagen sur le Rotor-Gene® Q. Les profils sont établis suite à l'électrophorèse capillaires effectuée sur l'instrument 3500 (Applied Biosystems) et à l'analyse par le logiciel Genemapper ID-X (ThermoFisher Scientific). La proportion des différents contributeurs des profils ADN pour chaque expérience est calculée en utilisant le logiciel STRmix v2.5. Pour certaines traces, le mélange d'ADN était trop complexe pour que la proportion des différents contributeurs puisse être déterminée. Ces traces n'ont pas été prises en compte. Cette étape permet d'obtenir la quantité d'ADN dont le profil correspond au donneur et la quantité d'ADN étranger. La quantité de bruit de fond correspond à cette quantité d'ADN étranger.

Afin d'obtenir, pour chaque volontaire, des données représentant la plus grande variation à disposer de leur ADN sur la surface de la main et à transférer leur ADN, des consignes ont été données avant chaque expérience. Pour 10 expériences, il a été demandé aux volontaires de se laver les mains justes avant le début de l'expérience. Pour 10 autres, ils ont porté des gants 30 minutes avant et enfin, pour les 10 dernières, aucune indication ne leur a été donnée.

Le plan expérimental est décrit dans la Figure 12.



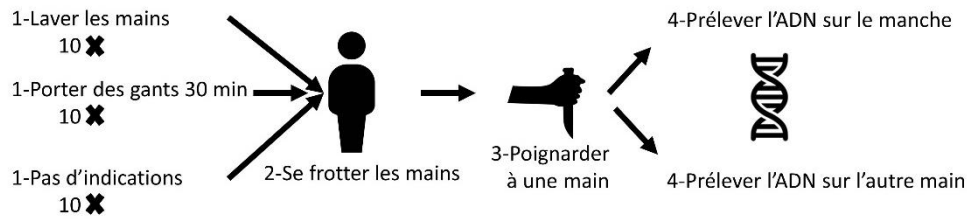


Figure 12 : Plan expérimental afin d'acquérir des données sur la quantité d'ADN sur les mains des participants ainsi que les quantités d'ADN, de chaque participant, transférées sur le couteau.

Les auteurs ont également étudié la capacité de deux individus à transférer secondairement leur ADN sur un couteau via une poignée de main. Un volontaire transférant plus d'ADN (Participant 6) et un individu transférant moins d'ADN (Participant 2) que les autres volontaires ont été sélectionnés. Le plan expérimental est présenté en Figure 13.

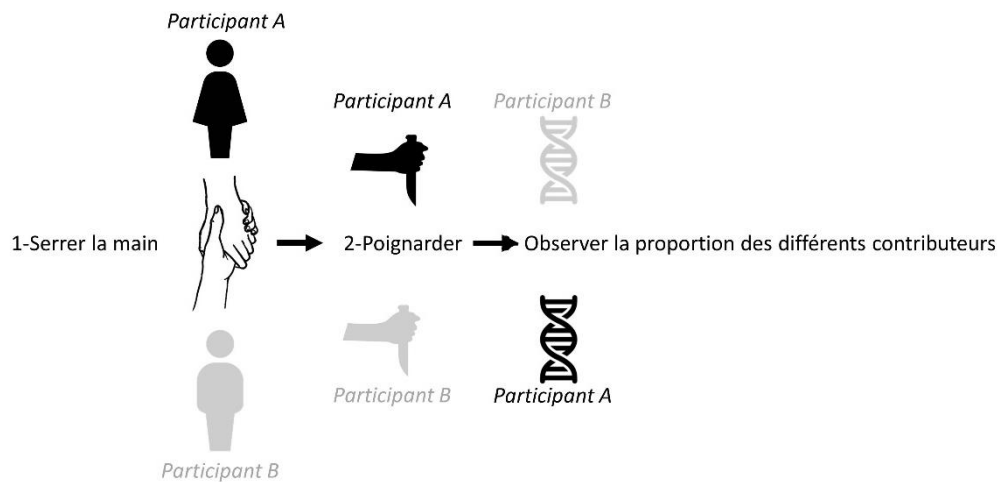
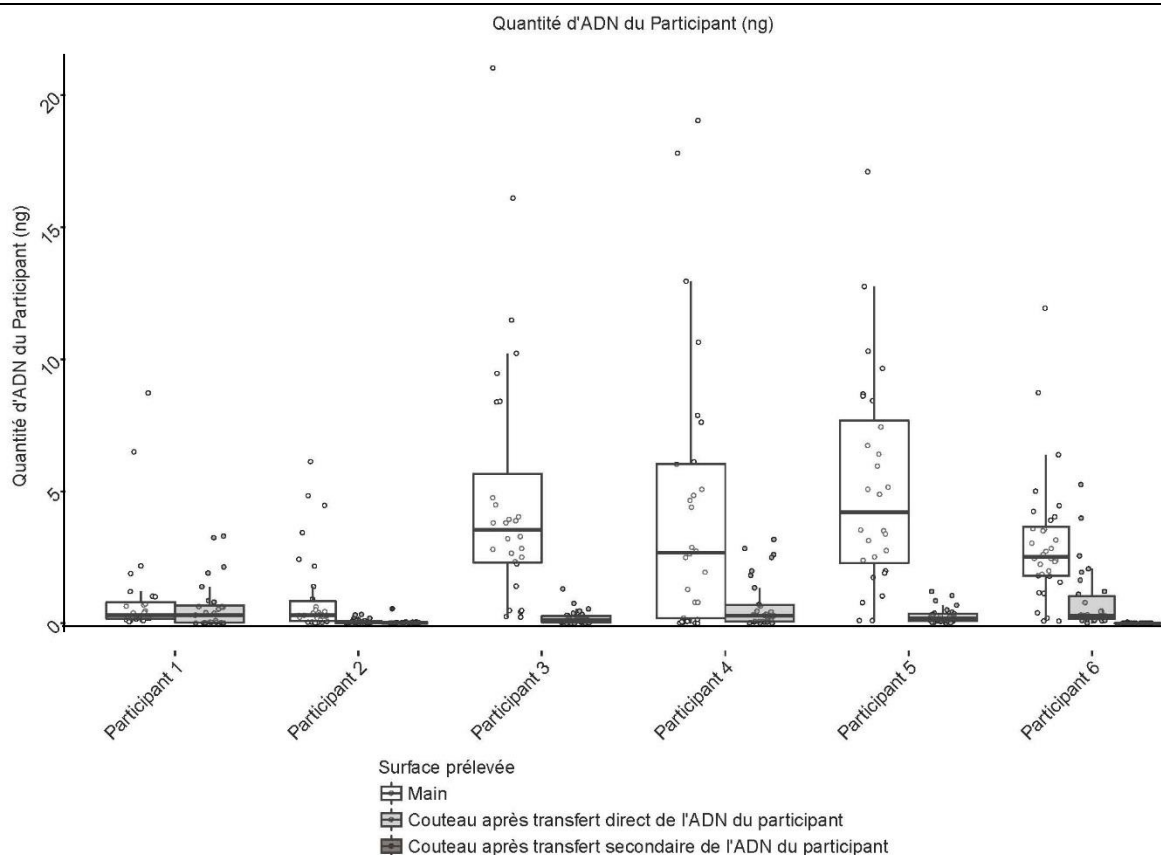


Figure 13 : Plan expérimental afin d'acquérir des données sur la quantité d'ADN de chaque participant, transférée secondairement sur le couteau via une poignée de main.

Les résultats des quantifications et des proportions de contribution pour chaque participant concernant la quantité d'ADN sur les mains, la quantité d'ADN directement transférées sur le couteau ainsi que la quantité d'ADN transférée secondairement sont données en Annexe V.

Les résultats des différentes expériences (Figure 14) montrent qu'un individu peut être qualifié de bon donneur dans un contexte donné et de mauvais donneur dans un contexte différent. Cependant, un individu peut donner en moyenne plus d'ADN ou moins d'ADN qu'un autre.

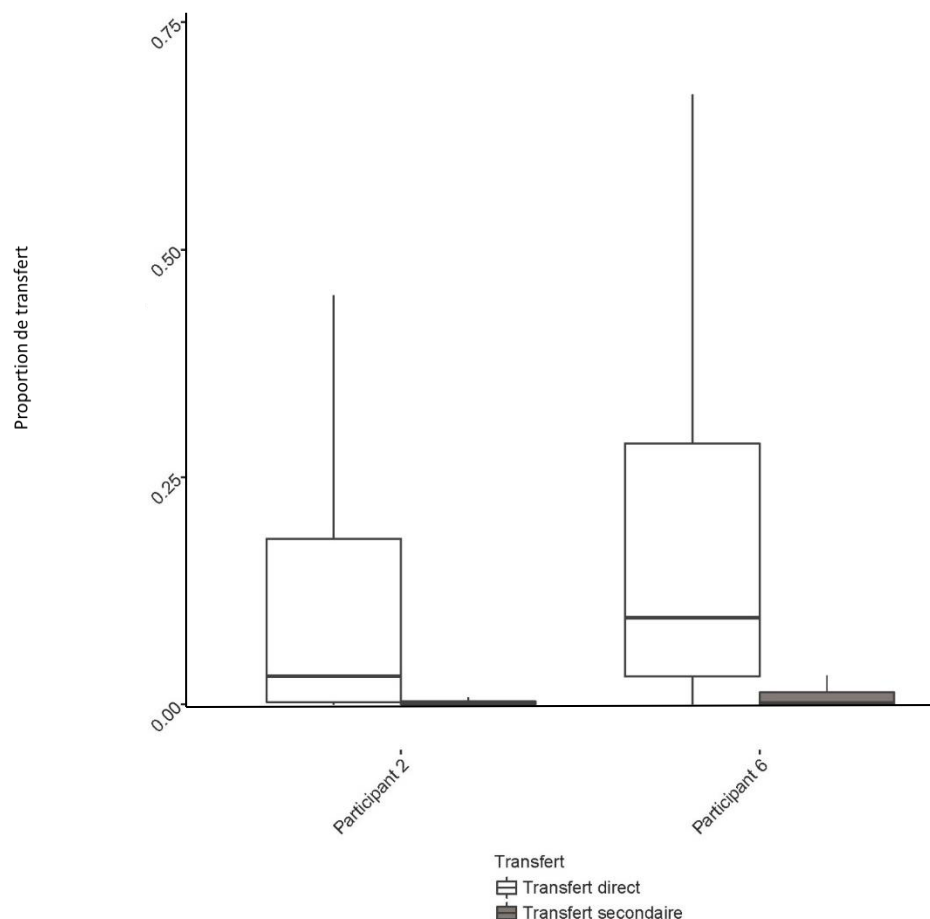


**Figure 14 : Quantités d'ADN (en ng) correspondant à chacun des participants qui ont été prélevées soit directement sur la main, soit sur le couteau suite à un transfert direct ou suite à un transfert secondaire via une poignée de main.**

La proportion de transfert est le rapport entre la quantité prélevée sur la main et la quantité transférée sur le couteau. Nous pouvons observer que la proportion de transfert direct dépend des individus. En effet, la Figure 14 montre qu'un individu peut donner en moyenne beaucoup d'ADN lorsque celui-ci est prélevé directement sur les mains mais très peu quand celui-ci est prélevé sur le couteau après transfert (Par exemple le participant 3). La proportion de transfert direct est en moyenne, dans ce cas, faible, comme le montre la Figure 15 pour deux participants. Inversement, la Figure 14 montre qu'un individu peut donner en moyenne peu d'ADN quand celui-ci est prélevé directement sur les mains mais transférer en moyenne une grande partie de cet ADN sur le couteau (Par exemple le participant 1). La proportion de transfert direct est alors, en moyenne, élevée, comme le montre la Figure 15 pour deux participants.

En revanche, la quantité d'ADN des participants retrouvée sur le manche du couteau après une poignée de main avec le participant qui a poignardé est tellement faible par rapport à la quantité d'ADN sur les mains que la proportion de transfert secondaire observée lors de cette

expérience est très faible, comme le montre la Figure 15. Du fait de ces proportions très faibles, ceci peut signifier que la proportion de transfert secondaire ne dépend pas de l'individu.



**Figure 15 : Variation des proportions de transfert direct et secondaire d'ADN de chaque participant (1000 points générés aléatoirement à partir des distributions modélisant les quantités d'ADN sur les mains et sur le manche du couteau après un transfert direct ou secondaire).**

Ces observations remettent donc en question la définition simpliste d'un bon ou mauvais donneur. Ainsi, tenir compte de l'ensemble des données, c'est-à-dire de cette intra variabilité, est plus adapté que de définir une variable « Statut du donneur » avec comme état possible : « Bon » ou « Mauvais » et d'associer ce statut à une seule probabilité en fonction du statut sélectionné.

Cette étude qualifie également la proportion d'ADN étranger sur les mains d'un individu et sur un couteau suite à un transfert direct ou suite à un transfert secondaire. La quantité d'ADN étranger est négligeable par rapport à la quantité totale d'ADN lorsqu'il est prélevé sur les mains ou sur le couteau après un transfert direct. Cette quantité n'est par contre plus négligeable lorsque l'ADN est prélevé sur le couteau suite à un transfert secondaire.

Les données issues de ces deux recherches vont informer les variables de « Quantité d'ADN sur les mains », « Proportion de transfert » et « Bruit de fond » (Annexe VI). Elles sont résumées ci-dessous :

- « Quantités d'ADN sur les mains »

Les quantités d'ADN retrouvées sur les mains des deux individus (Participant 2 et 6) après prélèvement et extraction sont présentées en Annexe VI.3. Les données qui informent le BN pour cette variable représentent, quant à elles, les quantités d'ADN retrouvées sur les mains avant qu'il y ait eu prélèvement et extraction d'ADN. Les paramètres de distribution représentant les données de quantités d'ADN sur les mains avant la perte d'ADN liée au prélèvement et à l'extraction sont définis à partir des paramètres des distributions représentant les données de quantités d'ADN sur les mains après prélèvement et extraction ainsi que les données combinées d'efficacité de prélèvement et d'extraction. Pour plus de détails, la méthode appliquée est détaillée dans l'article de Samie et al. (2019) (Annexe II.2(c)).

- « Proportion de transfert »

Pour la variable « Proportion de transfert » lorsque la proposition du Procureur est vraie, les données utilisées sont les quantités d'ADN sur les mains (Voir Annexe VI.3) et les quantités d'ADN transférées (Voir Annexe VI.4). Lorsque la proposition de la défense Hd1 est vraie les données utilisées sont les quantités d'ADN sur les mains (Annexe VI.3) et les quantités d'ADN transférées (Voir Annexe VI.5).

- « Bruit de fond »

Les données informant la variable « Bruit de fond » représentent les quantités d'ADN **avant** qu'il y ait eu prélèvement et extraction d'ADN. Les paramètres des distributions liés aux quantités d'ADN en bruit de fond avant prélèvement et extraction sont définis sur la base des paramètres des distributions représentant les données de quantités d'ADN en bruit de fond observées **suite** au prélèvement et à l'extraction d'ADN (présentées en Annexe VI.7) ainsi que les données combinées d'efficacité de prélèvement et d'extraction (présentées en Annexe VI.2). La méthode appliquée est détaillée dans l'article de Samie et al. (2019(c)) (Annexe II.2).

## IV.2 Impacts du type de données utilisé sur le rapport de vraisemblance

### IV.2.1 Introduction

L'objectif de cette partie est de définir dans quelle mesure les données d'efficacité d'extraction, d'efficacité de prélèvement, de quantité d'ADN sur les mains, de bruit de fond et de proportion de transfert doivent être spécifiques ou non au scénario 1 ( $H_{d1}$  : Transfert secondaire) étudié. Nous avons défini les données comme étant « spécifiques » au cas de M. X, si elles proviennent d'expériences reproduisant les circonstances du cas effectuées avec le POI et avec les moyens utilisés (kit d'extraction et moyen de prélèvement) pour analyser la trace ADN. Pour rappel, le scénario 2 ( $H_{d2}$  : Pas de transfert secondaire) n'est pas étudié dans cette partie.

### IV.2.2 Méthode

Pour chacune des cinq variables d'intérêt, le BN est successivement paramétré à partir de trois types de données :

- les données issues de la littérature (Chapitre II), utilisées lors de la paramétrisation initiale,
- les données spécifiques au cas de M. X issues des expériences (Chapitre IV),
- la combinaison de données issues de la littérature et des expériences.

L'estimation des paramètres suit une approche bayésienne (Taroni et al., 2010). Dans un premier temps, les données (ou paramètres de distribution) à priori basées sur les connaissances disponibles avant d'acquérir de nouvelles données sont définies. Celles-ci sont décrites par Samie et al. (2019<sub>(b)</sub>) - (Annexe II.1). Dans un second temps, des données (ou des paramètres de distribution) sont nouvellement acquises. Ici ces données nouvellement acquises sont soit issues de la littérature (Samie et al. 2019<sub>(b)</sub>-Annexe II.1), soit issues des expériences (données spécifiques), soit issues de la combinaison des données de la littérature et des expériences. Les données (ou paramètres de distribution) à priori et les données (ou paramètres de distribution) nouvellement acquises permettent de définir les données (ou paramètres de distribution) à posteriori. Ce sont elles qui viennent paramétriser le BN.

Les données spécifiques, pour les variables « Quantités d'ADN sur les mains », « Bruit de fond » et « Proportion de transfert » sont celles obtenues lorsque le Participant 2 est le POI et le participant 6 est l'AO (le POI est moins donneur que l'AO). Pour rappel, dans la partie IV.1.2 de ce travail, le participant 2 a été identifié comme étant globalement moins bon

donneur que les autres participants et le participants 6 comme étant globalement meilleur donneur que les autres.

Des simulations par ré-échantillonnage sont alors effectuées :

- Cent simulations sont effectuées au cours desquelles les données des variables non instanciées sont ré-échantillonnées, sur le même principe que dans l'article de Samie et al. (2019<sub>(b)</sub>) - (Annexe II.1), pour chaque variable d'intérêt étudiée et pour chacun des 3 types de données utilisés. Lors des simulations effectués pour une variable d'intérêt, les données des autres variables qui ne sont donc pas ré-échantillonnées proviennent des données issues de la littérature. Les données de la variable « Proportion de Transfert » étant obtenues à partir des données de la variable « Quantités d'ADN sur les mains », les données de ces deux variables ont été ré-échantillonnées conjointement.
- Cent simulations sont également effectuées au cours desquelles les données de l'ensemble des variables non instanciées sont ré-échantillonnées. Pour ces simulations, les variables sont informées seulement par des données spécifiques au cas en question (Résultat « Toutes les variables » sur la Figure 17).

L'hypothèse qui sous-tend cette expérimentation stipule que le LR qui devrait être utilisé dérive de l'utilisation de données spécifiques au cas en question. La question est alors de savoir si des données issues de la littérature ou des données combinant celles de la littérature et celles qui sont spécifiques au cas peuvent être utilisées.

La variation observée sur les  $\log_{10}(\text{LR})$  en fonction du type de donnée pour chaque observation  $E_i = \{E_1, E_2\}$  (avec  $E_1$  : quantité d'ADN dont le profil correspond à celui du POI et  $E_2$  : quantité d'ADN dont le profil est différent de celui du POI) pour chaque variable étudiée peut être visualisée en sélectionnant « Part 2 : Type of data analysis » sur le site [https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/).

L'impact du type de données utilisé (données issues de la littérature ou données combinant celles de la littérature et celles qui sont spécifiques) est dit significatif si le  $\log_{10}(\text{LR médian})$  obtenu à partir des simulations par ré-échantillonnage de ce type de données diffère d'un ordre de grandeur supérieure à 1 par rapport au  $\log_{10}(\text{LR médian})$  obtenu en utilisant les données spécifiques au cas en question.

IV.2.3 Résultats

En visualisant les résultats obtenus pour chaque observation sur « Part 2 : Type of data analysis » du site [https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/), nous observons que l'impact du type de données sur le  $\log_{10}(\text{LR})$  diffère en fonction des observations  $E_i = \{E_1, E_2\}$ .

Le résultat observé, lorsqu'une quantité  $E_1$  de 0.8-0.9 ng correspondant au profil ADN du POI et une quantité  $E_2$  de 0.06-0.07 ng d'ADN différent du profil ADN du POI sont obtenues, est présenté dans la Figure 16. Cet exemple montre la tendance qui se dégage de manière générale lorsque les résultats de chaque observation  $E_i = \{E_1, E_2\}$  sont analysés.

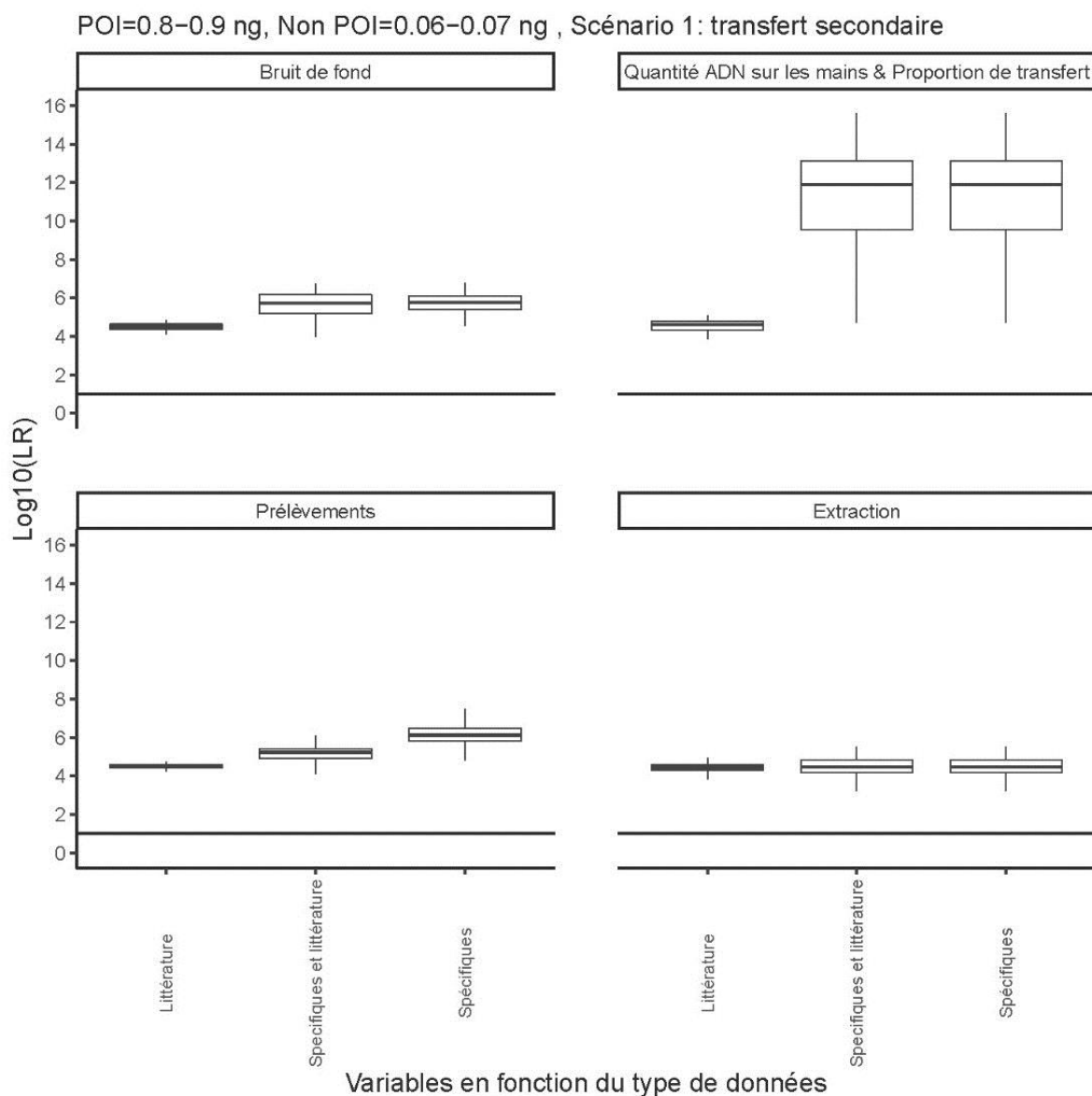


Figure 16 : Variation des  $\log_{10}(\text{LRs})$  en fonction du type de données informant les variables d'intérêt qui sont ré-échantillonnées par simulation.

La Figure 16 montre que :

- Pour la variable « Efficacité d'extraction », il n'y a aucune différence significative entre les  $\log_{10}$ (LRs médians) obtenus en utilisant seulement les données de la littérature, seulement les données des expériences, ou la combinaison des deux types de données. Il n'y a donc pas d'impact du type de données utilisé sur le LR pour cette variable.
- Pour la variable « Efficacité de prélèvement », il y a des différences significatives entre les  $\log_{10}$ (LRs médians) obtenus en utilisant seulement les données de la littérature, seulement les données des expériences, ou la combinaison des deux types de données. Il y a donc un impact du type de données utilisé sur le LR pour cette variable.
- Pour les variables « Bruit de fond », « Quantités d'ADN sur les mains » et « Proportion de transfert », une différence significative entre les  $\log_{10}$ (LRs médians) obtenus en utilisant seulement les données de la littérature et seulement les données des expériences est observée. Il y a donc un impact de l'utilisation des données provenant uniquement de la littérature pour cette variable. En revanche, aucune différence significative entre les  $\log_{10}$ (LRs médians) obtenus en utilisant la combinaison des deux types de données et seulement les données des expériences est observée. Il n'y a donc pas d'impact de l'utilisation de la combinaison des deux types de données pour cette variable.

De plus, nous observons Figure 17 que, lorsque le BN est paramétré uniquement avec les données spécifiques au cas de M. X et que des quantités  $E_1$ , de 0.8-0.9 ng correspondant au profil ADN du POI, et  $E_2$ , de 0.06-0.07 ng d'ADN différent du profil ADN du POI, sont obtenues, les données qui impactent le plus la valeur des LR sont celles des variables « Quantité d'ADN sur les mains » et « Proportion de transfert ».

En effet, le  $\log_{10}$ (LR) médian du *box plot* « Toutes les variables » est plus proche du  $\log_{10}$ (LR médian) du *box plot* « Quantité d'ADN sur les mains et Proportion de transfert » que celui des autres variables. Le  $\log_{10}$ (LR médian) du *box plot* « Toutes les variables » correspond au résultat obtenu des simulations par ré-échantillonnage lorsque les données spécifiques de l'ensemble des variables sont utilisées pour paramétriser le BN. Le  $\log_{10}$ (LR médian) du *box plot* « Quantité d'ADN sur les mains et Proportion de transfert » correspond au résultat obtenu lorsque les données spécifiques des variables « Quantité d'ADN sur les

---



« mains et Proportion de transfert » sont utilisées, les autres variables étant informées par des données issues de la littérature.

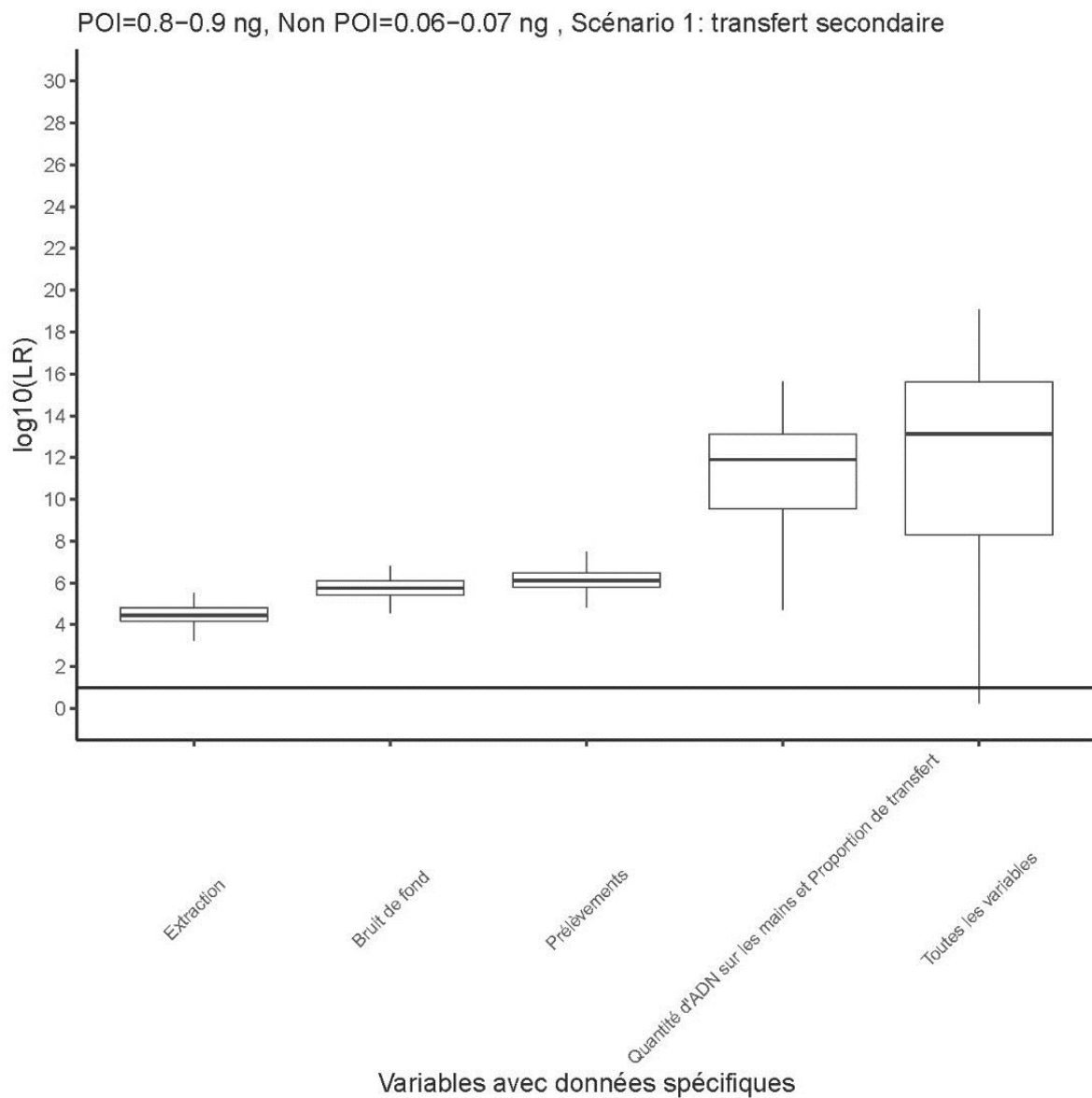


Figure 17 : Variation des  $\log_{10}(\text{LRs})$  en fonction des variables d'intérêt dont les données spécifiques sont ré-échantillonnées par simulation. Les autres variables sont informées par des données issues de la littérature.

#### IV.2.4 Discussion

Les résultats observés sur la Figure 17 indiquent qu'il est recommandé, lors de l'évaluation des résultats ADN d'un cas, de conduire des expériences reproduisant les activités alléguées afin d'obtenir les données correspondant aux **personnes d'intérêt**. En effet, **les variables «Quantité d'ADN sur les mains » et « Proportion de transfert » ont le plus d'impact sur la valeur du LR**. Cette observation confirme le fait que lors des simulations effectuées dans

le chapitre II, le BN étant initialement paramétrisé avec uniquement les données de la littérature, la différence attendue entre les quantités d'ADN sur les mains du POI et de l'AO et des proportions de transfert d'ADN associées à ces personnes n'était pas correctement paramétrées car nous avons adopté une simplification stipulant un comportement analogue entre le POI et l'AO.

Afin d'obtenir les données pour les variables « Quantité d'ADN sur les mains » et « Proportion de Transfert », les données d'efficacité d'extraction et de prélèvement sont utilisées. Pour se faire, il est donc également recommandé, quel que soit le cas, de définir d'une part, l'efficacité d'extraction du kit utilisé en laboratoire, et, d'autre part, l'efficacité des moyens de prélèvement utilisés sur le terrain. Une fois ces conditions définies, les données sont disponibles et peuvent être utilisées de manière générique quel que soit le cas considéré.

Dans le cas d'espèce, les résultats obtenus montrent que pour les variables « Bruit de fond », « Efficacité d'extraction », « Quantité d'ADN sur les mains » et « Proportion de transfert », les données combinées de la littérature et des expériences peuvent être utilisées. En Effet, les résultats montrent que si les données de la littérature ne sont pas adéquates au cas en question, elles constituent cependant une base de connaissance sur ces variables. De plus, l'utilisation de ces données combinées aux données issues des expériences n'a pas d'impact. En revanche, pour la variable « Efficacité de prélèvement », seule les données spécifiques au moyen de prélèvement utilisé doivent servir à paramétrer le BN. En Effet, les résultats montrent que les données de la littérature ne sont pas adéquates au cas en question et que l'utilisation de ces données, combinées aux données issues des expériences, a un impact sur le  $\log_{10}(\text{LR})$ . Cette conclusion est dépendante des données qui ont été obtenues de la littérature et des expériences. Une généralisation de ces résultats à tous les scénarios est impossible.

**En conclusion**, pour la suite de ce travail, les données de la littérature seront combinées aux données issues des expériences pour les variables « Bruit de fond », « Efficacité d'extraction » et « Quantité d'ADN sur les mains » et « Proportion de transfert ». Pour la variable « Efficacité du prélèvement », seules les données issues des expériences seront utilisées.



## Chapitre V- Etude de l'efficacité de la méthodologie

---

### V.1 Introduction

L'objectif premier de ce chapitre est d'analyser l'efficacité de la méthodologie : sachant quelle proposition  $H_p$  ou  $H_{d1}$  est vraie, est-ce que l'application de la méthodologie conduit à des résultats exprimés par un LR qui soutiennent la vraie proposition par rapport à la proposition alternative ?

Le second objectif vise à améliorer le résultat obtenu par le BN. Pour cela, l'efficacité de la méthodologie est étudiée en fonction de trois LR :

- Le LR dit « initial » qui est le LR obtenu par le BN.
- Le LR dit « calibré » qui est le LR obtenu lorsque le LR initial est ajusté à partir des LR des cas connus (cas observés dont on connaît le résultat) à l'aide d'un modèle de régression logistique (Morisson, 2013).
- Le LR dit « reporté ». Un LR reporté est un LR considéré comme robuste, c'est à dire un LR dans lequel nous accordons notre confiance en se basant sur des critères qui seront alors définis dans ce chapitre.

Pour évaluer l'efficacité de la méthodologie, le cas de M. X a empiriquement été reproduit 120 fois (nous obtenons 120 cas dit « cas simulés ») en suivant l'une ou l'autre des propositions suivantes :

- Hypothèse  $H_p$  : « Le POI a poignardé la victime ».
- Hypothèse  $H_{d1}$  : « Le POI a serré la main de l'AO qui a poignardé la victime avec le couteau retrouvé sur les lieux ».

Les résultats des prélèvements effectués sur le manche du couteau sont évalués suivant la méthodologie élaborée au travers de ce travail. Elle comprend en l'état les étapes suivantes :

- Construction d'un BN informé par les circonstances du cas, pour les variables instanciées, et des données issues de la littérature, pour les variables non instanciées (Chapitre II).
- Identification des variables d'intérêt : « Quantité d'ADN sur les mains », « Proportion de transfert », « Efficacité d'extraction », « Efficacité de prélèvement » et « Bruit de fond » et dont une acquisition de données serait nécessaire (Chapitre III).

- Acquisition de données spécifiques au cas de M. X pour les variables « Quantité d'ADN sur les mains », « Proportion de contact », « Efficacité d'extraction », « Efficacité de prélèvement » et « Bruit de fond » (Chapitre IV).

Enfin, les résultats des évaluations de chacun des cas simulés sont étudiés afin de déterminer dans quelle mesure les observations soutiennent la proposition qui a été produite, par rapport à la proposition alternative. Ces résultats permettront également d'effectuer une calibration de la méthode, de poser un critère définissant la robustesse d'un LR et de se questionner sur le report ou non du LR dans un rapport d'expertise.

## V.2 Méthode

### V.2.1 Obtention des rapports de vraisemblance dits « initiaux »

Nous avons établi dans le chapitre IV, sur l'étude de la quantité d'ADN sur les mains et la proportion de transfert, que le participant 6 est en moyenne meilleur donneur que le participant 2. Ces deux participants constituent les protagonistes, POI ou AO, utilisés dans la production des cas empiriques.

Deux situations ont été étudiées : la première lorsque le POI est le participant 2 et l'AO le participant 6 (le POI est généralement moins bon donneur que l'AO) et la deuxième lorsque le POI est le participant 6 et l'AO le participant 2 (le POI est généralement meilleur donneur que l'AO).

Les expériences réalisées dans le chapitre IV afin de déterminer la proportion des différents contributeurs lorsque que le POI poignarde le savon balistique ou serre la main de l'AO constituent 111 cas simulés dans des circonstances connues, soit :

- 53 cas simulés pour lesquels la proposition du procureur est reproduite dont 30 pour lesquels le POI est le participant 6 et 23 cas simulés pour lesquels le POI est le participant 2.
- 58 cas simulés pour lesquels la proposition de la défense est reproduite dont 30 pour lesquels le POI est le participant 6 et 28 cas simulés pour lesquels le POI est le participant 2.

Cent vingt cas simulés ont été initialement produits. Pour chacun des cas, un prélèvement est effectué sur le manche du couteau en utilisant le FLOQSwab™. L'ADN est ensuite extrait avec le kit QIAshredder-QIAamp Mini DNA en utilisant les colonnes Microcon® 30 puis

quantifié en utilisant le kit Investigator® Quantiplex de Qiagen sur le Rotor-Gene® Q. Les profils sont établis à partir d'une électrophorèse capillaire effectuée sur l'instrument 3500 (Applied Biosystems) et une analyse par le logiciel Genemapper ID-X (ThermoFisher Scientific). Pour chaque expérience, la proportion des différents contributeurs est calculée en utilisant le logiciel STRmix®. Pour 9 cas simulés, les mélanges d'ADN étaient trop complexes pour que les proportions d'ADN de chaque contributeur puissent être déterminées avec le logiciel STRmix®. Ces 9 cas simulés n'ont donc pas été évalués.

Nous avons identifié, dans le chapitre III, les variables « Efficacité d'extraction », « Efficacité de prélèvement », « Quantités sur les mains », « Bruit de fond » « Proportion de transfert », comme étant des variables à fort impact pour la plupart des observations. Nous avons aussi montré que pour la plupart des observations, il était conseillé d'acquérir des données spécifiques au cas pour les variables « Quantités sur les mains » et « Proportion de transfert ». Il est également préférable d'utiliser des données spécifiques au kit d'extraction et à la méthode de prélèvement utilisée.

Les résultats ADN sont évalués au moyen du BN informé avec les données spécifiques au cas (variables non instanciées) et aux circonstances du cas (variables instanciées), combinées avec les données de la littérature pour les variables « Bruit de fond », « Efficacité d'extraction » et « Quantité d'ADN sur les mains » et « Proportion de transfert », suite aux résultats du chapitre IV. Un LR dit « initial » est alors établi pour chacun des cas.

Le ratio entre  $E_2$  (Quantité d'ADN prélevée différent du profil ADN du POI) et  $E_1$  (Quantité d'ADN prélevée correspondant au profil ADN du POI) a été calculé. Une quantité de 0,001 ng a été utilisée pour calculer le ratio lorsqu'une quantité de 0 était obtenue. Ce ratio est utilisé afin d'étudier si les cas pour lesquels les observations ne soutiendraient pas la « bonne proposition » par rapport à l'alternative correspondraient à des situations avec des ratios particuliers.

#### V.2.2 Obtention des rapports de vraisemblance dits « calibrés »

Une calibration du système est effectuée à partir de cas dont la proposition est connue, c'est-à-dire que nous savons quelle est la « bonne » proposition. La calibration du système permet de déterminer les LRs calibrés d'après les LRs initiaux obtenus à partir des 111 cas dont la « bonne » proposition est connue. Le système est dit « calibré » si, par exemple, un LR de 10 est obtenu aussi souvent sous  $H_p$  que le LR de 0.1 sous  $H_d$  et si l'observation  $E_i = \{E_1, E_2\}$  du

cas en question a effectivement été observée 10 fois plus sous Hp que sous Hd dans les expériences.

La méthode de calibration choisie est la régression logistique (Morisson, 2013). La calibration par régression logistique présente de nombreux avantages, selon De Donno (2016), notamment par rapport à la « Pool Adjacent Violators Algorithm » (Ayeret al., 1995) :

- Elle ne produit pas de rapports de vraisemblance calibrés infinis ou nuls.
- Elle n'altère pas la structure des données et la construction du modèle probabiliste.
- Elle permet une estimation des LR calibrés en dehors de l'intervalle des LR initiaux qui sont les LR observés lors de cette étude de l'efficacité de la méthode.

Le LR calibré est calculé à partir du LR initial selon la formule suivante :

$$\log_{10}(LR)_{\text{calibré}} = a \log_{10}(LR)_{\text{initial}} + b$$

Les paramètres de calibration  $a$  et  $b$  sont estimés à partir des LR obtenus par le réseau bayésien pour des cas dont on connaît la nature des propositions. Sachant quelle proposition est vraie, il est possible d'ajuster les paramètres  $a$  et  $b$  afin de minimiser les erreurs et leurs coûts.

Le ratio entre  $E_2$  (Quantité d'ADN prélevé différent du profil ADN du POI) et  $E_1$  (Quantité d'ADN prélevé correspondant au profil ADN du POI) a été calculé.

### V.2.3 Obtention des rapports de vraisemblance dits « reportés »

Nous pourrions reporter le LR calibré tel quel. Cependant, la question qui se pose est la suivante : quelle est la confiance que nous accordons aux données utilisées et donc au LR obtenu ? Si nous n'avons pas confiance dans ce LR, la question du LR reporté dans un rapport se pose alors.

Cette confiance dépend des résultats obtenus si un autre set de 30 expériences, dans les mêmes conditions et avec les mêmes participants, avait été fait. Théoriquement, si le LR obtenu avec 30 résultats d'expériences est robuste, nous nous attendons à un LR, obtenu avec 30 autres résultats de la même expérience, d'un même ordre de grandeur.

Des simulations ont donc été réalisées, basées sur la même méthode qu'au chapitre IV, c'est-à-dire sur des ré-échantillonnages avec remplacement d'un nombre identique à celui des données initiales, à partir des données initiales informant le BN. Les données initiales sont la

combinaison des données issues des expériences et de la littérature pour les variables « Bruit de fond », « Efficacité d'extraction » et « Quantité d'ADN sur les mains » et « Proportion de transfert » et uniquement des données issues des expériences pour la variable « Efficacité de prélèvement ». Les LR ont ensuite été calibrés suivant une régression logistique dont les paramètres ont été définis précédemment.

Un critère de robustesse du LR permettant de définir cette confiance accordée au résultat a été établi sur la base des résultats obtenus suite à ces simulations par ré-échantillonnage. Une discussion sur ce qu'il conviendrait de faire concernant la transmission du LR ou non dans un rapport en se basant sur ce critère sera présentée dans ce chapitre (V.3.3).

Les résultats de l'efficacité de la méthodologie à partir des LR obtenus avant la calibration (LR initiaux), après la calibration (LR calibrés) et après application de ce critère de robustesse (LR reportés) seront présentés ci-après.

## V.3 Résultat

### V.3.1 Rapports de vraisemblance initiaux

Le Tableau 5 présente le nombre de cas simulés où les observations  $E_i = \{E_1, E_2\}$  soutiennent ou non la « bonne » proposition. La « bonne » proposition est celle qui était empiriquement reproduite par les protagonistes. Les observations  $E_i = \{E_1, E_2\}$ , les quantifications, les proportions des différents contributeurs et les LR obtenus sur ces 111 cas sont présentés en Annexe VII.1.

**Tableau 5 : Nombre de cas pour lesquels les observations soutiennent la bonne ou la mauvaise proposition en fonction de la capacité du POI à transférer son ADN par rapport à l'AO.**

POI	« Bonne » Proposition	Observations soutiennent la « bonne » proposition	Observations soutiennent la « mauvaise » proposition
Meilleur donneur que AO	Procureur	30	0
Moins bon donneur que AO	Procureur	22	1
Meilleur donneur que AO	Défense	12	16
Moins bon donneur que AO	Défense	29	1



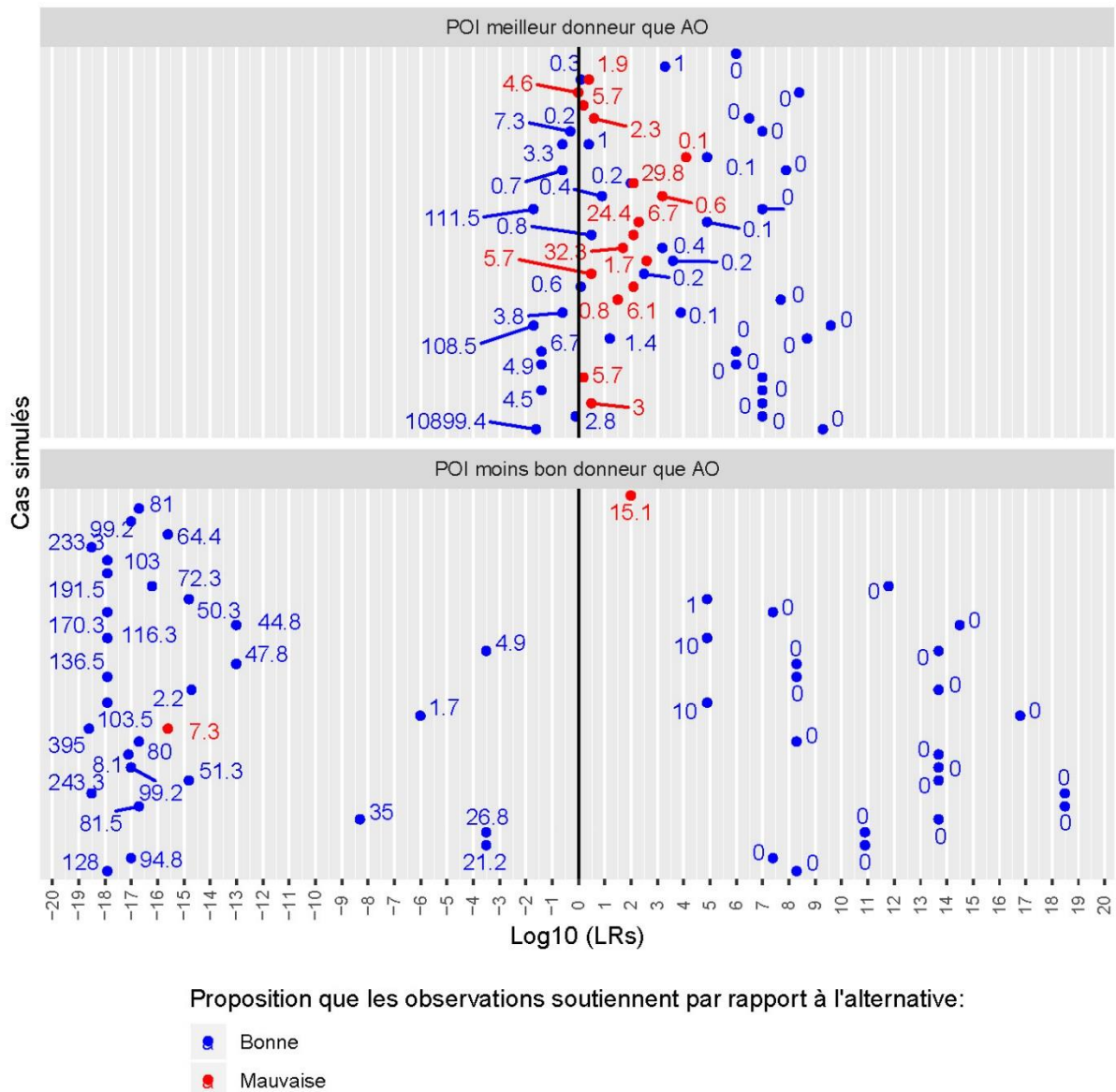


Figure 18 : Points illustrant le  $\log_{10}(LRs)$  obtenu pour chacun des cas simulés lorsque le POI est meilleur ou moins bon donneur que l'AO, associé au ratio  $E_2/E_1$ , indiqués en bleu si les observations soutiennent la bonne proposition et en rouge si les observations soutiennent la mauvaise proposition.

Le Tableau 5 et la Figure 18 montrent que lorsque la proposition du procureur est vraie, les observations soutiennent effectivement la proposition du procureur par rapport à celle de la défense dans 52 cas sur 53 soit dans 98% des cas. Dans ces cas-là, les ratios des quantités d'ADN prélevés  $E_2/E_1$  sont majoritairement proches de 0 (La valeur de 0 correspond au résultat obtenu arrondi au dixième). En revanche, les observations soutiennent de manière erronée la proposition de la défense par rapport à celle du procureur dans 1 cas sur 53 soit dans 2% seulement et le ratio des quantités d'ADN est de 7,3.

Lorsque la proposition de la défense est vraie, les observations soutiennent effectivement la proposition de la défense par rapport à celle du procureur dans 41 cas sur 58 soit dans 71%

des cas. En revanche, les observations soutiennent la proposition du procureur par rapport à celle de la défense dans 17 cas sur 58 soit dans 29%. Dans l'ensemble de ces cas-là, le ratio est plus grand que 0,1 (cf. labels sur les points rouges avec des  $\log_{10}(\text{LRs})$  supérieurs à 1 sur la Figure 18).

Le LR le plus bas transmis, alors que les observations soutiennent la mauvaise proposition, est de l'ordre de  $3^{E-16}$ , le plus haut transmis est de l'ordre de  $1^{E4}$  (cf. Annexe VII.1).

Nous pouvons ainsi dire, à partir de nos observations, que si un LR compris entre 1 et de l'ordre de  $1^{E4}$  est obtenu et que le ratio est plus grand que 0.1, le risque est plus grand que les observations soutiennent la mauvaise proposition par rapport à l'alternative.

Il n'est cependant pas possible de définir une zone de risque si un LR inférieur à 1 est obtenu, la valeur du LR étant très faible et le ratio ne se démarquant pas des ratios obtenus pour le cas dont les observations soutiennent la bonne proposition par rapport à l'alternative.

Ces résultats diffèrent selon le statut relatif des donneurs, en particulier si la proposition de la défense est vraie. En effet, si le POI est en moyenne meilleur donneur que l'AO, les observations soutiennent effectivement la proposition de la défense dans 29 cas sur 30 soit dans 96% des cas. En revanche, si le POI est en moyenne moins bon donneur que l'AO, les observations soutiennent la proposition de la défense dans 12 cas sur 28 soit seulement 43% des cas.

Nous remarquons que les valeurs de LR peuvent être extrêmement faibles ou hautes. Ceci sera discuté dans la partie V.4.1.

### V.3.2 Rapports de vraisemblance calibrés

Une calibration des LR a été effectuée d'après les LR initiaux obtenus à partir de cas dont la bonne proposition est connue.

Les paramètres de calibration obtenus sont les suivants :

- $a=0.20$
- $b=-0.41$

Le Tableau 6 présente le nombre de cas simulés où les observations  $E_i=\{E_1, E_2\}$  soutiennent ou non la « bonne » proposition avec les LR calibrés obtenus. Les observations  $E_i=\{E_1, E_2\}$ , les quantifications, les proportions des différents contributeurs et les LR calibrés obtenus sur ces 111 cas sont également présentés en Annexe VII.1.

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

Tableau 6 : Nombre de cas pour lesquels les observations soutiennent la bonne ou la mauvaise proposition en fonction de la capacité du POI à transférer son ADN par rapport à l'AO d'après les LR calibrés.

POI	« Bonne » Proposition	Observations soutiennent la « bonne » proposition	Observations soutiennent la « mauvaise » proposition
Meilleur donneur que AO	Procureur	23	7
Moins bon donneur que AO	Procureur	22	1
Meilleur donneur que AO	Défense	21	7
Moins bon donneur que AO	Défense	30	0

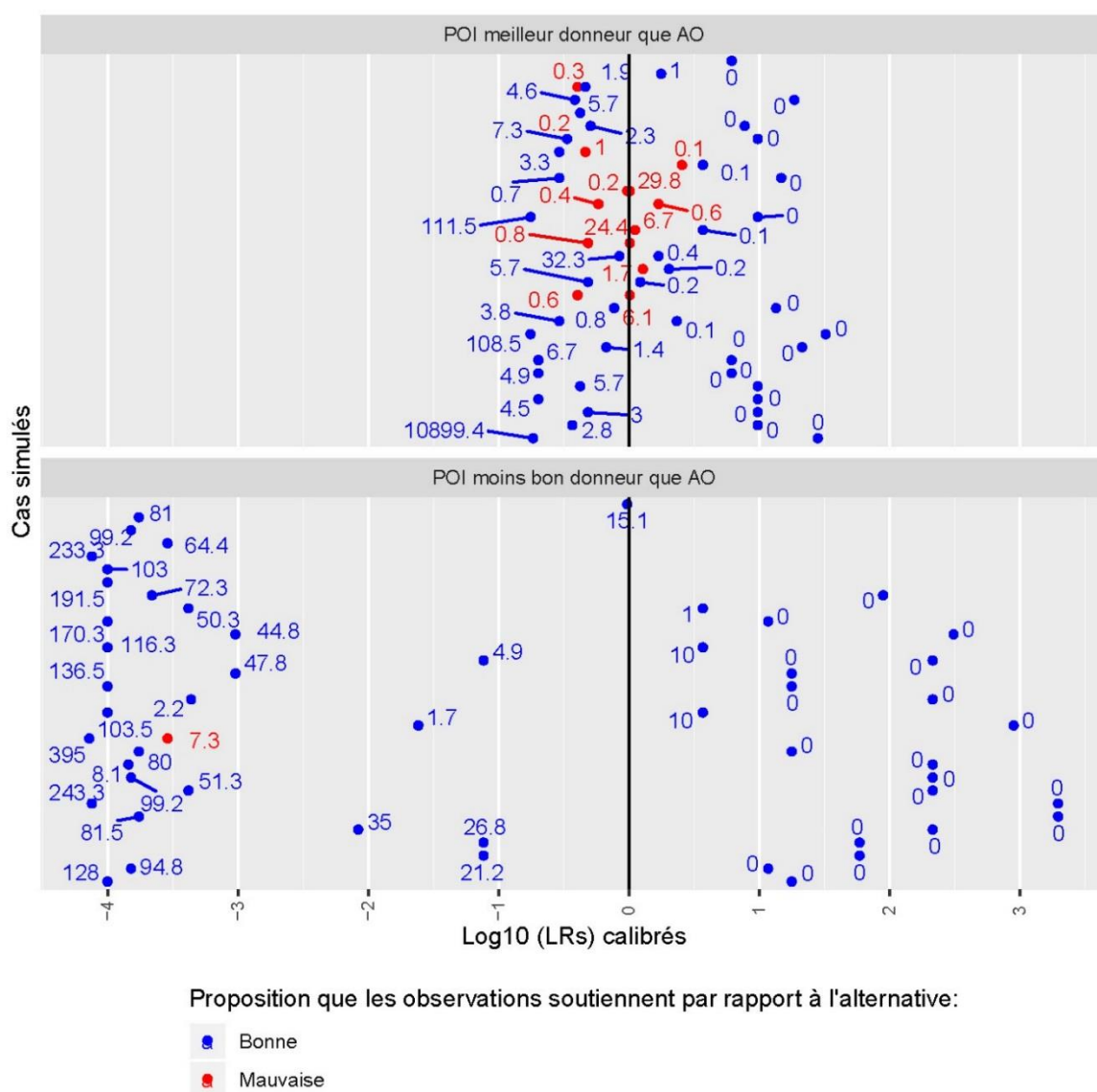


Figure 19 : Points illustrant le  $\log_{10}(\text{LRs})$  obtenu pour chacun des cas simulés lorsque le POI est meilleur ou moins bon donneur que l'AO, associé au ratio  $E_2/E_1$ , indiqués en bleu si les observations soutiennent la bonne proposition et en rouge si les observations soutiennent la mauvaise proposition d'après les LR calibrés.

Le Tableau 6 et la Figure 19 montrent que lorsque la proposition du procureur est vraie, les observations soutiennent effectivement la proposition du procureur par rapport à celle de la défense dans 45 cas sur 53 soit dans 85% des cas. Dans ces cas-là, les ratios des quantités d'ADN sont majoritairement proches de 0. (La valeur de 0 correspond au résultat obtenu arrondi au dixième). En revanche, les observations soutiennent la proposition de la défense par rapport à celle du procureur dans 8 cas sur 53 soit dans 15%, ce qui est plus élevé qu'avant la calibration du système. Les ratios des quantités d'ADN varient de 0,2 à 7,3.

Lorsque la proposition de la défense est vraie, les observations soutiennent effectivement la proposition de la défense par rapport à celle du procureur dans 51 cas sur 58 soit dans 88% des cas. Les observations soutiennent la proposition du procureur par rapport à celle de la défense dans seulement 7 cas sur 58 soit dans seulement 12% des cas, au lieu de 29% lorsque le système n'est pas calibré. Le ratio est plus grand que 0,1 pour cette situation.

Le LR calibré le plus bas transmis alors que les observations soutiennent la mauvaise proposition est de l'ordre de  $3^{E-4}$ , le plus haut transmis est de l'ordre de 2.

Nous pouvons également toujours dire que, selon nos observations, si un LR compris entre 0.4 et 2 est obtenu et que le ratio est plus grand que 0.1, le risque est plus grand que les observations soutiennent la mauvaise proposition par rapport à l'alternative.

Il n'est par contre pas possible de définir une zone de risque si un LR inférieur à 0,4 est obtenu, la valeur du LR étant très faible et le ratio ne se démarquant pas des ratios obtenus pour le cas dont les observations soutiennent la bonne proposition par rapport à l'alternative. Un seul cas a été observé.

Ces résultats montrent une amélioration de l'efficacité de la méthodologie après calibration, excepté si la proposition du procureur est vraie et que le POI est meilleur donneur que l'AO.

### V.3.3 Rapports de vraisemblance reportés

#### Robustesse du LR

Les résultats de la variation des  $\log_{10}$ (LRs calibrés) obtenus par ré-échantillonnage des données peuvent être visualisés en sélectionnant « Part 3 : Results of cases » sur le site [https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/).

La Figure 20 montre le résultat, lorsqu'une quantité d'ADN entre 0.1 et 0.2 ng correspondant au profil ADN du POI ( $E_1$ ) et une quantité d'ADN entre 0.01 et 0.02 ng différent du profil ADN du POI ( $E_2$ ) sont obtenues (Cas de M. X).

---

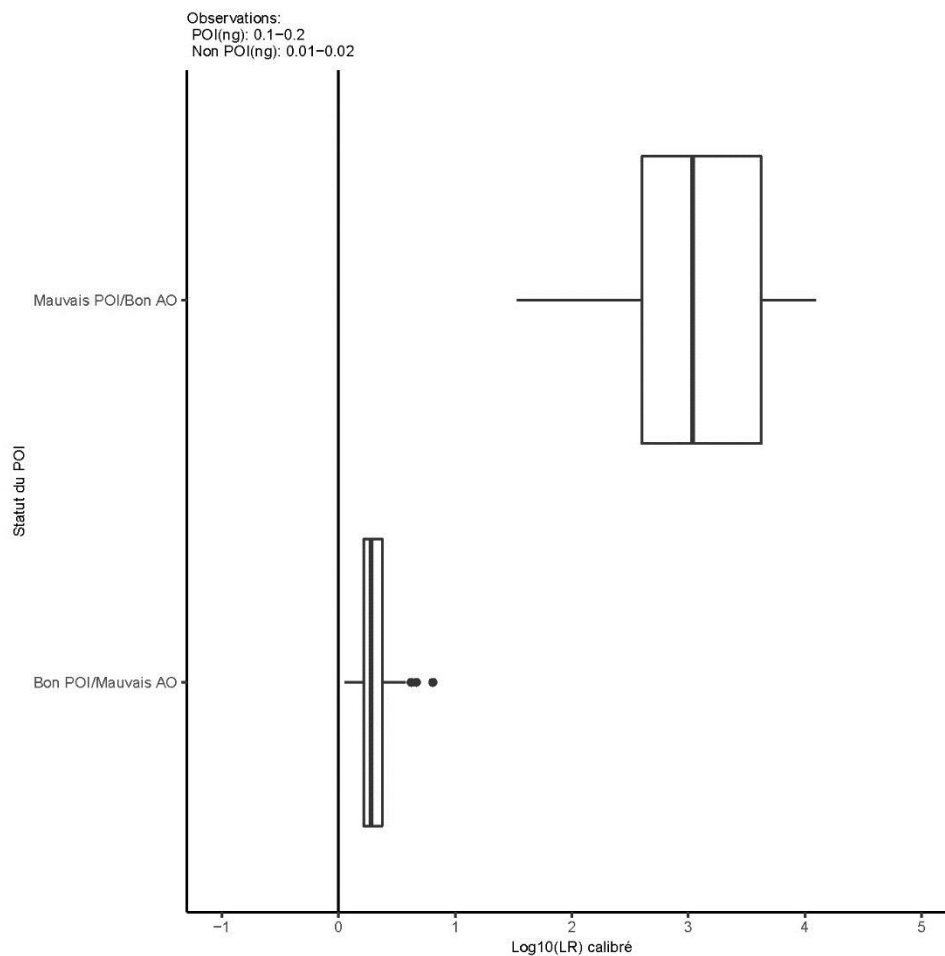


Figure 20 : Variation, obtenue par simulation, des  $\log_{10}(\text{LRs})$  lorsqu'une quantité d'ADN entre 0.1 et 0.2 ng correspondant au profil ADN du POI ( $E_1$ ) et une quantité d'ADN entre 0.01 et 0.02 ng différent du profil ADN du POI ( $E_2$ ) sont obtenues, en fonction de la capacité du POI ( $M.X$ ) à transférer son ADN par rapport à l'AO.

Si le POI est moins bon donneur que l'AO, le LR calibré minimum est proche de 80 alors que le LR calibré maximum est proche de 10'000. Une grande différence est observée entre les LR calibrés minimum et maximum. Dans cette situation, cela signifie que les observations obtenues avec un jeu de données soutiennent très fortement  $H_p$  par rapport à la  $H_{d1}$  alors qu'avec un autre jeu de données, ces mêmes observations soutiennent modérément  $H_p$  par rapport à  $H_{d1}$  (selon l'échelle verbale proposée en exemple par Willis et al., 2015 dans le guide de l'ENFSI). Or, ces jeux de données, de par la méthode de ré-échantillonnage, ne sont pas drastiquement différents. Au vu de ce résultat, il est légitime de remettre en question la confiance dans le LR calibré obtenu avec les données de la paramétrisation.

En revanche, si le POI est meilleur donneur que l'AO, une différence d'ordre 1 seulement est observée. Au vu de ce résultat, nous sommes d'avis que nous pouvons avoir confiance dans le LR calibré obtenu avec les données de la paramétrisation.

Ainsi, la confiance accordée au LR dépend de la variation des LR obtenus si un autre jeu de 30 expériences avait été fait. Ici, cet autre jeu de données a été simulé par ré-échantillonnage des données initiales avec remises. L'étendue de la variation (LRmax-LRmin) dépend essentiellement du nombre de simulations effectuées. Plus le nombre de simulations est grand, plus des valeurs extrêmes risquent d'être obtenues et celles-ci ne refléteraient pas correctement la réalité. En revanche observer l'étendue interquartile limite cet effet.

De plus, l'observation d'une variation des  $\log_{10}(\text{LRs})$  d'un ordre de 3 autour d'un  $\log_{10}(\text{LR})$  médian de 1, par exemple, n'a pas la même signification sur la robustesse du LR que si cette variation d'ordre de 3 était observée autour d'un  $\log_{10}(\text{LR})$  de 9. Dans le premier cas, nous ne serions pas confiant dans la valeur du LR à transmettre. En revanche dans le deuxième, nous le serions. Afin de tenir compte de cet aspect de la robustesse qui dépendrait également de la valeur du LR, le paramètre jugé approprié pour évaluer la robustesse du LR obtenu est l'étendue interquartile des  $\log_{10}(\text{LRs})$  par rapport au  $\log_{10}(\text{LR})$  médian).

Deux conditions, selon nous, peuvent définir un LR comme non robuste :

- une variation de l'étendue interquartile trop grande par rapport au LR médian obtenu,
- un grand nombre de LR signifiant que les observations soutiennent des propositions différentes par rapport au LR médian obtenu.

Le seuil de l'étendue interquartile par rapport au LR calibré médian (en valeur absolue) permettant de correspondre au mieux à ces deux conditions a été fixé à 0.8, basé sur l'observation des résultats qui peuvent être visualisés en sélectionnant « Part 3 : Results of cases » sur le site [https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/). Cette notion de robustesse dépend ici de notre impression personnelle, qui est subjective. Pour chaque observation  $E_i = \{E_1, E_2\}$ , nous avons déterminé si nous aurions confiance dans ce LR. Le seuil a alors été fixé sur la base de ces résultats.

Nous avons observé que, lorsque l'ensemble des LR ainsi que le LR médian sont proches de 1, malgré une variation très petite, l'étendue interquartile par rapport au LR calibré médian (en valeur absolue) peut être supérieure à 1 et être défini, à tort, comme non robuste.

Un LR robuste, c'est-à-dire dans lequel nous aurions confiance, serait donc un LR dont l'étendue interquartile des  $\log_{10}(\text{LRs})$  obtenus par un ré-échantillonnage des données initiales par rapport au  $\log_{10}(\text{LR})$  médian (en valeur absolue) n'excéderait pas une valeur seuil de 0.8 ou que le  $\log_{10}(\text{LR})$  médian est de l'ordre de 0.

Si un LR robuste (comme définie précédemment) est obtenu avec le jeu de données initiales, le LR peut être transmis dans un rapport d'expertise. Il est à rappeler qu'il n'est, ici, pas question de transmettre l'intervalle des LR comme résultats ou d'en écrire la variation et ceci pour deux raisons principales :

- Les LR sont obtenus avec des jeux de données différents. Or l'évaluation du cas ne devrait se faire que sur le jeu de données utilisées à l'évaluation.
- Transmettre un intervalle est fortement déconseillé par plusieurs auteurs (Taylor et al., 2016, Taroni et al., 2016)

En revanche, un LR non robuste, c'est-à-dire dans lequel nous n'aurions pas confiance, signifient alors que l'étendue interquartile des  $\log_{10}(\text{LRs})$  obtenus par un ré-échantillonnage des données par rapport au  $\log_{10}(\text{LR})$  médian (en valeur absolue) obtenu est supérieure à la valeur seuil de 0.8. Si tel est le cas, cela peut s'expliquer par une quantité de données insuffisantes, en partant du présupposé qu'elles étaient adéquates. L'utilité d'acquérir des données supplémentaires devrait donc être investigué, comme vu dans la partie III.2. Cependant, ceci nécessiterait peut-être l'acquisition d'un tel nombre de données qu'il ne serait pas possible de l'effectuer.

Si un LR non robuste est obtenu, la question du résultat à transmettre dans un rapport d'expertise se pose.

Trois options sont proposées dans ce travail :

- transmettre un LR de 1,
- transmettre le fait que les experts ont obtenus un résultat non robuste (dans lequel nous n'aurions pas confiance) qui ne permet pas de se prononcer sur les observations,
- ne pas reporter de LR, et indiquer qu'il n'est pas possible de se prononcer sur les observations.

Un LR de 1 signifie que les observations soutiennent autant les deux propositions. Ceci devrait, selon nous, être à différencier de la situation dans laquelle le LR obtenu n'est pas robuste.

Transmettre un résultat non robuste est risqué car il pourrait être mal compris. Transmettre ce LR implique potentiellement que nous avons une certaine confiance ce qui est alors contradictoire avec le résultat obtenu.

La troisième option est donc privilégiée.

---

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

Le Tableau 7 présente le nombre de cas sur les 111 pour lesquels un LR calibré robuste est obtenu signifiant d'une part que les observations soutiennent la « bonne » proposition par rapport à l'alternative, le nombre de cas pour lesquels un LR calibré robuste est obtenu signifiant d'autre part que les observations soutiennent la « mauvaise » proposition par rapport à l'alternative et enfin le nombre de cas pour lesquels un LR calibré non robuste est obtenu.

**Tableau 7 : Nombre de cas pour lesquels un LR calibré robuste (qui signifie que les observations soutiennent la bonne ou la mauvaise proposition) ou non est obtenu, en fonction de la capacité du POI à transférer son ADN par rapport à l'AO.**

POI	« Bonne » Proposition	LR robuste Observations soutenant la « bonne » proposition	LR robuste Observations soutenant la « mauvaise » proposition	LR non robuste
Meilleur donneur que AO	Procureur	21	5	4
Moins bon donneur que AO	Procureur	22	0	1
Meilleur donneur que AO	Défense	21	5	2
Moins bon donneur que AO	Défense	1	0	29



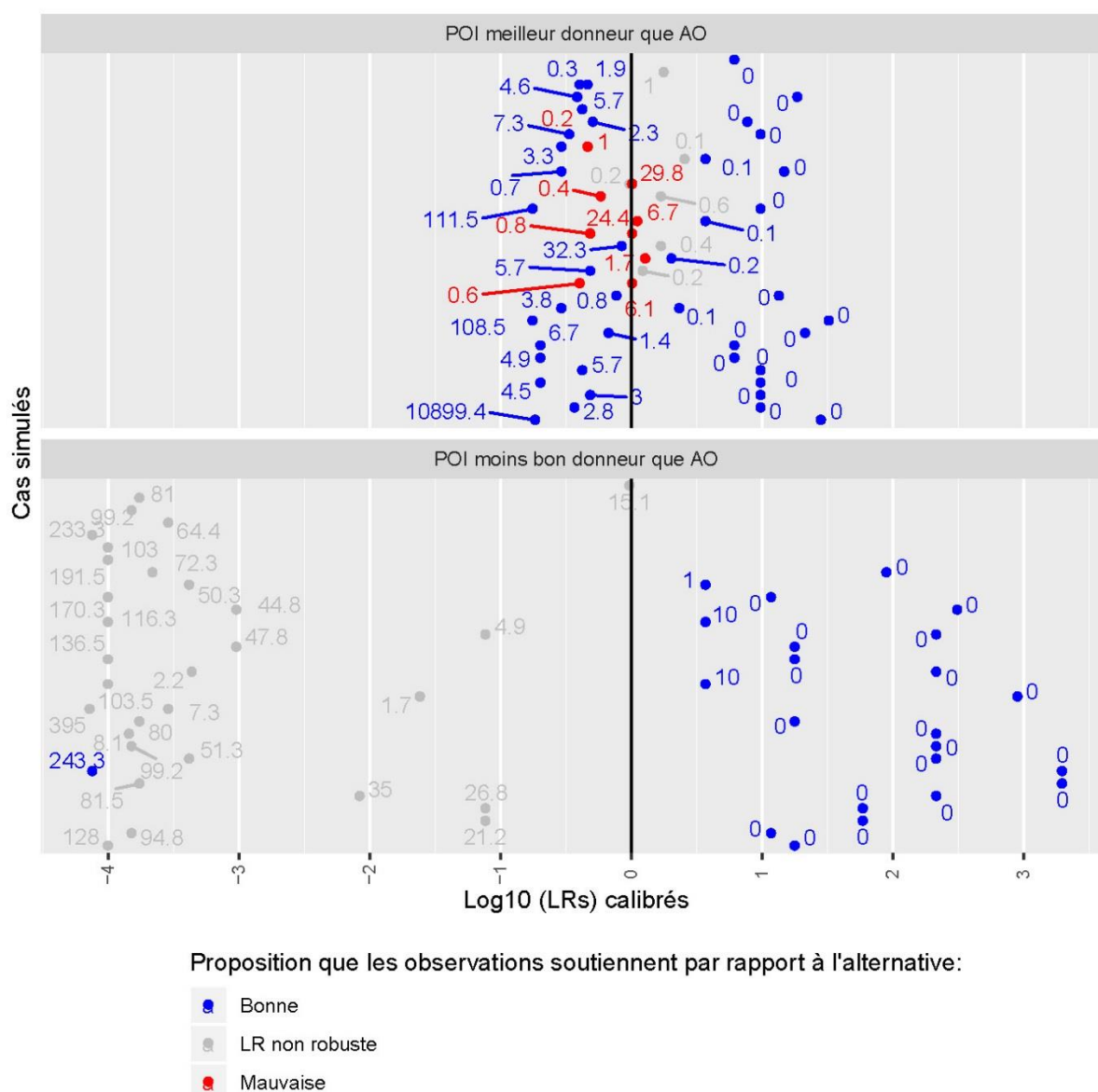


Figure 21 : Points illustrant le  $\log_{10}(\text{LR})$  calibré obtenu pour chacun des cas simulés lorsque le POI est meilleur ou moins bon donneur que l'AO, associé au ratio  $E_2/E_1$ , indiqués en bleu si les observations soutiennent la bonne proposition, en rouge si les observations soutiennent la mauvaise proposition et en gris s'il n'est pas robuste.

Le Tableau 7 et la Figure 21 montrent que la prise en compte du critère de robustesse permet de ne transmettre que 10 cas au lieu des 15 pour lesquels les observations soutiennent la mauvaise proposition par rapport à l'autre.

Si un LR robuste est supérieur à 1, le risque que les observations soutiennent la mauvaise proposition est de 5 sur 48 soit 10% des cas. Pour ces 5 cas, les LR sont légèrement supérieurs à 1. Si un LR robuste est inférieur à 1, le risque que les observations soutiennent la mauvaise proposition est de 5 sur 27 soit 18% des cas. Le LR le plus bas transmis alors que

les observations soutiennent la mauvaise proposition est de 0.4, le plus haut transmis est seulement de l'ordre de 1.

Dans 36 cas sur 111 soit 32% des cas, un LR non robuste est obtenu et ne serait donc pas reporté. Or dans ces 36 cas, un LR signifiant que les observations soutenaient la bonne proposition avait été obtenu dans 31 cas, en particulier si la proposition de la défense est vraie et que le POI est moins donneur que l'AO.

La plupart des cas non reportés (31 sur 36) sont des cas pour lesquels la proposition de la défense est la bonne. De plus, sur ces 31 cas, 29 cas étaient des cas pour lesquels un LR signifiant que les observations soutenaient la bonne proposition avait été obtenu. Il existe donc un biais en faveur de l'accusation. Ce résultat signifie que si la proposition de la défense est vraie, il faudrait mener plus d'expériences pour obtenir un LR robuste que si la proposition du procureur est vraie. Nous n'arrivons pas à expliquer cette observation.

#### Zone de risque

Nous avons observé dans la partie précédente (V.3.2) qu'un LR compris entre 0.4 et 2 obtenu avec un ratio  $E_2/E_1$  plus grand que 0.1 conduisait à un risque plus important que les observations soutiennent la mauvaise proposition par rapport à l'alternative. Par la suite, nous définissons la zone de risque par l'obtention d'un LR compris entre 0.4 et 2 avec un ratio  $E_2/E_1$  supérieur à 0.1.

La question de comment tenir compte de cette information dans la transmission des résultats se pose. Deux options sont proposées dans ce travail :

- Option 1 : Définir que le LR est reporté seulement s'il est robuste et qu'il n'est pas compris dans la zone de risque.
- Option 2 : Définir que le LR est reporté seulement s'il est robuste, en précisant la probabilité que les observations obtenues soutiennent la mauvaise proposition par rapport à l'alternative.

L'option 2 est préférée car si le LR est robuste, aucune donnée supplémentaire ne permettra d'affiner le résultat obtenu. Nous pensons donc que si le LR est robuste, il doit être reporté en précisant le risque que les observations soutiennent la mauvaise proposition. Les résultats obtenus pour chacune de ces deux options, sont néanmoins détaillés ci-après.

*Option 1.*

Le Tableau 8 montre les résultats obtenus lorsque le LR est reporté ou non en prenant en compte cette zone de risque (en noir) ou non (en rouge).

**Tableau 8 : Nombre de cas où un LR calibré robuste (qui signifie que les observations soutiennent la bonne ou la mauvaise proposition) ou non est obtenu, en fonction de la capacité du POI à transférer son ADN par rapport à l'AO, en tenant compte du risque (en noir) ou non (en rouge). Le risque est défini en fonction de la valeur du LR et du ratio  $E_2/E_1$ .**

POI	« Bonne » Proposition	LR reporté		LR non reporté
		Observations soutenant la « bonne » proposition	Observations soutenant la « mauvaise » proposition	
Meilleur donneur que AO	Procureur	21-20	5-0	4-10
Moins bon donneur que AO	Procureur	22-22	0-0	1-1
Meilleur donneur que AO	Défense	22-13	5-0	2-15
Moins bon donneur que AO	Défense	1-1	0-0	29-29

Les 10 cas pour lesquels les LR étaient reportés et les observations soutiennent la mauvaise proposition sont ceux pour lesquels les LR ne seront pas reportés en prenant en compte cette zone de risque. Ce résultat était attendu. Aucun point rouge indiquant que le LR peut être reporté et que les observations soutiennent la mauvaise proposition n'est alors observé (Figure 22).

Cependant, en prenant en compte la zone de risque, les LR de 9 cas pour lesquels les observations soutiennent la bonne proposition ne sont plus reportés.

Ceci signifie que pour 55 cas sur 111, soit 49% des cas, le LR n'est pas reporté si nous prenons en compte la zone de risque. Or les observations soutiennent la bonne proposition pour 40 cas sur ces 55.

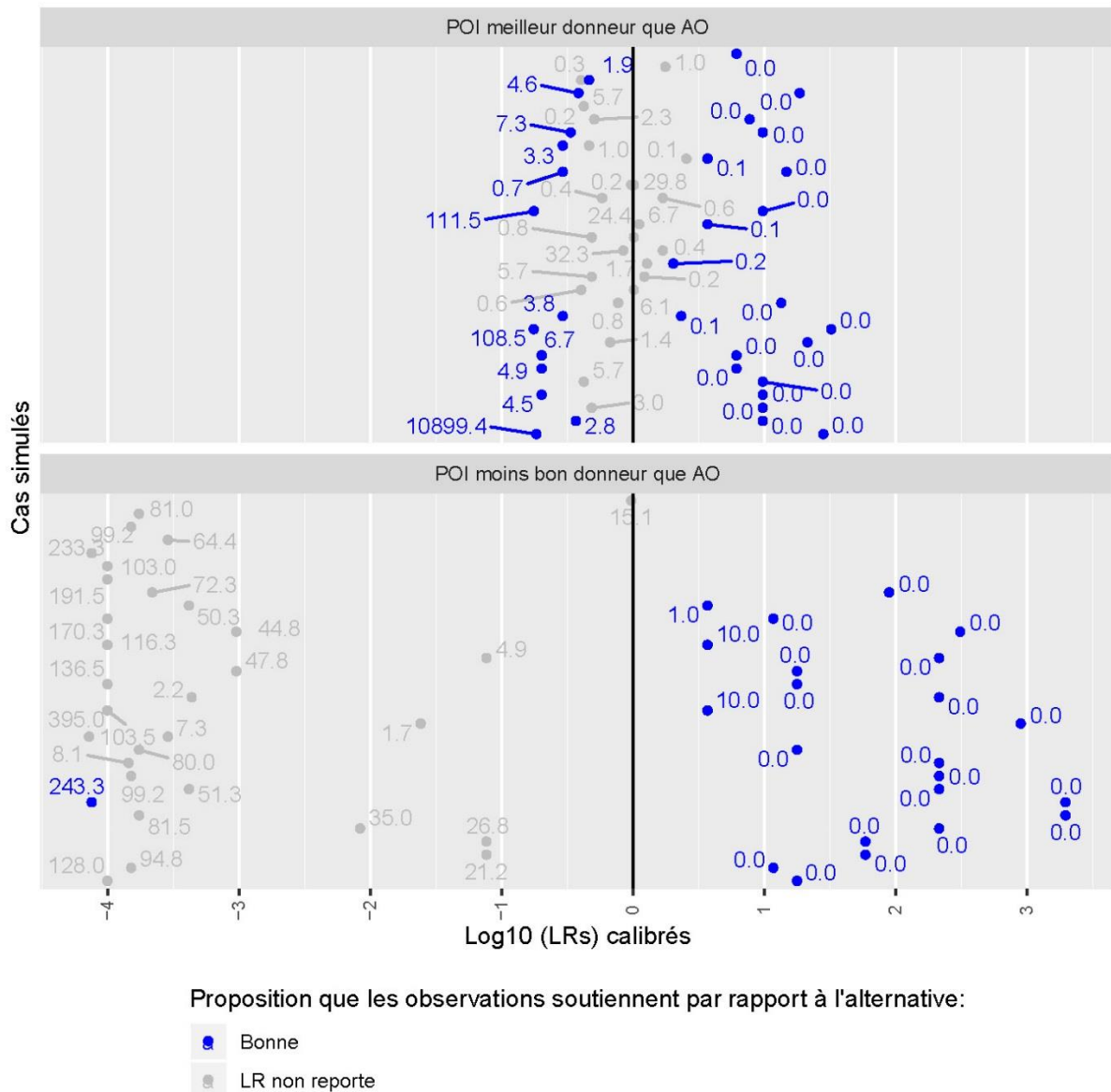


Figure 22 : Points illustrant le  $\log_{10}(\text{LR})$  calibré obtenu pour chacun des cas simulés lorsque le POI est meilleur ou moins bon donneur que l'AO, associé au ratio  $E_2/E_1$ , indiqués en bleu si le LR peut être reporté et que les observations soutiennent la bonne proposition, en rouge si le LR peut être reporté et que les observations soutiennent la mauvaise proposition et en gris si le LR ne peut pas être reporté.

### Option 2.

Pour cette option, le LR robuste est reporté et les informations quant au risque que les observations soutiennent la mauvaise proposition sont également reportées.

Pour rappel, nous avons vu que si un LR robuste est supérieur à 1, le risque que les observations soutiennent la mauvaise proposition est de 5 sur 48 soit 10%. Si un LR robuste est inférieur à 1, le risque que les observations soutiennent la mauvaise proposition est de 5 sur 27 soit 18%.

En prenant en compte les informations sur la zone de risque, c'est-à-dire si un LR est robuste et compris entre 0.4 et 2, le risque que les observations soutiennent la mauvaise proposition par rapport à l'alternative est de 52% (10 cas sur 19). Ceci est attendu, les LR étant faibles.

En revanche, pour les 40 autres cas pour lesquels un LR robuste non compris entre 0.4 et 2 avec un ratio  $E_2/E_1$  plus grand que 0,1 est obtenu, le risque est alors plus faible : ce cas n'a pas été observé.

Dans notre étude, la « zone de risque » est définie par le ratio  $E_2/E_1$  et la valeur des LR qui sont obtenus à partir de l'étude de ces 111 cas comprenant 2 couples de donneurs. Ces éléments (ratio  $E_2/E_1$  et valeur du LR) devraient être définis à partir de plus de cas.

**En conclusion**, en pratique, le LR calibré peut être transmis dans un rapport seulement s'il est robuste, ce qui implique que l'on dispose de suffisamment de données pour avoir confiance en celui-ci. Lorsque le LR calibré est proche de 1, compris entre 0.4 et 2, il est observé que les observations soutiennent la mauvaise proposition par rapport à l'alternative dans environ 50% des cas, ce qui est attendu. En revanche, lorsque le LR calibré est non compris entre 0.4 et 2, il n'a jamais été remarqué que les observations soutenaient la mauvaise proposition par rapport à l'alternative. Si le LR calibré n'est pas robuste, il est nécessaire d'acquérir des données supplémentaires.

## V.4 Discussion

Deux points sont discutés dans cette partie :

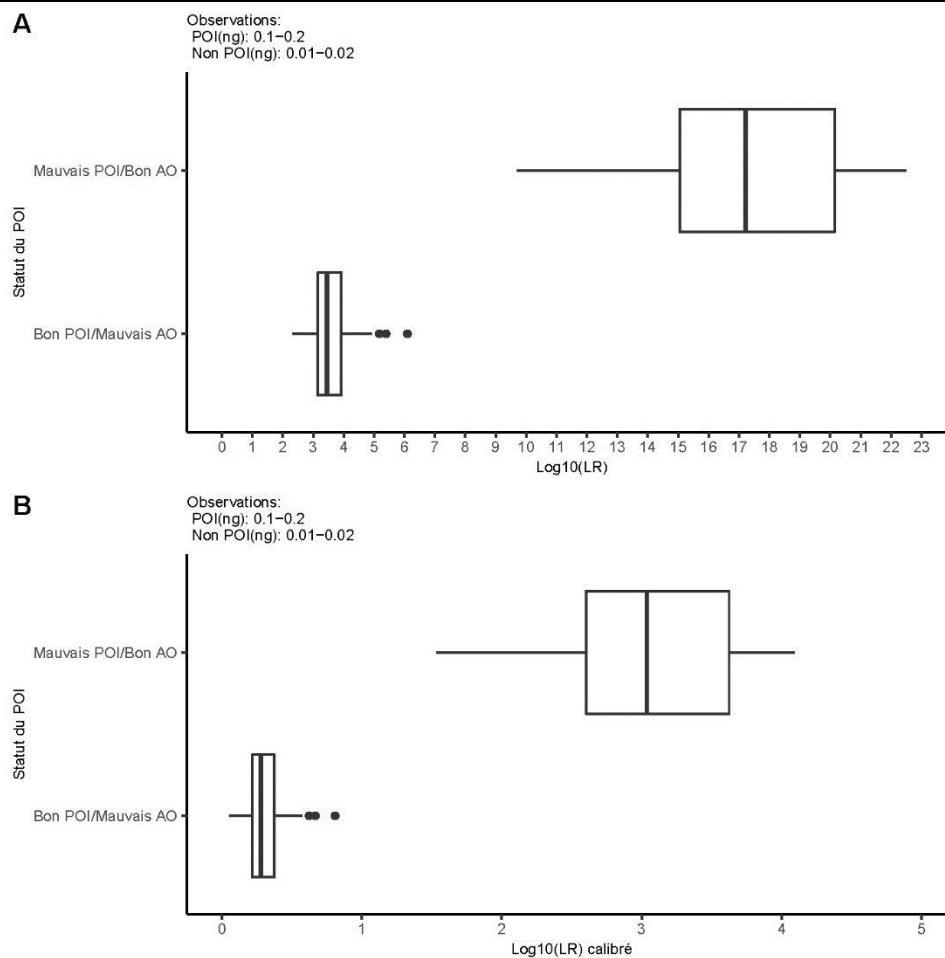
- la valeur du LR,
- le statut de donneur de l'AO.

### V.4.1 Valeur du rapport de vraisemblance

Nous avons observé que des LR initiaux extrêmement grands ou extrêmement petits pouvaient être obtenus (cf. partie V.3.1). Afin de faciliter la discussion, reprenons l'exemple du cas de M.X.

La Figure 23 présente les résultats de l'évaluation statistique obtenus pour le cas de M.X. en tenant compte des simulations par ré-échantillonnage, avant (LR initiaux non calibrés) et après (LR calibrés) les avoir calibrés.

## Evaluation des résultats ADN considérant des propositions au niveau de l'activité 2019



**Figure 23 : Variation des log<sub>10</sub>(LRs) non calibrés (A) et calibrés (B) et pour le cas de M.X, en fonction de la capacité du POI (M.X) à transférer son ADN par rapport à l'AO.**

La Figure 23 montre que, si le POI est moins bon donneur que l'AO, le LR médian obtenu est de l'ordre de  $10^{17}$ . Ce LR est-il réaliste ? La valeur élevée du LR est liée à la paramétrisation du BN. En effet, les distributions issues des données informant le réseau BN ont pour conséquence d'obtenir des probabilités extrêmement faibles d'observer les états extrêmes par rapport aux états qu'il est le plus probable d'observer. Pour exemple, le Tableau 9 présente la table des probabilités conditionnelles obtenue pour la variable « Efficacité d'extraction » dans le cas de M.X.

**Tableau 9 : Table des probabilités conditionnelles associées à la variable « Efficacité d'extraction » utilisée pour le cas de M.X lors de la paramétrisation.**

Etat	Probabilité associée
[0.00,0.05)	1.E-03
[0.05,0.10)	3.E-02
[0.10,0.15)	1.E-01
[0.15,0.20)	2.E-01
[0.20,0.25)	2.E-01
[0.25,0.30)	2.E-01
[0.30,0.35)	1.E-01
[0.35,0.40)	6.E-02
[0.40,0.45)	3.E-02
[0.45,0.50)	1.E-02
[0.50,0.55)	3.E-03
[0.55,0.60)	7.E-04
[0.60,0.65)	1.E-04
[0.65,0.70)	1.E-05
[0.70,0.75)	1.E-06
[0.75,0.80)	6.E-08
[0.80,0.85)	1.E-09
[0.85,0.90)	9.E-12
[0.90,0.95)	6.E-15
[0.95,1.00]	1.E-19

Le Tableau 9 montre que la probabilité d'observer l'état le plus extrême par rapport à l'état le plus probable, c'est-à-dire d'observer une efficacité d'extraction comprise entre 95% et 100% est de  $1^{E-19}$ .

Afin d'obtenir des probabilités plus réalistes d'un point de vue pratique, une solution serait de borner à une valeur minimale qui semble raisonnable. Dans ce cas, les questions des bornes à utiliser et pour quelle(s) raison(s) se posent. Lorsque l'évaluation se fait en considérant des propositions au niveau de la source, il est recommandé de borner le LR à 1 milliard (Foreman et Evett, 2001, Hopwood et al., 2012), représentant un nombre raisonnable de comparaisons.

L'analyse de sensibilité repose entièrement sur le nombre de données à disposition pour chaque variable. Une proposition pourrait être alors de borner les probabilités en définissant un nombre raisonnable de données qui pourraient être à disposition. Par exemple, si nous jugeons qu'il est raisonnable de penser que 4000 données pour chaque variable pourraient

être à disposition, les probabilités associées aux états des variables ne pourraient pas être plus petites que 1/4000.

Une autre solution serait d'utiliser les distributions permettant de prendre en compte ces états extrêmes de sorte que la probabilité de les observer ne soit pas extrêmement faible et donc déraisonnable (Saralees Nadarajah, 2003).

Calibrer le système a entraîné une diminution de la valeur des LR (Figure 23) ce qui était attendu au vu des LR obtenus pour étudier l'efficacité de la méthodologie. Les LR calibrés reflètent de manière plus réaliste la valeur attendue lorsque peu de données informent le BN.

Il serait utile d'étudier les distributions sous-jacentes plus en détail, afin que les probabilités représentent de manière calibrée ce qui est observé. Le système serait alors calibré directement sans avoir besoin d'effectuer une calibration de ce système après coup.

#### V.4.2 Statut du donneur de l'AO

Dans un cas pratique, l'AO n'est pas connu. Or nous avons vu que le statut relatif de la qualité des donneurs impacte fortement le résultat de l'évaluation.

L'AO étant inconnu, plusieurs options peuvent être envisagées :

- Utiliser par défaut les données d'une personne moins bonne donneuse que le POI. En effet, nous avons observé précédemment que le LR est plus petit lorsque l'AO est en moyenne moins bon donneur que le POI. Cette mesure serait donc conservatrice.
- Utiliser les données issues d'une population générale. Pour chaque individu composant cette population, des données sur les quantités d'ADN sur les mains et les proportions de transfert seraient obtenues. Cette mesure permettrait de prendre en compte la probabilité que le POI soit moins bon ou meilleur que l'AO.

La deuxième option semble plus adaptée car elle permet de tenir compte de l'incertitude quant au statut relatif de l'AO par rapport au POI.

Afin d'étudier la deuxième option, les simulations par ré-échantillonnage sont effectuées telles que décrites dans la partie V.2 en regroupant cette fois les données lorsque l'AO est le participant 2 et l'AO est le participant 6 pour informer les variables « quantités d'ADN sur les mains » et « Proportion de transfert ». Les LR ont ensuite été calibrés suivant la régression logistique dont les paramètres ont été fixés précédemment.



Pour chacune des observations  $E_i = \{E_1, E_2\}$ , les résultats peuvent être consultés en sélectionnant « Part 3 : Results of cases » sur le site [https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/). Ils peuvent être comparés aux résultats obtenus en utilisant les données spécifiques de l'AO.

Si les observations soutiennent la proposition du procureur, nous observons que, de manière générale, le LR est plus faible dans le cas où le POI est moins bon donneur et le LR est plus haut dans le cas où le POI est meilleur donneur si la population générale (composée d'un bon et d'un mauvais donneur) est utilisée plutôt que les données spécifiques de l'AO comme l'illustre l'exemple du cas de M.X (Figure 24), ce qui était attendu.

Si les observations soutiennent la proposition de la défense (Transfert secondaire allégué), nous observons que, de manière générale, le LR est plus élevé dans le cas où le POI est moins bon donneur et le LR est plus faible dans le cas où le POI est meilleur donneur si la population générale (composée d'un bon et d'un mauvais donneur) est utilisée plutôt que les données spécifiques de l'AO, ce qui était attendu.

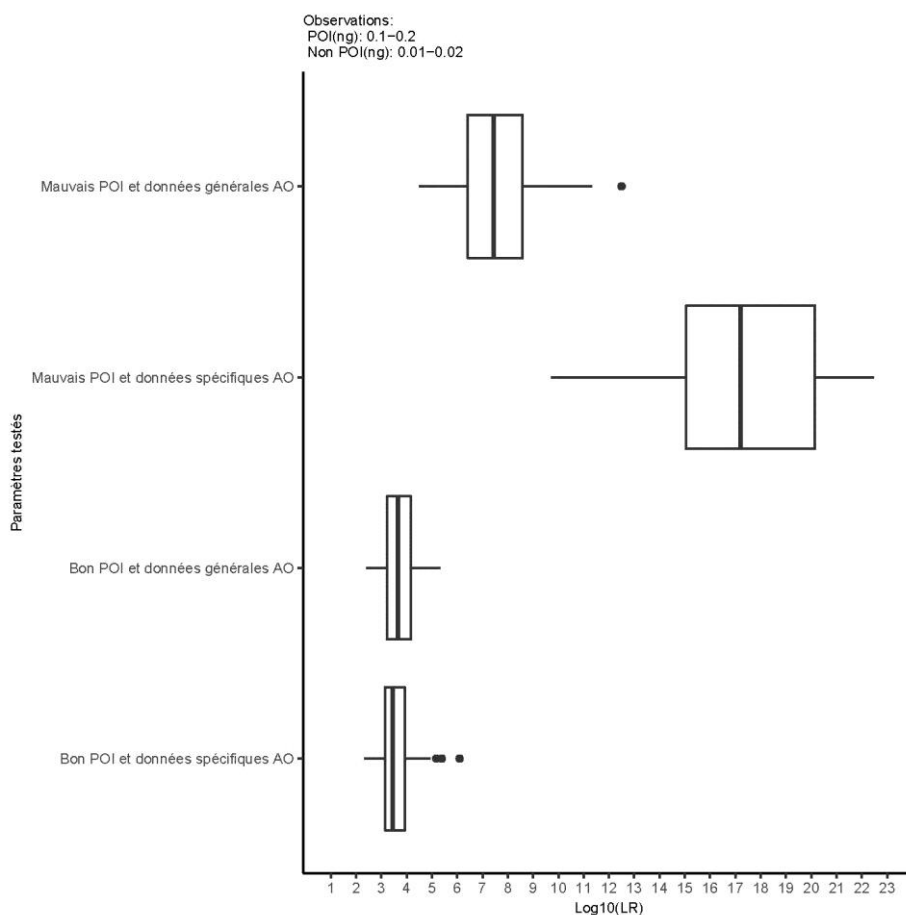


Figure 24 : Variation des  $\log_{10}(\text{LRs})$  calibrés pour le cas de M.X, en fonction de la capacité du POI (M.X) à transférer son ADN par rapport à l'AO et des données (spécifiques ou d'une population générale) de l'AO.

On observe également qu'en utilisant les données spécifiques de l'AO, la variation des LR est plus faible dans le cas où le POI est moins bon donneur que l'AO par rapport au cas où le POI est meilleur que l'AO, comme l'illustre l'exemple du cas de M.X (Figure 24). Ceci peut expliquer que la variation des LR soit plus faible lorsque les données de l'AO d'intérêt, meilleur donneur que le POI, et d'une autre personne, moins bonne donneuse que l'AO d'intérêt, compose la population générale que lorsque seules les données de l'AO d'intérêt est pris en compte. A l'inverse, nous observons que la variation des LR est un peu plus importante en prenant en compte les données de l'AO d'intérêt, moins bon donneur que le POI, et d'une autre personne, meilleure donneuse que l'AO d'intérêt, dans la population générale que lorsque seules les données de l'AO d'intérêt est pris en compte.

Dans ce travail, seuls deux individus ont été pris en compte, qui étaient soit aussi bon que le POI soit aussi mauvais. Nous nous attendons à ce que ces effets de variation soient atténués avec un grand nombre d'individus plus ou moins bons donneurs pris en compte dans la population générale. Le Tableau 10 présente, lorsque les données combinées de plusieurs personnes sont prises en compte pour définir l'AO :

- le nombre de cas où les observations soutiennent la bonne ou la mauvaise proposition et dont le LR calibré pourra être transmis car jugé robuste,
- le nombre de cas ayant un LR calibré non robuste, dont une acquisition de données supplémentaire serait nécessaire.

**Tableau 10 : Nombre de cas pour lesquels un LR robuste ou non est obtenu, en fonction de la capacité du POI à transférer son ADN par rapport à l'AO, lorsque les données spécifiques de l'AO (en rouge) ou d'une population générale représentant l'AO sont utilisées (en bleu).**

POI	« Bonne » Proposition	LR robuste Observations transmises soutenant la « bonne » proposition	LR robuste Observations transmises soutenant la « mauvaise » proposition	LR non robuste
Meilleur donneur que AO	Procureur	21-20	5-5	4-5
Moins bon donneur que AO	Procureur	22-17	0-1	1-5
Meilleur donneur que AO	Défense	21-26	5-1	2-1
Moins bon donneur que AO	Défense	1-25	0-4	29-1

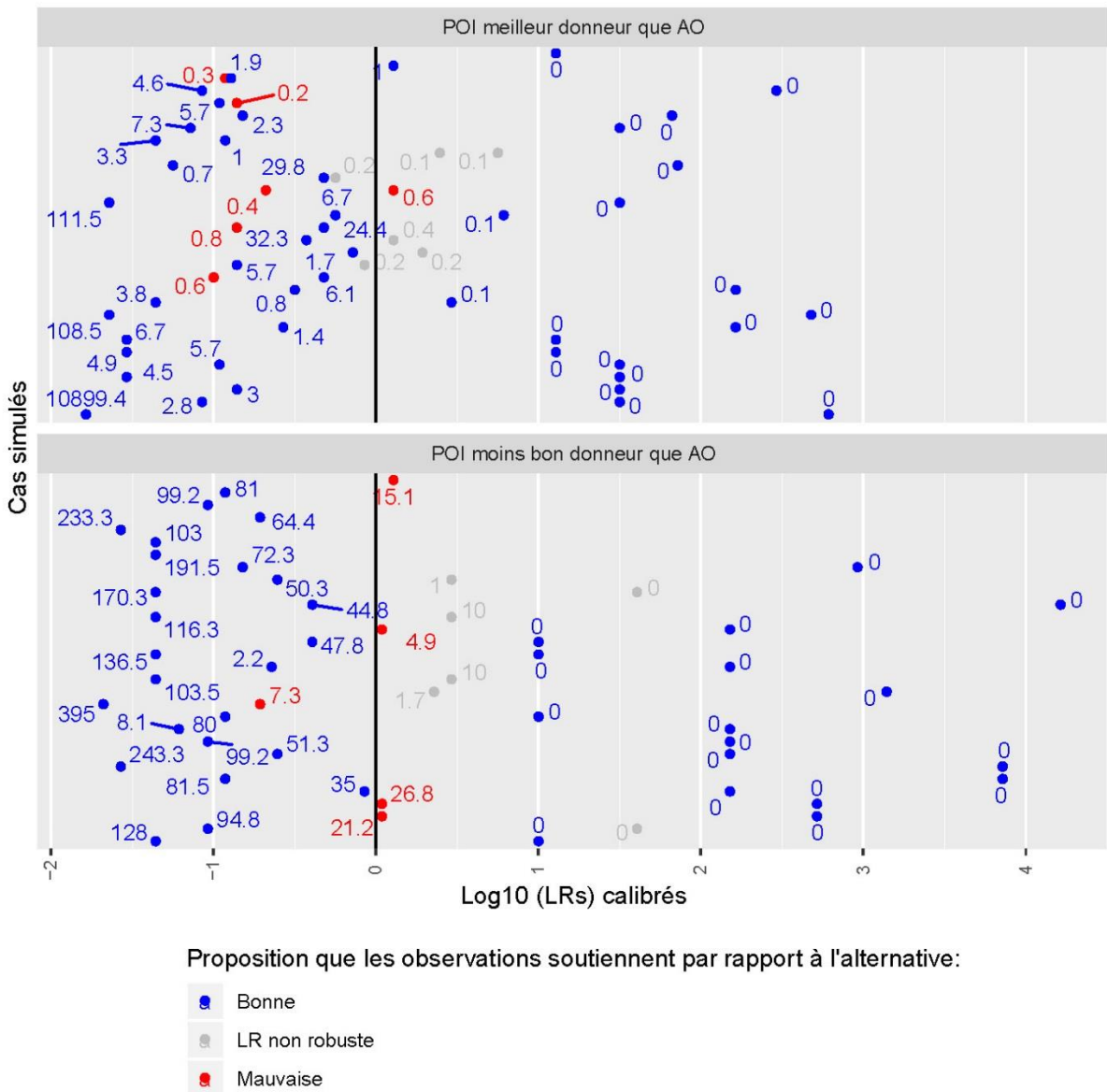


Figure 25 : Points illustrant le  $\log_{10}(\text{LR})$  calibré obtenu pour chacun des cas simulés lorsque le POI est meilleur ou moins bon donneur que l'AO, associé au ratio  $E_2/E_1$ , indiqués en bleu si les observations soutiennent la bonne proposition, en rouge si les observations soutiennent la mauvaise proposition et en gris s'il n'est pas robuste lorsque les données d'une population générale représentant l'AO sont utilisées.

Le Tableau 10 montre que les résultats sont similaires, que l'ensemble des données des deux participants soit pris en compte pour les variables relatives à l'AO ou que seules les données spécifiques à l'AO soient prises en compte, excepté lorsque la proposition de la défense est vraie et que le POI est moins bon donneur que l'AO. Dans ce cas, un LR robuste de 28 cas est obtenu, sur les 29 dont un LR non robuste avait été obtenu, dont 4 pour lesquelles les observations soutiennent la mauvaise proposition.

Ces 28 cas pour lesquels le LR peut être reporté lorsque la population générale d'AO est pris en compte, résultent de la diminution de la variation des LR lorsqu'un AO moins bon donneur que l'AO spécifique fait partie de cette population générale.

Si un LR robuste plus grand que 2 est obtenu, le risque est faible que les observations soutiennent la mauvaise proposition par rapport à l'alternative. Si un LR robuste inférieur à 1 est obtenu et que le ratio obtenu est entre 0,2 et 7,3, le risque que les observations soutiennent la mauvaise proposition par rapport à l'alternative est faible (Figure 25). Le LR le plus bas transmis alors que les observations soutiennent la mauvaise proposition est de 0.1, le plus haut transmis est seulement de l'ordre de 2.

Dans 12 cas sur 111 soit 10% des cas, un LR non robuste est obtenu et ne serait donc pas reporté. Or dans ces 12 cas, un LR signifiant que les observations soutenaient la bonne proposition a été obtenu dans 8 cas.

Les données spécifiques de l'AO ayant également été prises en compte dans la « population générale », ce résultat ne peut être généralisé à celui qui aurait été obtenu en prenant une population générale comptant plus d'individus et ne comptant pas l'AO. Cependant, plus la population générale est grande, plus il est attendu que les LR obtenus ainsi que la variation des LR, soient proches de ceux obtenus si seules les données de l'AO en question avaient été pris en compte.

**En conclusion**, le LR qui devrait être reporté doit être robuste, au sens défini dans ce chapitre, et calibré. En appliquant la méthodologie pour évaluer les résultats ADN obtenus des 111 cas empiriquement reproduits, le LR ne peut être reporté dans 36 cas (soit 32% des cas). Dans 65 cas (soit 58% des cas), les observations soutiennent la bonne proposition par rapport à l'alternative et le LR peut être reporté.

En revanche dans 10 cas (soit 9% des cas), le LR calibré peut être reporté alors que les observations soutiennent la mauvaise proposition par rapport à l'alternative. Cependant ces LR calibrés sont compris entre 0.4 et 2 seulement. Le ratio  $E_2/E_1$  ainsi que la valeur du LR peut aider à détecter ces cas. Plus d'expériences avec plus de couples différents seraient nécessaires pour affiner les conditions délimitant une « zone de risque » où le LR reporté signifierait que les observations soutiennent la mauvaise proposition.

Dans les cas empiriquement reproduits, l'AO était connu et les données lui correspondant ont pu être utilisées. Toutefois, celui-ci étant généralement inconnu, il faudrait pouvoir utiliser les données issues d'une population générale, qui ne sont pas disponibles à l'heure actuelle.

---

Nous avons également montré la nécessité d'avoir un système calibré. Afin d'améliorer cette calibration, une étude approfondie des distributions sous-jacentes paramétrant le BN serait à envisager.

## Chapitre VI - Vers une application pratique de la méthodologie

Pour rappel, l'objectif de cette recherche est d'élaborer une méthodologie qui permet aux experts d'évaluer des résultats ADN lorsque différentes activités sont alléguées pour caractériser les propositions et qui soit applicable à de nombreux cas.

L'objectif de ce chapitre est de définir, en pratique, les différentes étapes de la méthodologie élaborée suite aux résultats obtenus, d'étudier les possibilités de généralisation et d'identifier également la plus-value de ce travail, les limites pratiques ainsi que les perspectives futures.

### VI.1 Les étapes de la méthodologie en pratique

Dans cette partie, un protocole d'application, qui découle des résultats obtenus dans ce travail est présenté. La Figure 26 détaille ce protocole.

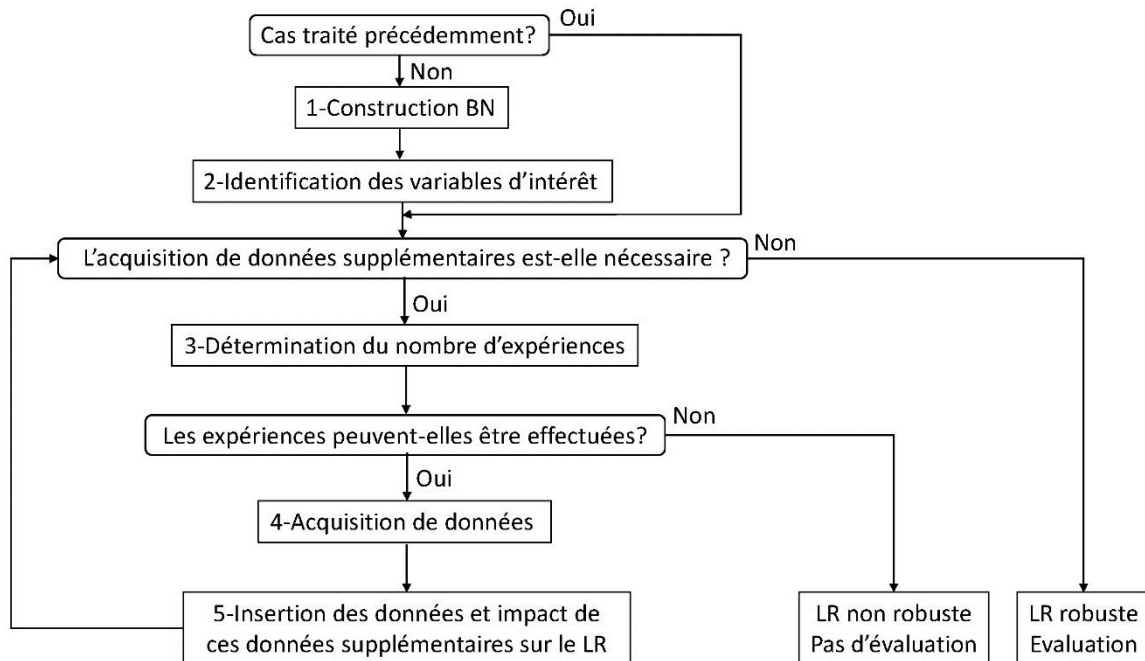


Figure 26 : Protocole d'application de la méthodologie

La méthodologie se décompose en plusieurs étapes détaillées ci-après :

- Etape 1 : si un BN n'a pas déjà été construit, il faut donc le faire. Il s'agit d'élaborer un modèle probabiliste qui tienne compte des propositions, des observations et de toutes les variables qui influencent les observations. Le BN permet de calculer un LR à partir des données disponibles dans la littérature et /ou issues d'expériences.

- Etape 2 : identification des variables d'intérêt à partir d'une analyse de sensibilité du BN. Cette analyse permet, par la technique de ré-échantillonnage, de déterminer les variables ayant le plus d'impact sur le LR et, en même temps, de déterminer la robustesse du LR.
  - Transmission d'un LR robuste dans le rapport d'expertise, s'il n'est pas nécessaire d'acquérir des données supplémentaires.
- Etape 3 : Déterminer le nombre d'expériences supplémentaires, s'il est nécessaire d'acquérir des données.
  - Les experts ne peuvent se prononcer sur le résultat du LR car ces données supplémentaires sont nécessaires mais ne peuvent être acquises.
- Etape 4 : acquisition des données car ces données sont nécessaires et peuvent être acquises. Les données associées aux variables sont obtenues à partir d'expériences plus ou moins spécifiques au cas, dépendant de la variable pour laquelle les expériences sont faites.
- Etape 5 : Implémentation des données acquises dans le BN et analyse de l'impact de ces nouvelles données sur le LR. L'objectif est de définir s'il est nécessaire de mener des expériences supplémentaires.

Si le cas peut être modélisé par un BN déjà élaboré, les étapes 1 et 2 ont déjà été effectuées donc le protocole peut être commencé à l'analyse des résultats de l'étape 2 afin de définir si des données supplémentaires doivent être acquises ou non.

Les limites pratiques de ce protocole seront discutées dans ce chapitre (voir Partie VI.)

Nous avons montré dans ce travail que si un transfert secondaire est allégué par la défense dans un cas d'agression utilisant une arme à main nue, l'évaluation nécessite de faire des expériences, a minima pour définir la variabilité intra individuelle concernant la quantité d'ADN transféré par le POI. Cela signifie que les données spécifiques supplémentaires doivent être acquises et le nombre peut être déterminé à l'étape 3.

Un exemple d'application de ce protocole est proposé dans la partie suivante.

## VI.2 Généralisation de la méthodologie

En théorie, la méthodologie proposée peut être généralisable et utilisée quel que soit le cas. Cependant, celle-ci repose sur la modélisation graphique probabiliste des cas, c'est-à-dire la construction du BN, qui peut être plus ou moins compliquée, ainsi que sur des simulations nécessitant des connaissances de R et dont la réalisation demande du temps de calcul.

En revanche, ce BN peut être réutilisé pour tous les cas, impliquant des traces ADN de contact pour lesquels un transfert direct d'ADN du POI est allégué sous  $H_p$ , un transfert secondaire d'ADN du POI et un transfert direct d'ADN de l'AO est alléguée sous  $H_{d1}$ , soit un transfert direct d'ADN du POI est allégué sous  $H_p$ , aucun transfert direct d'ADN du POI et un transfert direct d'ADN de l'AO est alléguée sous  $H_{d2}$ .

De plus les résultats de l'étape 2, qui visait à l'identification des variables d'intérêt, peuvent être réutilisés à partir du moment où les données (notamment du bruit de fond et de la quantité d'ADN sur la surface de départ, qui était la main dans notre cas) sont adéquates pour le cas en question.

Dans la pratique, l'origine de l'ADN peut être un réservoir d'ADN dont la quantité peut parfois ne pas être déterminée. Par exemple, un sachet de drogue est retrouvé dans le sac d'une personne et cette personne indique que quelqu'un l'a mis dans son sac, elle ne l'a jamais manipulé. Sous la proposition de la défense, l'ADN qui s'est déposé sur le sachet peut donc provenir de son sac. Si la quantité d'ADN qui s'est accumulé dans le sac ne peut pas être déterminée, l'évaluation statistique ne pourra être réalisée. Si celle-ci peut être déterminée, les données de quantités d'ADN sur les mains ne seront pas adéquates et doivent être remplacée par les données sur la quantité d'ADN qui s'est accumulée dans le sac.

En résumé, il est donc possible d'utiliser le BN construit dans ce travail ainsi que les résultats obtenus concernant l'identification des variables d'intérêt et la calibration des LR pour tout type de cas impliquant des traces ADN de contact pour lesquelles :

- un transfert direct d'ADN du POI est allégué sous  $H_p$  et un transfert secondaire d'ADN du POI et un transfert direct d'ADN de l'AO sont allégués sous  $H_{d1}$ ,
- un transfert direct d'ADN du POI est allégué sous  $H_p$  et aucun transfert direct d'ADN du POI et un transfert direct d'ADN de l'AO sont allégués sous  $H_{d2}$ ,



et ceci, quel que soit le type d'arme (donc de surface) utilisé, le type de contact, l'environnement (voir les résultats présentés dans le chapitre III) dont les données utilisées dans ce travail pour informer le BN sont jugées adéquates à ce nouveau cas.

Un exemple d'application du protocole est présenté dans cette partie concernant un cas d'agression au couteau pour lequel la proposition de la défense est différente de  $H_{d1}$ .

Le cas, appelé « Cas Mme Z », est le suivant : les secours sont appelés par Mme Z pour intervenir dans un immeuble vers 20h. Mme Z a, selon elle, découvert le corps de son mari M. Z poignardé de deux coups de couteau en rentrant à son domicile. Un couteau à manche lisse, dont la lame est ensanglantée, est retrouvé près de la victime. Celui-ci est conservé dans un sachet en papier. Le couteau est ensuite transmis au laboratoire qui procède au prélèvement de l'ADN au moyen d'un écouvillon puis analyse ce prélèvement. La mort remonte à quelques heures avant la découverte du corps. Durant l'enquête, Mme. Z est mise en cause, il s'agit donc de la personne d'intérêt appelée POI.

La mission confiée par le procureur à l'expert pourrait être la suivante :

« Veuillez procéder aux analyses ADN de l'arme, un couteau, afin de déterminer si l'échantillon remis comporte du matériel ADN. Le cas échéant, nous vous demandons de comparer le profil obtenu avec le profil ADN de Mme Z. Veuillez finalement indiquer dans quelle mesure vos résultats appuient l'une ou l'autre des propositions ».

Les propositions formulées jouent un rôle fondamental dans l'évaluation des résultats. Le couteau appartenait au couple Z, selon les dires de Mme Z. Les propositions alléguées sont donc les suivantes :

- Hypothèse  $H_p$  : « La POI a poignardé la victime »
- Hypothèse  $H_{d3}$  : « Quelqu'un d'autre a poignardé la victime, la POI a utilisé le couteau lors du déjeuner vers 12h ».

#### 1<sup>ère</sup> étape : Construction du réseau bayésien

La Figure 27 illustre l'OOBN construit pour le cas de Mme Z. Il s'agit d'une adaptation du modèle générique utilisé dans cette recherche.

Selon la proposition du procureur, l'ADN peut provenir :

- **De la POI**, qui est la personne qui a utilisé l'objet d'intérêt au moment de l'action délictueuse. Il y a alors eu un phénomène de transfert d'ADN de la POI directement

sur le couteau au moment de l'agression et un phénomène de persistance de l'ADN sur le manche du couteau entre l'action d'intérêt et la mise sous scellés soit un bloc dit « TP ».

- **Du bruit de fond**, qui provient des personnes, autre que la POI, ayant utilisé de manière légitime l'objet d'intérêt avant l'action. Le bruit de fond peut être différent ou correspondre fortuitement au profil ADN de la POI, soit un bloc dit « M ».

Ces deux voies sont indiquées en bleue sur la Figure 27.

Si la proposition de la défense  $H_{d3}$  : « Quelqu'un d'autre a poignardé la victime, la POI a utilisé le couteau le midi » est vraie, les quantités d'ADN obtenues peuvent provenir :

- **De la POI**, qui a utilisé le couteau de manière légitime lors de son déjeuner vers 12h. Dans ce cas, il y a un phénomène de transfert et de persistance de l'ADN de la POI sur le couteau soit un seul bloc TP (Figure 27), contrairement aux scénarios précédents utilisés dans ce travail.
- **De l'AO**, qui a utilisé l'objet d'intérêt au moment de l'action délictueuse. Il y a eu un phénomène de transfert de l'ADN de l'AO sur le couteau lors de l'agression et un phénomène de persistance de l'ADN soit un bloc TP.
- **Du bruit de fond**, qui provient des personnes, autre que la POI ayant utilisé de manière légitime l'objet d'intérêt avant l'action. Le bruit de fond peut être différent ou correspondre fortuitement au profil ADN de la POI et/ou de l'AO soit un bloc M.

Ces voies sont indiquées en orange sur la Figure 27. Les états des variables ainsi que les données utilisées issues de la littérature sont les même que pour l'OON du cas de M. X dans le chapitre II.

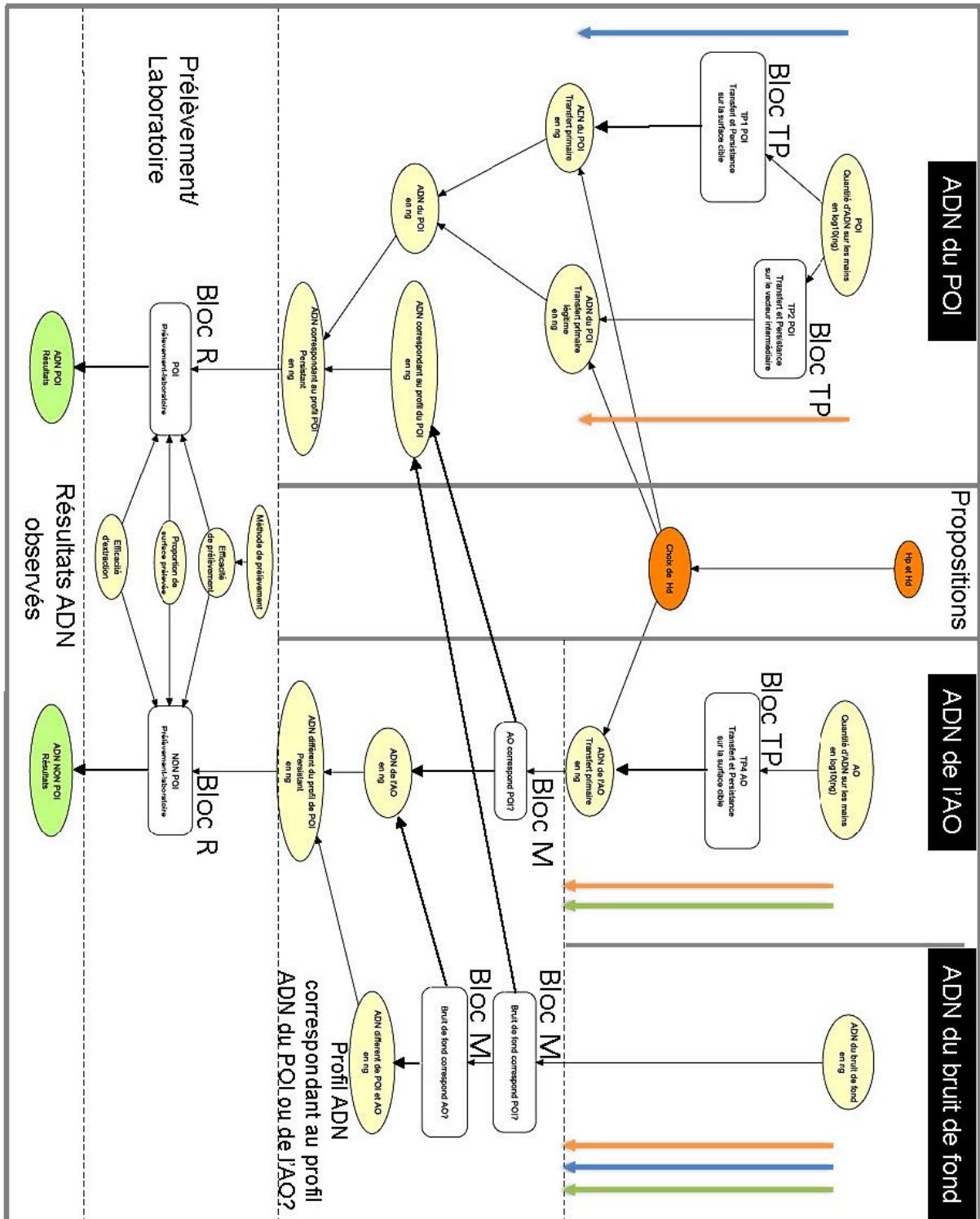


Figure 27 : OOBN utilisé pour évaluer les observations ADN considérant des propositions,  $H_p$  et  $H_3$ .

2<sup>ème</sup> étape : identification des variables d'intérêt (variables ayant le plus d'impact sur le LR)

Afin d'identifier ces variables, des simulations par ré-échantillonnage, à partir des données initiales issues de la littérature, sont effectuées en utilisant la méthode présentée par Samie et al (2019<sub>(b)</sub>) (Annexe II.1). Seul le résultat obtenu pour les observations du cas en question est pertinent. Cependant il a été choisi de présenter dans ce travail les résultats pour l'ensemble des observations possibles dans la Figure 28. Les résultats détaillés pour chaque observation sont disponibles en sélectionnant « Part 1 : Sensitivity analysis » sur le site [https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/).

La Figure 28 illustre la variation des étendues des  $\log_{10}(\text{LR})$  pour chacune des variables, ceci pour l'ensemble des observations  $E_i = \{E_1, E_2\}$ .

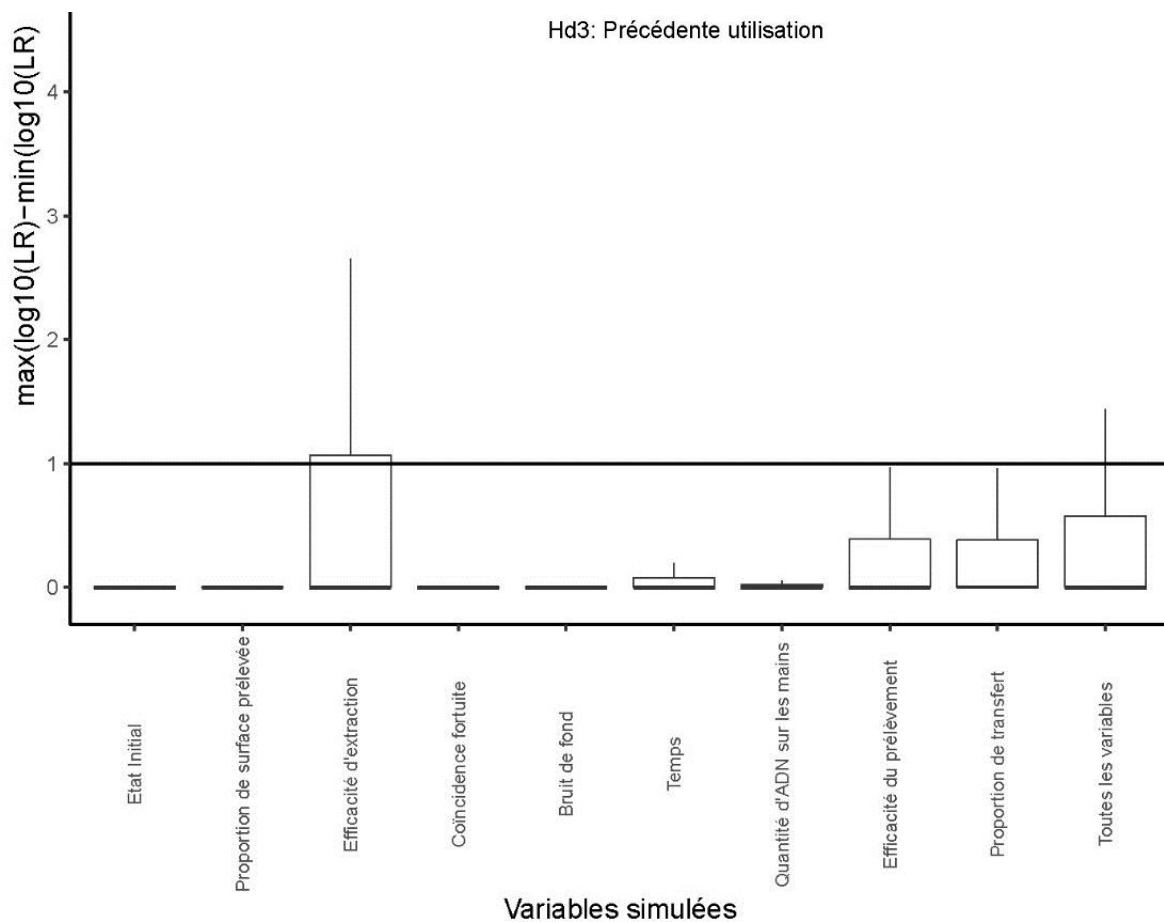


Figure 28 : Variation des étendues des  $\log_{10}(\text{LRs})$  pour chaque variable étudiée suite au changement des données de ces variables pour le scénario en question ( $H_p/H_{d3}$ ).

Considérant l'ensemble des observations, la seule variable ayant un impact significatif sur le LR est l'efficacité d'extraction. Au vu du résultat observé en Figure 28, des données supplémentaires sur l'efficacité d'extraction devraient être acquises.

De plus, les résultats du chapitre IV ont montré qu'il est recommandé :

- d'utiliser des données spécifiques de l'efficacité de la méthode de prélèvement ;
- de conduire des expériences reproduisant les activités alléguées afin d'obtenir les données **des personnes d'intérêt**.

#### 3<sup>ème</sup> étape : Détermination du nombre d'expériences

Des simulations par ré-échantillonnage d'un plus grand nombre de données, suivant la méthode proposée dans le chapitre III.3 pourraient être faites. Cependant en se basant sur les résultats obtenus pour le scénario 1 (Chapitre III), il a été jugé pertinent d'acquérir 30 données sur l'efficacité d'extraction. Les expériences peuvent être effectuées.

#### 4<sup>ème</sup> étape : Acquisition des données

Les 30 données obtenues par des expériences sur l'efficacité d'extraction ainsi que sur l'efficacité de la méthode de prélèvement (Voir Chapitre IV) ont été utilisées pour mettre à jour la paramétrisation du BN.

Concernant les données sur « Quantités d'ADN sur les mains » et « Proportion de Transfert », deux situations ont été étudiées. La première considère un POI qui, en moyenne, est un meilleur donneur que l'AO et la deuxième un POI en moyenne moins bon donneur que l'AO. Les mêmes participants 2 et 6 ont été considérés. Les données relatives à la quantité d'ADN sur les mains sont donc reprises du chapitre III et présentées en Annexe VI.3.

La proposition du procureur est la même que pour le cas de M. X donc les données du cas de M. X sur les proportions de transfert lorsque la proposition du procureur est vraie sont utilisées. Pour obtenir les données de la proportion de transfert et du bruit de fond du cas de Mme Z lorsque la proposition de la défense est vraie, 30 expériences ont été réalisées pour chacun des couples (Le participant 2 est le POI avec le participant 6 qui est l'AO et inversement). Le plan d'expérience est le suivant :

Le POI coupe une pomme avec un couteau puis l'AO se saisit du couteau et poignarde un bloc de savon balistique. Un prélèvement ADN avec le FLOQSwab™ est ensuite effectué sur le manche du couteau. L'ADN est extrait avec kit QIAshredder-QIAamp Mini DNA en utilisant les colonnes Microcon® 30 puis quantifié en utilisant le kit Investigator® Quantiplex de Qiagen sur le Rotor-Gene® Q. Les profils sont établis suite à l'électrophorèse capillaires effectuée sur l'instrument 3500 (Applied Biosystems) et à l'analyse par le logiciel Genemapper ID-X (ThermoFisher Scientific).

La proportion des différents contributeurs est calculée en utilisant STRmix™. La déconvolution des profils ADN par STRmix™ a pu être effectuée pour 113 expériences.

Les résultats des données de « Proportion de transfert » et de « Bruit de fond » sont présentés en Annexe VI.6 et VI.8.

Le BN, utilisé pour évaluer ces cas, a été paramétrisé avec ces données.

#### 5<sup>ème</sup> étape : Insertion des données et impact des données supplémentaires sur le BN

Les 113 expériences effectuées à l'étape 4 constituent 113 cas dont nous savons si la proposition du procureur est vraie ou si la proposition de la défense est vraie. Les résultats de ces 113 expériences sont évalués en utilisant les données obtenues à l'étape 4 pour paramétriser le BN. Les résultats détaillés sont présentés en Annexe VII.2.

Afin de déterminer si le LR obtenu est robuste, 100 simulations ont été effectuées, basées sur les mêmes principes que les simulations du chapitre V.

Le Tableau 11 présente le nombre de cas pour les lesquels les observations soutiennent la bonne ou la mauvaise proposition et le nombre de cas pour lesquels un LR calibré non robuste est obtenu.

**Tableau 11 : Nombre de cas dont les observations soutiennent la bonne ou la mauvaise proposition ou dont un LR calibré non robuste est obtenu, en fonction de la capacité du POI à transférer son ADN par rapport à l'AO.**

POI	« Vraie » Proposition	LR calibré robuste Observations transmises soutenant la « bonne » proposition	LR calibré robuste Observations transmises soutenant la « mauvaise » proposition	LR calibré non robuste
Meilleur donneur que AO	Procureur	10	10	10
Moins bon donneur que AO	Procureur	14	1	8
Meilleur donneur que AO	Défense	29	1	0
Moins bon donneur que AO	Défense	20	2	8

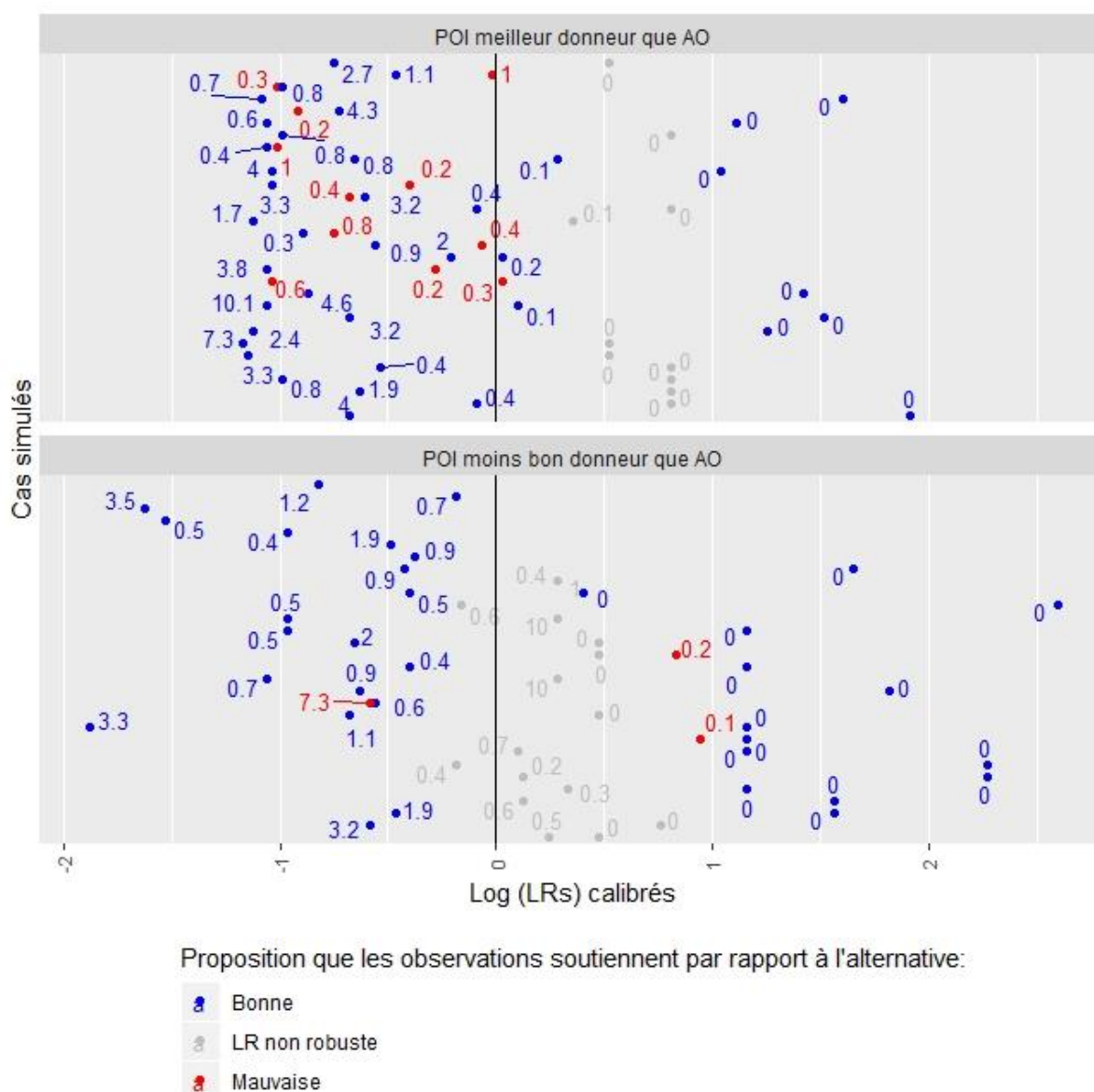


Figure 29: Points illustrant les  $\log_{10}(\text{LRs})$  calibrés obtenus pour chacun des cas simulés lorsque le POI est meilleur ou moins bon donneur que l'AO, associé au ratio  $E_2/E_1$ , indiqués en bleu si les observations soutiennent la bonne proposition, en rouge si les observations soutiennent la mauvaise proposition et en gris s'il n'est pas robuste.

Le Tableau 11 et la Figure 29 montrent que lorsque la proposition du procureur est vraie, les observations soutiennent la proposition du procureur par rapport à celle de la défense dans seulement 24 cas sur 53. Les ratios des quantités d'ADN prélevés  $E_2/E_1$  sont majoritairement proches de 0. En revanche, le LR est jugé non robuste sur 18 cas et les observations soutiennent la proposition de la défense dans 11 cas sur 53 soit dans 20% des cas.

Lorsque la proposition de la défense est vraie, les observations soutiennent la proposition de la défense par rapport à celle du procureur dans 49 cas sur 60 soit dans 81% des cas. En revanche, les observations soutiennent la proposition du procureur par rapport à celle de la

défense dans 3 cas sur 60 soit dans 5% des cas. Dans l'ensemble de ces cas, le ratio des quantités d'ADN prélevée  $E_2/E_1$  est plus grande que 0.1. Huit LR non robustes sont obtenus.

Le LR le plus bas transmis alors que les observations soutiennent la mauvaise proposition est de 0.09, le plus haut transmis est de 8.

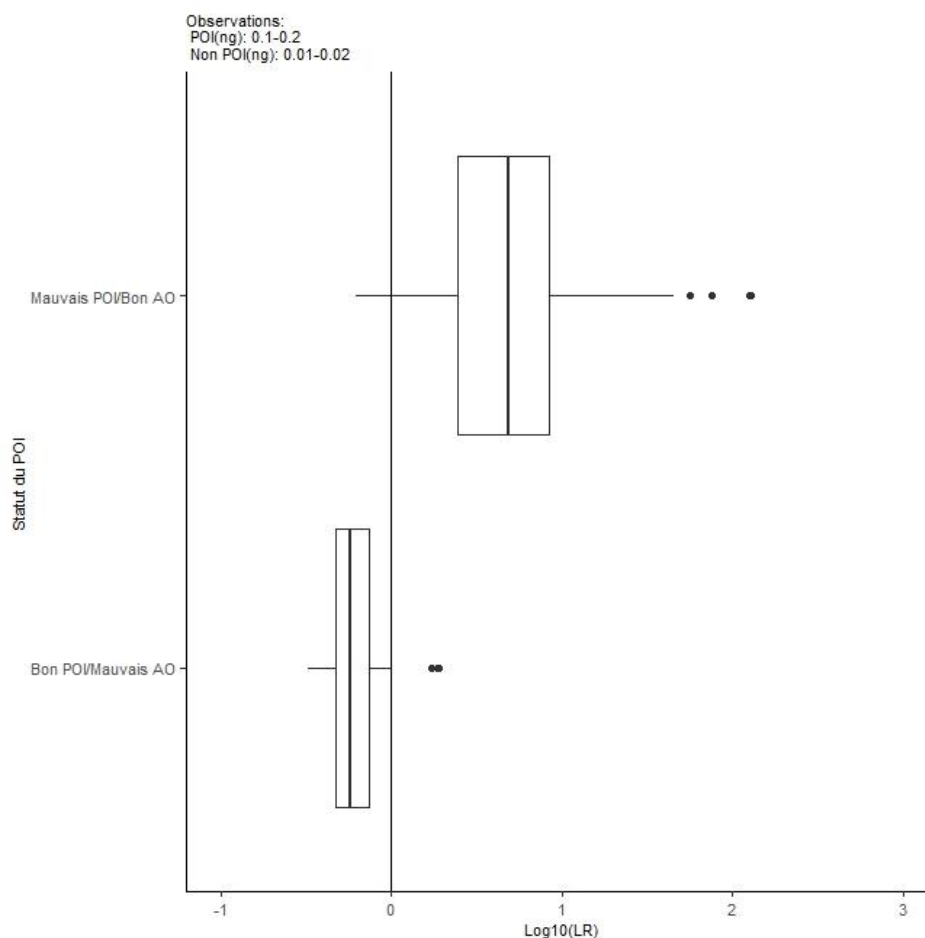
Si un  $LR > 1$  est obtenu et que le ratio est plus grand que 0.1, le risque est plus grand que les observations soutiennent la mauvaise proposition par rapport à l'alternative.

Sur 60 LR robuste inférieurs à 1, 11 LR signifient que les observations soutiennent la mauvaise proposition. Le risque d'erreur si un LR inférieure à 1 est obtenu est de 18%.

Dans 26 cas sur 113, soit 23% des cas, un LR non robuste est obtenu. Sur ces 26 cas, les observations soutenaient la bonne proposition dans 20 cas : 18 cas lorsque la proposition du procureur est vraie et 2 cas lorsque la proposition de la défense est vraie et que le POI est moins bon donneur que l'AO.

La Figure 30 illustre le résultat des 100 simulations effectuées à partir du BN informé avec les données de la littérature et des données spécifiques à l'AO lorsqu'une quantité de 0.1-0.2 pg correspondant au profil du POI et une quantité 0.01-0.02 différente du profil du POI sont obtenues. Les quantités ont été choisies pour refléter un cas d'espèce.





**Figure 30 : Variation des  $\log_{10}(\text{LRs})$  calibrés, en fonction de la capacité du POI (bon ou mauvais donneur) à transférer son ADN par rapport à l'AO, lorsque les données d'une population générale représentant l'AO sont utilisées dans un cas pour lequel l'utilisation légitime du couteau est alléguée par la défense.**

Dans cet exemple, pour ces observations ADN  $\{E_1 : 0.1-0.2 ; E_2 : 0.01-0.02\}$ , aucune acquisition de nouvelles données supplémentaires n'est nécessaire, un LR robuste peut être transmis. En revanche, cet exemple illustre la nécessité d'obtenir des données spécifiques au POI afin d'établir s'il est plus probable que celui-ci soit meilleur donneur ou moins bon donneur que la population générale.

### VI.3 Plus-value de la démarche proposée

#### VI.3.1 Comparaison avec les résultats obtenus par Pun (2016)

Si un transfert secondaire de l'ADN du POI ou un transfert direct légitime est allégué par la défense, les variables d'intérêt identifiées par Pun (2016) suite à l'analyse de sensibilité sont :

- le bruit de fond
- le transfert.

En revanche, la persistance et la pollution ne sont pas des variables d'intérêt, identifiées comme ayant un impact sur le LR. Ces conclusions, malgré un scénario (vol de voiture), des propositions et un BN différents, sont similaires à nos résultats. Nous avons en effet identifié les variables suivantes :

- « Quantités d'ADN sur les mains », « Proportion de transfert »,
- « Bruit de fond »,
- « Efficacité d'extraction »
- « Efficacité du prélèvement » (seulement si un transfert secondaire est allégué et non un transfert direct légitime)

Les variables liées à la persistance de l'ADN n'ont, comme Pun (2016), pas été retenues comme étant des variables d'intérêt. Pour rappel, la pollution est comprise dans la variable « Bruit de fond ».

La plus-value du BN, au vu des résultats obtenus concernant l'analyse de sensibilité, est la prise en compte de l'impact des variables associées aux laboratoires soit l'efficacité de la méthode de prélèvement et l'efficacité du kit d'extraction.

Si aucun transfert de l'ADN du POI n'est allégué par la défense, la variable d'intérêt identifiée dans ce travail et par Pun (2016) est la variable « Coïncidence fortuite ».

Ce travail apporte des contributions additionnelles :

- l'impact des variables associées aux laboratoires soit l'efficacité de la méthode de prélèvement et l'efficacité du kit d'extraction est étudiée, comme dit précédemment,
- l'information concernant les fractions majeures/fractions mineures dans un profil ADN est prise en compte dans l'évaluation grâce à l'observation constituée de deux éléments E1 (quantité d'ADN dont le profil correspond au profil ADN du POI) et E2 (Quantité d'ADN dont le profil ADN est différent de celui du POI),
- l'application [https://lydie-samie.shinyapps.io/DNA\\_Activity\\_Ph/](https://lydie-samie.shinyapps.io/DNA_Activity_Ph/) permet la visualisation rapide des résultats obtenus pour chaque observation possible. Elle peut être réutilisée pour certains cas dont le BN et les données utilisées pour la paramétrisation sont adaptés,
- une pré-évaluation de ces cas, grâce à l'application, peut être réalisée, comme décrit dans la partie qui suit.

### VI.3.2 Pré-évaluation

Pour un cas où les résultats de l'étape 1 (construction du BN) et l'étape 2 (identification des variables d'intérêt) peut s'appliquer, il a été établi que l'évaluation nécessite de faire des expériences a minima pour définir la variabilité intra individuelle concernant la quantité d'ADN transférée par le POI.

Cependant, il est possible de réaliser une pré-évaluation avant d'acquérir des données au moyen de l'application [https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/).

La pré-évaluation permet d'aider à apprécier notre capacité à réaliser une évaluation dans l'affaire en question. Elle permet de visualiser l'ordre de grandeur du LR potentiellement attendu, si celui-ci peut être robuste et si cette évaluation nécessite potentiellement l'acquisition d'un grand nombre de données. Etre en mesure de définir **au préalable** la nécessité d'acquérir d'autres données supplémentaires et en quelle quantité est un avantage considérable pour limiter l'investissement dans une expertise.

Prenons comme exemple un deuxième cas d'agression mais cette fois à la hache plutôt qu'un couteau. Les propositions alléguées prendront les formes les suivantes :

- Hypothèse  $H_p$  : « Le POI a frappé à la hache la victime »,
- Hypothèse  $H_{d1}$  : « Le POI a serré la main de l'AO qui a agressé la victime avec la hache retrouvée sur les lieux ».

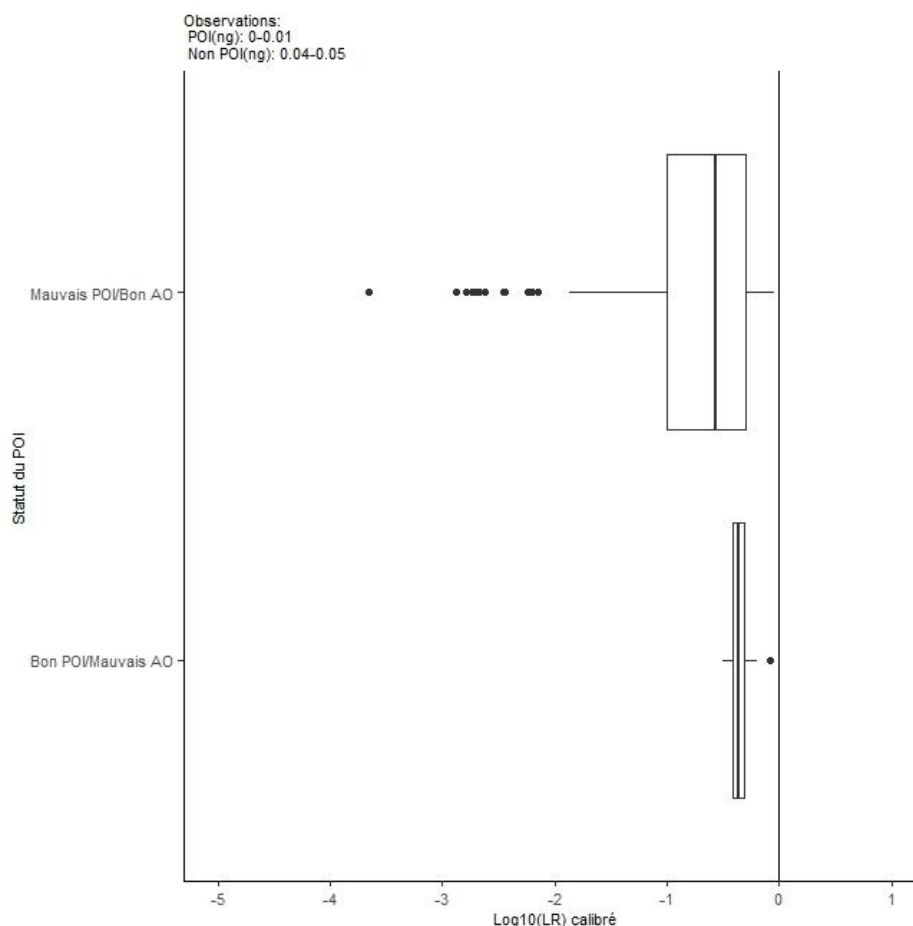
Une trace ADN de mélange composée d'une fraction majeure de vraisemblablement deux personnes et d'une fraction mineure pas interprétable a été obtenue à partir du prélèvement effectué sur le manche de la hache. Le profil ADN du POI est compatible avec le profil ADN de la trace. Une quantité totale d'ADN de 0.054 ng dont 17 % correspond au profil du POI est obtenue en utilisant le logiciel STRmix™. Des quantités de 0.009 ng d'ADN correspondant au profil du POI et de 0.045 ng d'ADN différent du profil du POI sont donc obtenues. Nos observations sont alors :

- $E_1$  : 0.00-0.01 ng
- $E_2$  : 0.04-0.05 ng

Les observations sont donc différentes de celles obtenues pour le cas de M. X. Cependant, le réseau bayésien construit pour le cas de M. X peut être exploité pour évaluer les résultats de ce nouveau cas. Les données utilisées pour la paramétrisation du BN sont jugées adéquates pour ce nouveau cas. Nous pouvons donc exploiter les résultats obtenus lorsque les données

---

spécifiques au POI, à l'AO et au bruit de fond ainsi que les données d'efficacité d'extraction et d'efficacité de prélèvement sont prises en compte (Chapitre V), présentés dans la Figure 31. Ils peuvent également être visualisés dans « Part 3 : Results of cases » sur le site [https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/).



**Figure 31 : Variation des log<sub>10</sub>(LRs) calibrés lorsque des quantités entre 0.00 et 0.01, et 0.04 et 0.05 ng d'ADN correspondant au profil ADN du POI, respectivement de l'AO sont obtenues.**

Dans cette situation, un LR médian calibré et robuste, au sens des critères adoptés dans ce travail, de 0.4 est attendu si le POI est meilleur donneur que l'AO. En revanche, un LR calibré et non robuste de l'ordre de 0.2 est attendu si le POI est moins bon donneur que l'AO.

Ne connaissant pas le statut du donneur du POI par rapport à la population générale, il faut investiguer le nombre d'expériences nécessaires pour conduire l'évaluation.

Si des simulations par ré-échantillonnage sont faites en augmentant le nombre de données associées aux variables d'intérêt (voir chapitre II), il est possible de déterminer si l'apport de données supplémentaires permet de diminuer la variation du LR et déterminer également le nombre de données supplémentaires nécessaires.

Si l'ajout de données d'une ou des variables d'intérêt conduit à une diminution de la variation des  $\log_{10}(\text{LRs})$  jusqu'à ce que l'étendue interquartile des  $\log_{10}(\text{LRs})$  par rapport au  $\log_{10}(\text{LR})$  médian n'excède pas le seuil défini de 0.8, alors il est utile d'acquérir des données supplémentaires pour cette ou ces variables d'intérêt.

La connaissance de ce nombre minimum d'expériences qui seraient à effectuer et du LR potentiellement obtenu (0,4 ou 0,2 en fonction du statut du donneur par rapport à l'AO) permet d'aider à se déterminer sur l'apport du résultat de cette l'évaluation dans l'affaire et donc de se décider sur la réalisation ou non de celle-ci, avant d'acquérir des données.

La conclusion de la pré-évaluation ne tient compte ici que des résultats obtenus dans ce travail. Cependant, cette décision devrait être prise en tenant compte également des moyens disponibles pour le cas d'espèce.

Il arrive que les questions des activités soient posées une fois au tribunal et non dans le mandat d'expertise. Il sera alors difficile de répondre sachant qu'une évaluation aussi détaillée et potentiellement des expériences devraient être réalisées. Cependant, il serait rapide et facile de visualiser les résultats obtenus à partir d'un cas similaire en utilisant l'application shiny et d'ainsi faire une pré-évaluation afin de guider, quant aux attentes, dans le cas d'espèce.

Pour certaines observations, un LR inférieur à 1 est obtenu si le POI est meilleur donneur que l'AO alors qu'un LR supérieur à 1 est obtenu si le POI est moins bon donneur que l'AO. Ceci illustre alors la nécessité d'acquérir les données spécifiques au POI et peut constituer un argument pour que les expériences puissent être entreprises.

#### VI.4 Limitations pratiques de la méthode

La robustesse du LR obtenu et des données concernant l'AO ont été discutées précédemment. Nous avons identifié quelques limitations pratiques additionnelles :

- (1) Le besoin d'avoir des hypothèses définies.
- (2) L'investissement nécessaire à l'élaboration de tout nouveau BN.
- (3) La calibration du système.
- (4) La gestion d'impondérables. Nous avons observé que les variables « Quantité d'ADN sur les mains » et « Proportion de transfert » nécessitent des données spécifiques. Or ces données dépendent de circonstances qui peuvent être inconnues comme par exemple le port de gants. Comment gérer cet impondérable ?

(5) Les conditions d'expériences pour acquérir les données.

(6) L'application à un nombre restreint de cas

#### VI.4.1 Besoin d'hypothèses bien définies

Nous avons vu que les résultats d'une expertise sont appréciés au regard d'au moins deux propositions alternatives, mutuellement exclusives (Evet, 1998), proposées par les différentes parties et qui dépendent des circonstances du cas (Buckleton et al., 2014). En pratique, les parties, lorsqu'elles sont confrontées à des problématiques liées aux phénomènes de transfert ou de persistance, se posent la question de savoir si de tels événements sont possibles. Par exemple, l'une des parties demandera : « Est-il possible d'observer un profil ADN s'il y a eu un transfert secondaire ». Cependant, le transfert secondaire est un facteur qui est pris en compte dans l'évaluation des résultats. D'après la recommandation 6 de la commission ADN de la Société internationale de génétique forensique (Gill et al., 2020), les résultats ou facteurs qui sont pris en compte dans l'évaluation ne doivent pas être imbriqués dans les propositions. De plus, la réponse à cette formulation : « Est-il possible... ? » sera toujours oui. Gill et al., (2020) font plusieurs recommandations sur la manière dont ces propositions au niveau de l'activité devraient être indiquées. Cependant, nous pouvons nous interroger sur la façon de poser ces propositions si celles-ci ne suivent pas les recommandations, ce qui est le cas le plus fréquemment rencontré en pratique à l'heure actuelle.

Nous sommes d'avis qu'en l'absence de propositions correctement posées par les différentes parties, une première étape serait de pouvoir définir en commun les propositions en fonction des différentes allégations.

Si cette discussion ne peut pas se faire, nous suggérons de répondre à leurs questions et de leur expliquer pourquoi il est essentiel d'avoir des propositions définies telles que recommandées par Gill et al., (2020). Puis, s'il y a suffisamment d'informations à disposition sur le cas, de poser des propositions en partant de présupposés et en indiquant que, si ces informations venaient à changer, une nouvelle évaluation devrait être faite. S'il n'y a pas suffisamment d'informations, nous serions d'avis d'indiquer qu'une évaluation ne peut être faite en l'absence de propositions et d'informations supplémentaires.

#### VI.4.2 Construction d'un réseau bayésien et effectuer son analyse de sensibilité

Nous avons vu que le BN proposé dans ce travail peut être facilement adapté à d'autres cas et que les résultats des analyses de sensibilité et de calibration peuvent être utilisés pour de nombreux cas également. En revanche, un nouveau cas nécessitant la construction d'un nouveau BN signifie forcément la construction du BN, d'une nouvelle analyse de sensibilité, et d'une calibration du système.

Ces étapes demandent toutes un investissement non négligeable.

##### Construction d'un BN

Les étapes de la construction d'un BN sont détaillées par Taylor et al. (2018). Pour rappel, ils ont défini 7 étapes, déjà présentées par Aitken et al. (1996), permettant de construire un BN utilisable pour les cas de biologie forensique comportant des propositions au niveau de l'activité :

- définir les propositions,
- définir les activités,
- définir les groupes de variables dont les observations sont similaires,
- définir les observations,
- définir les variables du transfert et de la persistance,
- définir les relations entre les nœuds,
- informer le réseau et vérifier la cohérence de celui-ci.

Néanmoins, des connaissances approfondies sont nécessaires à la compréhension du fonctionnement d'un BN et sont donc nécessaires afin d'être en mesure de construire un BN cohérent. La mise en œuvre de cette méthode requiert une formation spécifique.

##### Analyse de sensibilité

L'analyse de sensibilité, telle que proposée dans ce travail nécessite des connaissances de programmation dans R ou un logiciel d'analyse.

Il est possible d'imaginer une interface dans laquelle il serait réalisable de sélectionner le BN adapté à son cas à partir d'une base de données de BN. Cette base de données de BN contiendrait différents BN pour chacun des scénarios les plus rencontrés et dont une analyse de sensibilité aurait été effectuée au préalable. Les résultats des étapes 1 et 2 seraient donc à disposition via cette interface. A partir de celle-ci, il serait également possible de construire

un nouveau BN, de sélectionner des données pour l'informer et d'effectuer l'analyse de sensibilité par un système de ré-échantillonnage automatique dépendant des données sélectionnées pour informer le BN. Cet environnement intégré n'existe pas aujourd'hui.

#### VI.4.3 Calibration du système

Réaliser une calibration pour chaque nouveau BN et chaque nouveau jeu de proposition est difficile et constitue une contrainte non négligeable. Elle nécessite de reproduire un certain nombre de cas suivant chacune des propositions.

Cependant, ces données peuvent être reprises de celles issues de l'étape de l'acquisition des données. De plus, une solution qui faciliterait la réalisation de cette calibration serait d'avoir à disposition dans cette interface, les résultats de la calibration obtenus à partir d'autres cas similaires par d'autres laboratoires forensiques en suivant cette méthodologie. Nous avons également vu (cf. partie V.4.1) qu'afin d'éviter de faire cette calibration, il faudrait étudier les distributions sous-jacentes plus en détails, afin que les probabilités représentent de manière calibrée ce qui est observé.

#### VI.4.4 Gestion d'impondérables : Exemple du port des gants

Le potentiel port des gants est une situation qui n'a pas été étudiée dans ce travail.

Si les éléments d'enquête indiquent l'utilisation possible de gants comme des traces de gants, ce paramètre doit être pris en compte aussi bien pour le POI que pour l'AO. Le BN devra donc être adapté à ce scénario spécifique.

Si aucun élément d'enquête ne permet de se déterminer sur l'utilisation de gants, deux options sont possibles :

- prendre en compte les gants quelle que soit la proposition des parties,
- se baser sur les propositions fournies sans référence à l'usage des gants.

En prenant en compte l'utilisation de gants, il est attendu que les observations soient aussi probables sous une proposition que sous l'autre pour un plus grand nombre de cas. Cependant si aucun élément d'enquête ne permet de définir si des gants ont pu être utilisés, il est préférable de se baser uniquement sur ce que les parties allèguent sans considérer toutes les options alternatives.



#### VI.4.5 Expériences : POI, type de surface et type de contact

Nous avons établi qu'afin d'évaluer les résultats ADN, des expériences spécifiques au POI sont nécessaires. La question des modalités si le POI n'est pas disponible, de l'action à reproduire ainsi que de l'objet à utiliser pour acquérir les données, se pose. Nous les passons en revue ci-après.

##### **Absence du POI**

Une possibilité serait d'effectuer les expériences dans les deux situations les plus extrêmes, c'est-à-dire de réaliser les expériences avec le meilleur donneur à disposition par rapport à la population générale avec le moins bon donneur à disposition par rapport à la population générale et inversement, comme cela a été fait dans ce travail.

Si un LR inférieure à 1 est obtenu lorsque le POI est meilleur donneur que l'AO alors qu'un LR supérieur à 1 est obtenu lorsque le POI est moins bon donneur que l'AO, alors la nécessité d'acquérir les données spécifiques au POI sera démontrée. Si cela n'est vraiment pas possible, alors aucune évaluation ne pourra être fournie au tribunal.

##### **Action à reproduire**

La littérature montre que le type de contact (pression, passif, friction) impacte le transfert (Goray et al., 2010). Dans la mesure du possible, il est conseillé de reproduire le plus fidèlement possible l'action alléguée. Si l'activité alléguée ne peut pas être reproduite, il faut que l'activité faite pour les expériences mette en œuvre un type de contact similaire à celui qui a été opéré durant l'activité alléguée.

##### **Objet à utiliser**

La littérature indique que le type de surface peut avoir une influence sur les observations ADN, en terme de quantité d'ADN et de qualité des profils ADN (Daly et al., 2012 ; Goray et al., 2010 ; Fonnøløp A. et al., 2015), en particulier si les surfaces sont radicalement différentes, par exemple un tissu et une surface en verre. Cependant cette influence serait moindre dans le cas de deux types de surfaces différentes mais faites dans le même matériau, telles que deux surfaces en plastique, l'une plutôt lisse et l'autre plutôt rugueuse (Cale et al., 2015).

Dans la mesure du possible, il est conseillé d'utiliser un objet du même modèle que l'objet en question pour mener des expériences. En revanche si cela n'est pas possible, un objet équivalent de même matériau peut être utilisé.

#### VI.4.6 L'application à un nombre restreint de cas

Malgré des formations sur la construction et le fonctionnement d'un BN et malgré la réutilisation de certains résultats de l'analyse de sensibilité et de certaines données, l'investissement reste important. La question des situations pour lesquelles cette méthodologie peut être utilisée en routine dans un laboratoire, se pose. Théoriquement, nous sommes d'avis que l'évaluation des résultats ADN considérant des propositions au niveau de l'activité en utilisant cette méthodologie devrait se faire, comme indiqué par le guide de l'ENFSI sur la rédaction des rapports évaluatifs (Willis et al., 2015), pour tous les cas impliquant des traces de contact. Or, ces cas représentaient en 2018 environ 80% de l'ensemble de traces analysées au Centre Universitaire Romand de Médecine Légale (CURML) de Lausanne (3645 traces de contact sur 4624 traces analysées d'après leur base de données). D'un point de vue pratique, il est donc difficile d'évaluer tous ces cas en considérant des propositions au niveau de l'activité et une sélection doit être envisagée. Une option serait dans un premier temps de faire une évaluation en considérant des propositions au niveau de la source à des fins d'investigation puis, dans un deuxième temps, de faire une évaluation en considérant des propositions au niveau de l'activité seulement pour des cas jugés au tribunal.

Si cette option avait été choisie, dans l'affaire Anderson (Osagie, 2013) (Voir la partie I.1), l'origine de la trace aurait été considérée dans un premier temps. Dans un deuxième temps, le Procureur aurait mandaté le laboratoire pour une évaluation en considérant des propositions au niveau de l'activité, alibi ou non. En effet, si M. Anderson n'avait pas eu un « alibi », le procureur aurait sans doute poursuivi son action, persuadé que l'ADN sous les ongles de la victime ne pouvait s'y trouver que suite à l'agression. L'évaluation considérant des propositions au niveau de l'activité aurait permis d'éviter ce raisonnement.

Cependant, sachant que malgré son alibi, M. Anderson a passé 5 mois en prison, il est difficile d'imaginer que, malgré une évaluation en considérant des activités, M. Anderson n'aurait pas été considéré comme prévenu et ne serait pas passé par la prison.

## VI.5 Perspectives

A la lumière des travaux entrepris dans cette recherche, des opportunités et des limitations identifiées, les perspectives suivantes peuvent être envisagées.

### VI.5.1 Problématique de l'AO

L'AO étant inconnu, une des options proposées est d'utiliser les données de la quantité d'ADN sur les mains et de la proportion de transfert représentant la variabilité entre chaque individu d'une population générale.

Si les données sur les efficacités d'extraction et de méthode de prélèvement sont connues et fixes entre laboratoires, il est possible de regrouper les données, au fur et à mesure des nouvelles expertises dans un seul jeu, quel que soit le laboratoire ayant acquis ces données.

De plus, grâce à cette base de données, il serait également intéressant d'étudier l'impact de l'origine ethnique sur le transfert d'ADN et si cette variable a un impact sur le LR.

### VI.5.2 Besoin de formation

Prenons les propositions suivantes, au niveau de la source :

- La personne d'intérêt est à la source de la trace
- Une autre personne, non apparentée à la personne d'intérêt, est à la source de la trace

Les LR obtenus, lorsqu'un profil ADN correspond au profil ADN de la personne sur plus de 10 loci est de l'ordre du milliard (Foreman et Evett, 2001, Hopwood et al., 2012). Les évaluations en considérant des propositions au niveau de la source étant plus courantes que les évaluations au niveau de l'activité, nous sommes donc habitués à pouvoir obtenir des LR de l'ordre du milliard. Or, dans la section V.3.3, les LR obtenus pour les 120 cas dont les résultats sont évalués en considérant les activités sont compris entre l'ordre de 0.00001 et 10'000. Il existe alors le risque qu'un LR de 10 000 soit perçu comme un LR représentant un poids indiciaire plus faible que ce qu'il est en réalité.

Afin d'éviter cet effet, il est essentiel de former les personnes du système judiciaire afin qu'elles puissent être à même de comprendre ce LR et les enjeux de l'évaluation d'une trace quelle qu'elle soit. Cette formation permettrait également de sensibiliser les différentes parties sur le besoin d'avoir des propositions et de savoir comment les formuler, que ce soit au niveau de la (sous)-source ou de l'activité.

### VI.5.3 Partage d'information

Au-delà de cette base de données sur l'AO, il serait utile de pouvoir combiner l'ensemble des résultats obtenus par tous les laboratoires pour toutes les expertises. Une volonté de partager les données obtenues par les différents laboratoires existe déjà dans la communauté scientifique (Kokshroorn et al., 2018). Pour cela, il est possible d'envisager un système expert permettant de regrouper l'ensemble des données existantes et d'évaluer les résultats ADN considérant des propositions au niveau de l'activité.

Ce système permettrait de :

- sélectionner le BN adapté à partir d'une base de données de BN,
- de sélectionner les données informant le BN en indiquant le type d'objet en question associé aux données de bruit de fond qui nous intéresse, le kit d'extraction, la méthode de prélèvement, les surfaces d'intérêt, les conditions environnementales, le temps écoulé entre les différentes actions,
- de faire une analyse de sensibilité par ré-échantillonnage si celle-ci n'a pas déjà été faite,
- de visualiser les résultats des analyses de sensibilité du BN sélectionné,
- d'importer et conserver les données spécifiques du cas,
- d'importer ou sélectionner les paramètres de calibration,
- de visualiser les résultats finaux.

Le regroupement de l'ensemble des connaissances dans un système intégré aurait pour avantage :

- d'avoir à disposition une base de données de BN pour différents « types de cas »,
- d'avoir à disposition des bases de données permettant d'informer les variables, y compris la base de données AO,
- d'avoir à disposition une base de données sur l'efficacité des différentes méthodes de prélèvements et des kits d'extraction ADN,
- d'augmenter les connaissances à disposition pour les variables d'intérêt,
- d'établir des pré-évaluations pour tous les types de cas,
- de visualiser les résultats qui peuvent faciliter la compréhension de ceux-ci au tribunal,
- de constituer une base de connaissances centralisées à disposition des experts.

Le système proposé dans ce travail est un KBS (« Knowledge-based system »). Un KBS est un programme informatique qui permet de résoudre des problèmes complexes à partir de bases de connaissances. Des exemples de KBS utilisés en sciences forensiques sont présentés par Curran (2009).

Une ébauche d'un tel système expert, a été mise en place dans le domaine des microtraces par Pinchin et Buckleton (1988) en créant leur logiciel nommé CAGE (« Computerised Assistance for Glass Evidence ») permettant à des non spécialistes d'évaluer les traces de verre. Cependant, celui-ci n'a pas été implémenté en routine. Plusieurs propositions de projets afin de développer ce type de système expert dans les sciences forensiques ont été soumises sans avoir pu aboutir (communication personnelle) mais qui soulignent le besoin de développer de tels systèmes.

#### VI.5.4 Sensibilité des analyses ADN

Les techniques étant de plus en plus sensibles, de potentiels contributeurs à la trace, même sans lien avec les faits en examen, deviennent détectables du fait de transferts complexes. De plus, en dessous d'une certaine quantité d'ADN, dépendante des instruments utilisés, de nombreux artefacts peuvent être observés, comme par exemple des pics désignés comme étant des allèles alors qu'ils n'en sont pas (drop in) ou des pics qui ne sont pas désignés comme étant des allèles alors qu'ils devraient l'être (drop out). Il apparaît alors difficile de savoir si ce qui est observé provient de la trace du contact direct d'intérêt ou d'artefacts.

Certains laboratoires utilisent un seuil de quantité en dessous duquel ils n'effectuent pas d'analyse limitant ainsi la possibilité d'avoir ces artefacts.

Champod (2013) pose une question plus générale à savoir si l'évaluation des résultats au niveau de l'activité serait plus utile au tribunal en phase évaluative ou bien à l'enquête dans sa phase d'investigation si les techniques continuent à être de plus en plus sensibles. En effet, la plus grande sensibilité des techniques augmente la possibilité d'observer des artefacts et des profils ADN résultant de transferts complexes, entraînant ainsi une augmentation de l'incertitude sur les observations. Aussi, l'évaluation avec des propositions au niveau de l'activité risque de ne plus permettre d'aider le tribunal. Cependant, l'information résultante de ces évaluations pourrait être utile en phase d'investigation pour orienter l'enquête et écarter ou consolider certaines pistes.

Concernant les perspectives, nous pouvons donc nous poser la question sur l'opportunité d'analyser et d'évaluer des traces de quantités de plus en plus infimes (Champod, 2013 ; Benschop et al., 2015). Nous sommes d'avis que, d'un point de vue de l'évaluation des résultats ADN, tant qu'il est difficile d'exploiter les résultats ADN obtenus à partir de ces traces ADN en quantité infime, augmenter la sensibilité des techniques ne serait d'aucune utilité. Il serait préférable de se concentrer d'abord sur l'amélioration de la méthodologie proposée dans ce travail et son développement à d'autres cas impliquant, par exemple, des traces de sperme, de sécrétions vaginales ou de salive, riches en ADN.

#### VI.5.5 Investiguer le scénario où un contact légitime est allégué par la défense

Nous avons observé un fort taux de cas dont un LR non robuste est obtenu lorsque la défense allègue un contact précédent avec l'objet. Cette proposition de la défense étant posée en pratique, il serait utile d'étudier ce scénario plus en détail. Lors d'un contact précédent le phénomène de transfert d'ADN de la surface cible vers la surface donneuse peut potentiellement avoir un impact dans ce scénario et devrait être pris en compte dans le BN.

#### VI.5.6 Intégrer les erreurs de laboratoire

Nous avons vu dans la partie II.1.6 que les erreurs de laboratoire, dues à des erreurs humaines, constituent des événements très particuliers, susceptibles d'arriver mais qui, dans la plupart des cas, peuvent être mis en évidence par les éléments de l'enquête et/ou résultats de laboratoire et par les incohérences qui résultent de cette erreur.

Pour ce travail, les erreurs de laboratoire n'ont pas été prises en compte afin de se concentrer essentiellement sur la problématique des propositions au niveau de l'activité. Cependant, il serait nécessaire de pouvoir les intégrer à l'évaluation des résultats ADN que les propositions soient au niveau de la (sous)-source ou de l'activité.



## VII - Conclusion

---

Le guide de l'ENFSI sur la rédaction des rapports évaluatifs (Willis et al., 2015) préconise d'évaluer les résultats ADN considérant des propositions au niveau de l'activité notamment quand l'expert se doit de considérer des facteurs, tels que le transfert, la persistance ou le bruit de fond, qui peuvent avoir un impact sur l'interprétation des résultats relatifs aux activités alléguées. Ceci est particulièrement le cas pour des traces constituées d'infimes quantités d'ADN, qui sont des traces de plus en plus prélevées.

Or, dans la pratique, les experts analysent et évaluent leurs observations principalement au niveau de la (sous-) source et rarement au niveau de l'activité et ceci pour trois raisons principales :

- (1) Les évaluations considérant des propositions au niveau de l'activité sont rarement demandées.
- (2) Les variables intervenant dans les phénomènes de transfert, de persistance et de bruit de fond sont nombreuses et il est difficile de concevoir un plan expérimental qui permettrait de prendre en compte tous les facteurs pouvant intervenir dans un cas réel.
- (3) L'incertitude quant aux types de données à utiliser, les données issues de nos connaissances ou d'expériences dans des conditions contrôlées ne reflètent pas la réalité ou sont en quantité limitée.

La méthodologie développée et proposée par ce travail permet de passer outre les contraintes (2) et (3) précédemment citées. Cette méthodologie repose sur l'élaboration d'un réseau bayésien dans lequel les interdépendances entre les observations ADN, les propositions alléguées ainsi que les différentes variables intervenant dans le transfert, la persistance, le bruit de fond sont modélisées. Il permet également de tenir compte de la dépendance entre les variables.

Afin d'établir une méthodologie, deux problématiques ont été plus particulièrement étudiées :

- 1- Comment identifier les variables impactant l'appréciation permettant ainsi de limiter ou contraindre l'acquisition de données structurées ?

Les variables ayant une influence sur le transfert et la persistance des traces ainsi que sur les observations sont nombreuses.



Cependant, toutes n'ont pas la même importance sur le résultat de l'évaluation. Nous avons montré, grâce à une analyse de sensibilité du BN par des simulations par ré-échantillonnage, que certaines variables ont un impact plus grand sur le rapport de vraisemblance (LR) que d'autres (Chapitre II).

Nous avons posé les définitions permettant l'identification de ces variables.

C'est donc sur ces variables qu'il faut se concentrer et acquérir, si nécessaire, des données supplémentaires et/ou spécifiques.

Celles-ci sont dépendantes des observations obtenues et des propositions utilisées. Cependant elles ne dépendent pas des circonstances du cas.

## 2- Dans quelle mesure devons-nous envisager l'acquisition de données ?

L'hypothèse sous-jacente est que le LR qui devrait être utilisé dérive de l'utilisation de données spécifiques. Mais peut être que des données issues de la littérature peuvent être utilisées. Les données « spécifiques » au cas en question sont des données qui correspondent au cas (les données proviennent d'expériences reproduisant les circonstances du cas effectuées avec le POI) et aux moyens utilisés pour analyser la trace ADN (les données utilisées sont celles issues du kit d'extraction et du moyen de prélèvement utilisé).

Or, les résultats présentés au chapitre IV montrent qu'il est préférable d'utiliser des données spécifiques pour les variables « Quantité d'ADN sur les mains » et « Proportion de transfert », celles-ci ayant le plus d'impact sur la valeur du LR. Les données de ces variables étant dépendantes de données des variables « Efficacité d'Extraction » et « Efficacité de prélèvement », il est conseillé d'utiliser également des données d'efficacité spécifiques du kit d'extraction et de la méthode de prélèvement utilisés. Les circonstances du cas sont utilisées pour informer les variables instanciées.

Suite à ces résultats, le protocole d'une méthodologie a été posé. La méthodologie élaborée se décompose en plusieurs étapes résumées ci-après :

- Etape 1 : construction d'un réseau bayésien (BN) : il s'agit d'élaborer un modèle probabiliste qui tient compte, des propositions, des observations et de toutes les variables qui influencent les observations. Le BN permet de calculer un LR à partir des données disponibles dans la littérature et /ou issues d'expériences.

- Etape 2 : identification des variables ayant un impact sur le LR à partir d'une analyse de sensibilité du BN effectuée par des simulations par ré-échantillonnage. Cette analyse permet de déterminer les variables ayant le plus d'impact sur le LR. Dans le cas du scénario 1 (Transfert secondaire allégué), il faut acquérir au minimum les données de « Quantité d'ADN sur les mains » et « Proportion de transfert » relatif au POI ainsi que les données d'efficacité du kit d'extraction et de la méthode de prélèvement si cela n'a pas été fait dans les cas précédents. Cette étape permet également de déterminer si le LR obtenu avec les données à disposition est robuste et peut donc être reporté. Les critères définissant la robustesse d'un LR ont été posés.
- Etape 3 : définir, d'après les résultats des simulations, le nombre de données supplémentaires à acquérir si cela est nécessaire.
- Etape 4 : acquérir ces données supplémentaires.
- Etape 5 : déterminer l'impact des données sur le LR afin de déterminer si le LR est robuste et peut donc être reporté.

Cette méthodologie a été appliquée pour des cas fictifs d'agressions aux couteaux élaborés en laboratoire pour deux scénarios différents :

- Le premier :
  - a. Hp : « Le POI a poignardé la victime »
  - b. Hd1 : « Le POI a serré la main de l'AO qui a poignardé la victime avec ce couteau »

Lorsque la proposition du procureur est vraie, dans 81% des cas, les observations soutiennent la proposition du procureur par rapport à celle de la défense. Elles soutiennent la mauvaise proposition dans 9,5% des cas et le LR ne peut être reporté dans 9,5% des cas.

Lorsque la proposition de la défense est vraie, les observations soutiennent la proposition de la défense par rapport à celle du procureur dans 39% des cas. Elles soutiennent la mauvaise proposition dans 8% des cas. Le LR ne peut être reporté dans 53% des cas.

- Le deuxième :
  - a. Hp : « Le POI a poignardé la victime »
  - b. Hd3 : « Quelqu'un d'autre a poignardé la victime, le POI a utilisé le couteau »

le midi ».

Lorsque la proposition du procureur est vraie, les observations soutiennent la proposition du procureur par rapport à celle de la défense dans 45% des cas seulement. Elles soutiennent la mauvaise proposition dans 21% des cas. Le LR ne peut être reporté dans 34% des cas.

Lorsque la proposition de la défense est vraie, les observations soutiennent la proposition de la défense par rapport à celle du procureur dans 82% des cas. Elles soutiennent la mauvaise proposition dans 5% des cas. Le LR ne peut être reporté dans 13% des cas.

Ces résultats d'efficacité dépendent du statut du donneur du POI par rapport à l'AO, c'est pourquoi il est essentiel d'acquérir à minima des données relatives au transfert d'ADN du POI.

Les ratios de quantités d'ADN permettent d'évaluer le risque de transmettre un LR qui pourrait induire en erreur.

Cette méthodologie se veut **applicable à de nombreux cas** : le raisonnement de construction du BN permet une généralisation, cependant les quantités initiales d'ADN doivent être connues.

Elle définit également un mécanisme permettant de **déterminer la robustesse du LR** obtenu et donc s'il peut être reporté ou non.

Elle apporte **une solution opérationnelle** : la méthodologie proposée ne requiert que l'acquisition de données spécifiques au cas relatives au statut de la personne d'intérêt (POI). Dans la mesure du possible, il est conseillé de reproduire le plus fidèlement possible l'action alléguée avec le POI en utilisant l'objet en question. Si cela n'est pas possible, il faut que l'activité choisie pour les expériences reproduise le même type de contact que l'activité alléguée, en utilisant un objet de même type dont la surface est de même nature que l'objet d'intérêt. Elle requiert également d'effectuer une calibration pour chaque BN et chaque jeu de propositions.

Ce travail apporte quatre contributions principales :

- La démonstration que les variables liées au laboratoire, « Efficacité de la méthode de prélèvement » et « Efficacité du kit d'extraction », impactent le LR.

- La démonstration qu'il est pertinent et nécessaire de prendre en compte dans l'évaluation l'information concernant les fractions majeures/fractions mineures dans un profil ADN grâce à l'observation constituée de deux éléments  $E_1$  (quantité d'ADN dont le profil correspond au profil ADN du POI) et  $E_2$  (Quantité d'ADN dont le profil ADN est différent de celui du POI).
- Les prémisses d'un système intégré voire d'un système expert sont posées au travers d'une application [https://lydie-samie.shinyapps.io/DNA\\_Activity\\_PhD/](https://lydie-samie.shinyapps.io/DNA_Activity_PhD/) qui permet la visualisation rapide des résultats obtenus pour chaque observation possible. Elle en facilite la compréhension également. Elle peut être réutilisée pour certains cas dont le BN et les données utilisées pour la paramétrisation sont adaptés.
- Une pré-évaluation de ces cas qui peut être facilement réalisée grâce à cette application.

Cependant certaines limitations existent, qui résultent souvent des impondérables des circonstances d'un dossier. En effet, l'AO est inconnu. Cette thèse propose de considérer les données de variabilité intra individu des individus composant une population générale. Ceci permettra de tenir compte de la possibilité que le POI soit meilleur ou moins bon donneur que l'AO dans une certaine population. De plus le fait qu'un transfert secondaire a effectivement eu lieu, ou que l'agresseur a utilisé des gants, est inconnu. Deux options sont proposées concernant cette problématique :

- prendre en compte par défaut la possibilité d'un transfert secondaire, ou le port les gants, quelle que soit la proposition des parties,
- se baser sur les propositions et les circonstances fournies.

La perspective principale de cette thèse est de regrouper l'ensemble des informations (données d'expériences, résultats des simulations et BN) dans une plateforme interactive qui permettrait d'effectuer des pré-évaluations et des évaluations pour un très grand nombre de cas et de limiter encore plus le nombre d'expériences à effectuer à chaque nouveau cas. Seules les données spécifiques des variables « Quantités d'ADN sur les mains » et « Proportion de transfert » pour le POI seraient alors à acquérir.



## Bibliographie

---

Aitken C.G.G., Taroni F., *Statistics and the evaluation of evidence for forensic scientists*, John Wiley & Sons, Ltd, Chichester, UK (2004).

Alessandrini F., Cecati M., Pesaresi M., Turchi C., Carle F., Tagliabracci A., *Fingerprints as evidence for a genetic profile: morphological study on fingerprints and analysis of exogenous and individual factors affecting DNA typing*, *Journal of Forensic Sciences*. 48 (2003) 586-592.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W., Silverman, E. *An empirical distribution function for sampling with incomplete information*, *The Annals of Mathematical Statistics*. 26 (1955) 641–647.

Basset P., Castella V., *Lessons from a study of DNA contaminations from police services and forensic laboratories in Switzerland*, *Forensic Science International: Genetics* 33 (2018) 147–154.

Basset P., Castella V., *Positive impact of DNA contamination minimization procedures taken within the laboratory*, *Forensic Science International: Genetics* 38 (2019) 232–235.

Biedermann A., Taroni F., *Bayesian networks for evaluating forensic DNA profiling evidence: a review and guide to literature*, *Forensic Science International: Genetics*. 6 (2012) 147–157.

Biedermann A., Champod C., Jackson G., Gill P., Taylor D., Butler J., Morling N., Hicks T., Vuille J., Taroni F., *Evaluation of Forensic DNA Traces When Propositions of Interest Relate to Activities: Analysis and Discussion of Recurrent Concerns*, *Frontiers in Genetics*. 7 (2016) 1–12.

Benschop C.C.G., Haned H., Yoo S.Y., Sijen T., *Evaluation of samples comprising minute amounts of DNA*, *Science & Justice*. 55 (2015) 316–322.

Bontadelli L. *Study of DNA shedder quality*, Msc thesis, School of Criminal Justice of Lausanne, University of Lausanne. Lausanne; (2009).

Breathnach M., Williams L., McKenna L., Moore E., *Probability of detection of DNA deposited by habitual wearer and/or the second individual who touched the garment*, *Forensic Science International: Genetics*. 20 (2016) 53–60.

Bright J.-A., Petricevic S.F., *Recovery of trace DNA and its application to DNA profiling of shoe insoles*, *Forensic Science International*. 145 (2004) 7–12.

Buckleton J., Bright J.-A., Taylor D., Evett I., Hicks T., Jackson G., Curran J.M., *Helping formulate propositions in forensic DNA analysis*, *Science & Justice*. 54 (2014) 258–261.

Bundesamt für statistik, *Polizeiliche Kriminalstatistik (PKS) Jahresbericht 2012, 2013*.

## Evaluation des résultats ADN considérant des propositions au niveau de l'activité 2019

---

Bussard G., Étude du transfert de l'ADN par les gants dans les cas de vol de véhicule. Évaluation des risques de contamination lors du prélèvement d'une trace de contact, Travail de master, Ecole des Sciences Criminelles, Université de Lausanne. Lausanne, 2008.

Cale C.M., Forensic DNA evidence is not infallible, *Nature*. 526 (2015) 611.

Cale C.M., Earll M.E., Latham K.E., Bush G.L., Could secondary DNA transfer falsely place someone at the scene of a crime?, *Journal of Forensic Sciences*. 61 (2015) 196–203.

Cemper-Kiesslich J., Tutsch-Bauer E., Neuhuber F., Another phantom from the morgue—a case of instrument-born sample contamination in the course of identifying an unknown deceased, *Forensic Science International: Genetics*. 7 (2013) 405–407.

Champod C., DNA transfer: informed judgment or mere guesswork?, *Frontiers in Genetics*. 4 (2013) 300 in Biedermann A., Vuille J., Taroni F., *DNA, Statistics and the Law: A Cross-Disciplinary Approach to Forensic Inference*, *Frontiers Media*. 4 (2014) 22-24.

Coquoz R., Comte J., Hall D., Hicks T., Taroni F., *Preuve par l'ADN: la génétique au service de la justice*, Presses polytechniques et universitaires romandes, Lausanne, Switzerland 3ème édition (2013).

Cook R., Evett I.W., Jackson G., Jones P.J., Lambert J.A., A hierarchy of propositions: deciding which level to address in casework, *Science & Justice*. 38 (1998) 231–239.

Cook O., Dixon L., The prevalence of mixed DNA profiles in fingernail samples taken from individuals in the general population, *Forensic Science International: Genetics*. 1 (2007) 62–68.

Curran J.M., *Use of Knowledge-Based Systems in Forensic Science*, *Wiley Encyclopedia of Forensic Science*, John Wiley & Sons (2009).

Daly D.J., Murphy C., McDermott S.D., The transfer of touch DNA from hands to glass, fabric and wood, *Forensic Science International: Genetics*. 6 (2012) 41–46.

David G., Clayson N., Jones S., Lewis J., Boyce M., Fraser I., et al., A response to Meakin and Jamieson DNA transfer: Review and implications for casework, *Forensic Science International: Genetics*. 21 (2016) 117–118.

De Donno M., *De la validation de calculateurs de rapports de vraisemblance en sciences forensiques- Application au cas de la dactyloscopie*. Mémoire de thèse. Ecole des Sciences Criminelles. Université de Lausanne. Lausanne, 2016.

Dufresne D., The Beta Product Distribution with Complex Parameters, *Communications in Statistics —Theory Methods*. 39:5 (2010) 837–854.

Evett I.W., Lambert J.A., Buckleton J.S., A Bayesian approach to interpreting footwear marks in forensic casework, *Science & Justice*. 38 (1998) 241–247.

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

---

Evett I.W., Gill P.D, Jackson G., Whitaker J., Champod C., Interpreting small quantities of DNA: The hierarchy of propositions and the use of bayesian networks, *Journal of Forensic Sciences*. 47 (2002) 520–530.

Farmen R.K., Jaghø R., Cortez P., Frøyland E.S., Assessment of individual shedder status and implication for secondary DNA transfer, *Forensic Science International: Genetics Supplement Series*. 1 (2008) 415–417.

Flanagan N., McAlister C., The transfer and persistence of DNA under the fingernails following digital penetration of the vagina, *Forensic Science International: Genetics*. 5 (2011) 479–483.

Fonneløp A.E., Egeland T., Gill P., Secondary and subsequent DNA transfer during criminal investigation, *Forensic Science International: Genetics*. 17 (2015) 155–162.

Fonneløp A.E., Johannessen H., Gill P., Persistence and secondary transfer of DNA from previous users of equipment, *Forensic Science International: Genetics Supplement Series*. 5 (2015) 191–192.

Fonneløp A.E., Johannessen H., Egeland T., Gill P., Contamination during criminal investigation: Detecting police contamination and secondary DNA transfer from evidence bags, *Forensic Science International: Genetics Supplement Series*. 5 (2015) 191–192.

Fonneløp A.E., Ramse M., Egeland T., Gill P., The implications of shedder status and background DNA on direct and passive transfer in an attack scenario, *Forensic Science International: Genetics*. 29 (2017) 48–60.

Foreman L.A., Evett I.W., Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system, *International Journal of Legal Medicine*. 114 (2001) 147–155.

Gill P., The utility of 'substrate controls' in relation to 'contamination', *Forensic Science International*. 85 (1997) 105–111.

Gill P., A. Kirkham, Development of a simulation model to assess the impact of contamination in casework using STRs, *Journal of Forensic Sciences*. 49 (2004) 485–491.

Gill P., Hicks T., Butler J.M., Connolly E., Gusmão L., Kokshoorn B., Morling N., van Oorschot R.A.H, Parson W., Prinz M., Schneider P.M., Sijen T., Taylor D., DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence - Guidelines highlighting the importance of propositions. Part II: Evaluation of biological traces considering activity level propositions, *Forensic Science International: Genetics* 44 (2020) 102186.

Goray M., Eken E., Mitchell R.J., Van Oorschot R.A.H, Secondary DNA transfer of biological substances under varying test conditions, *Forensic Science International: Genetics*. 4 (2010) 62–67.

Goray M., Mitchell J.R., Van Oorschot R.A.H, Evaluation of multiple transfer of DNA using mock case scenarios, *Legal Medicine*. 14 (2012) 40–46.

Goray M., Van Oorschot R.A.H., Mitchell J.R., DNA transfer within forensic exhibit packaging: Potential for

---



DNA loss and relocation, *Forensic Science International: Genetics*. 6 (2012) 158–166.

Goray M., Van Oorschot R.A.H., DNA transfer during social interactions, *Forensic Science International: Genetics Supplement Series*. 4 (2013) e101–e102.

Goray M., Van Oorschot R.A.H., The complexities of DNA transfer during a social setting, *Legal Medicine*. 17 (2015) 82–91.

Goray M., Fowler S., Szkuta B., Van Oorschot R.A.H., Shedder status—An analysis of self and non-self DNA in multiple handprints deposited by the same individuals over time, *Forensic Science International: Genetics*. (2016) 190–196.

Graham E.A.M., Rutty G.N., Investigation into “normal” background DNA on adult necks: implications for DNA profiling of manual strangulation victims, *Journal of Forensic Sciences*. 53 (2008) 1074–1082.

Hesman Saey T., Shaking hands could transfer your DNA — leaving it on things you never touched ; People can transfer DNA from others after they touch them, potentially complicating a crime scene, new data show, *ScienceNewsforStudents*, 17.04.2019. <https://www.sciencenewsforstudents.org/article/long-handshake-spreads-your-dna-things-never-touched> (Dernier accès le 13.05.2019).

Hopwood A.J., Puch-Solis R., Tucker V.C., Curran J.M., Skerrett J., Pope S., Tully G., Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts, *Science and Justice*. 52 (2012) 185–190.

Hu.N, Cong B., Li S., Ma C., Fu L., Zhang X., Current developments in forensic interpretation of mixed DNA samples (Review), *Biomedical Reports*. 2 (2014) 309–316.

Jackson G., Jones S., Booth G., Champod C., Evett I.W., The nature of forensic science opinion—a possible framework to guide thinking and practice in investigation and in court proceedings, *Science & Justice*. 46 (2006) 33–44.

Jones S., Scott K., Lewis J., Davidson G., Allard J.E., Lowrie C., McBride B.M., McKenna L., Teppet G., Rogers C., Clayson N., Baird A., DNA transfer through nonintimate social contact, *Science & Justice*. 56 (2016) 90–95.

Kamphausen T., Fandel S.B., Gutmann J.S., Bajanowski T., Poetsch M., Everything clean? Transfer of DNA traces between textiles in the washtub, *International Journal of Legal Medicine*. 129 (2015) 709–14.

Kanokwongnuwut P., Martin B., Kirkbride K.P., Linacre A., Shedding light on shedders, *Forensic Science International: Genetics*. 36 (2018) 20–25.

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

---

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

---

Kokshoorn B., Blankers B.J., Zoete J., Berger C.E.H., Activity level DNA evidence evaluation: On propositions addressing the actor or the activity, *Forensic Science International* 278 (2017) 115–124.

Kokshoorn B., Aarts L.H.J., Ansell R., Connolly E., Drotz W., Kloosterman A.D., McKenna L.G., Szkuta B., van Oorschot R.A.H., Sharing data on DNA transfer, persistence, prevalence and recovery: Arguments for harmonization and standardization, *Forensic Science International: Genetics* 37 (2018) 260–269.

Konis K, RHugin: RHugin. R package version 8.4 (2017). <http://rhugin.r-forge.r-project.org> (Dernier accès le 22.08.2019).

Kumar P., Gupta R., Singh R., Jasuja O.P., Effects of latent fingerprint development reagents on subsequent forensic DNA typing: A review, *Journal of Forensic and Legal Medicine*. 32 (2015) 64-69.

Lacerenza D., Aneli S., Omedei M., Gino S., Pasino S., Berchiolla P., Robino C., A molecular exploration of human DNA/RNA co-extracted from the palmar surface of the hands and fingers, *Forensic Science International: Genetics*. 22 (2016) 44–53.

Lapointe M., Rogic A., Bourgoin S., Jolicoeur C., Séguin D., Leading-edge forensic DNA analyses and the necessity of including crime scene investigators, police officers and technicians in a DNA elimination database, *Forensic Science International: Genetics*. 19 (2015) 50-55.

Lee H.C., Ladd C., Scherzinger C.A., Bourke M.T., Forensic applications of DNA typing, *American Journal of Forensic Medicine and Pathology*. 19 (1998) 10–18.

Lee H.C., Ladd C., Preservation and collection of biological evidence, *Croatian Medical Journal*. 42 (2001) 225–228.

Lehmann V.J., Mitchell R.J., Ballantyne K.N., van Oorschot R.A.H., Following the transfer of DNA: How does the presence of background DNA affect the transfer and detection of a target source of DNA?, *Forensic Science International: Genetics*. 19 (2015) 68–75.

Lowe A., Murray C., Whitaker J., Tully G., Gill P., The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces, *Forensic Science International*. 129 (2002) 25–34.

Malsom S., Flanagan N., McAlister C., Dixon L., The prevalence of mixed DNA profiles in fingernail samples taken from couples who co-habit using autosomal and Y-STRs, *Forensic Science International: Genetics*. 3 (2009) 57–62.

Matte M., Williams L., Frappier R., Newman J., Prevalence and persistence of foreign DNA beneath fingernails, *Forensic Science International: Genetics*. 6 (2012) 236–243.

McKenna L., Understanding DNA results within the case context: Importance of the alternative proposition, *Frontiers of Genetics*. 4 (2013) 1–3.

---

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

---

Meakin G., Jamieson A., DNA transfer: Review and implications for casework, *Forensic Science International: Genetics*. 7 (2013) 434–443.

Meakin G., Butcher E.V., Van Oorschot R.A.H., Morgan R.M., The deposition and persistence of indirectly-transferred DNA on regularly-used knives, *Forensic Science International: Genetics Supplement Series*. 5 (2015) e498–e500.

Meakin G., Butcher E.V., van Oorschot R.A.H., Morgan R.M., Trace DNA evidence dynamics: An investigation into the deposition and persistence of directly- and indirectly-transferred DNA on regularly-used knives. *Forensic Science International: Genetics*. 29 (2017) 38-47.

Morrison, G. S. Tutorial on logistic-regression calibration and fusion : Converting a score to alikelihood ratio, *Australian Journal of Forensic Sciences*. 45 (2013) 173–197.

Moretti T.R., Justa R.S., Kehl S.C. et al., Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles, *Forensic Science International: Genetics* 29 (2017) 126–144.

Noël S., Lagace K., Rogic A. et al., DNA transfer during laundering may yield complete genetic profiles, *Forensic Science International: Genetics*. 23 (2016) 240–247.

Oldoni F., Castella V., Hall D., Exploring the relative DNA contribution of first and second object's users on mock touch DNA mixtures, *Forensic Science International: Genetics. Supplement Series*. 5 (2015) 300–301.

Oldoni F., Castella V., Hall D., Shedding light on the relative DNA contribution of two persons handling the same object, *Forensic Science International: Genetics*. 24 (2016) 148-157.

Osagie K. O. High-Risk Forensics. *New York Times*. July 24, 2013.

Palmer R., The Evaluation of Fibre Evidence in the Investigation of Serious Crime, PhD thesis, School of Criminal Justice, University of Lausanne, 2016.

Pesaresi M., Buscemi L., Alessandrini F., Cecati M., Tagliabracci A., Qualitative and quantitative analysis of DNA recovered from fingerprints, *International Congress Series*. 1239 (2003) 947–951.

Petty R.E., Cacioppo J.T., *Attitudes and Persuasion: Classic and Contemporary Approaches*, Westview Press, Boulder, CO, US, 1996.

Pieffer C.M., Wiegand P., Persistence of touch DNA on burglary-related tools, *International Journal of Legal Medicine*. 131:4 (2017) 941-953.

Phipps M., Petricevic S., The tendency of individuals to transfer DNA to handled items, *Forensic Science International*. 168 (2007) 162–168.

Poetsch M., Bajanowski T., Kamphausen T., Influence of an individual's age on the amount and interpretability of DNA left on touched items, *International Journal of Legal Medicine*. 127 (2013) 1093–1096.

---

## Evaluation des résultats ADN considérant des propositions au niveau de l'activité 2019

---

Polley D., Mickiewicz P., Vaughn M., Miller T., Warburton R., Komonski D., Kantautasa C., Reida B., Frappiera R., Newman J., An investigation of DNA recovery from firearms and cartridge cases, *Canadian Society of Forensic Science Journal*. 39 (2006) 217–228.

Port N.J., Bowyer V.L., Graham E.A.M., Batuwangala M.S, Rutty G.N, How long does it take a static speaking individual to contaminate the immediate environment?, *Forensic Science Medicine and Pathology*. 2 (2006) 157–163.

Poy A., Van Oorschot R.A.H., Beware; gloves and equipment used during the examination of exhibits are potential vectors for transfer of DNA-containing material, *International Congress Series*. 1288 (2006) 556–558.

Proff C., Schmitt C., Schneider P.M., Foerster G., Rothschild M.A, Experiments on the DNA contamination risk via latent fingerprint brushes, *International Congress Series*. 1288 (2006) 601–603.

Pun K-M. Interprétation des profils génétiques obtenus à partir de traces de contact. Thèse de doctorat. Ecole des Sciences Criminelles, Université de Lausanne. Lausanne, 2016.

Raymond J.J, Walsh S.J., Van Oorschot R.A.H., Gunn P.R., Evans L., Roux C., Assessing trace DNA evidence from a residential burglary: Abundance, transfer and persistence, *Forensic Science International: Genetics Supplement Series*. 1 (2008) 442–443.

Raymond J.J., Van Oorschot R.A.H., Gunn P.R., Walsh S.J., Roux C., Trace evidence characteristics of DNA: A preliminary investigation of the persistence of DNA at crime scenes, *Forensic Science International: Genetics*. 4 (2009) 26–33.

R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2018). <https://www.R-project.org/> (Dernier accès le 22.08.2019).

RStudio Team, RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. (2016). <http://www.rstudio.com/> (Dernier accès le 22.08.2019).

Rutty G.N., Watson S., Davison J., DNA contamination of mortuary instruments and work surfaces: a significant problem in forensic practice?, *International Journal of Legal Medicine*. 114 (2000) 56–60.

Rutty G.N., An investigation into the transference and survivability of human DNA following simulated manual strangulation with consideration of the problem of third party contamination, *International Journal of Legal Medicine*. 116 (2002) 170–173.

Ryan S., Touch DNA. What is it? Where is it? How much can be found? And, how can it impact my case?. [http://ryanforensicdna.com/yahoo\\_site\\_admin/assets/docs/Touch\\_DNA\\_article.59101956.pdf](http://ryanforensicdna.com/yahoo_site_admin/assets/docs/Touch_DNA_article.59101956.pdf) (accessed December 22, 2015).

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

---

Samie L., Hicks T., Castella V., Taroni F., Stabbing simulations and DNA transfer, *Forensic Science International: Genetics*. 22 (2016) 73–80.

Samie L., Champod C., Glutz V., Garcia M., Castella V., Taroni F., The efficiency of DNA extraction kit and the efficiency of recovery techniques to release DNA using flow cytometry, *Science & Justice* 59. (2019(a)) 405-410.

Samie L., Champod C., Taylor D., Taroni F., The use of Bayesian Networks and simulation methods to identify the variables impacting the LR in stabbing cases evaluated given activity level propositions, *Forensic Science International: Genetics*. (2019(b)). Accepté avec révision mineure.

Samie L., Champod C., Taroni F., Estimating the quantity of transferred DNA in primary and secondary transfers, *Science & Justice*. (2019(c)). Accepté, en production.

Nadarajah S., *Stochastic Processes: Modelling and Simulation, Handbook of Statistics*, Shanbhag D.N. and Rao C.R. 21 (2003) 607-691.

Saravo L., Spitaleri S., Piscitello D., Travali S., DNA typing from steel cable, *International Congress Series*. 1261 (2004) 473–475.

Scherzinger C.A., Ladd C., Bourke M.T., Adamowicz M.S., Johannes P.M., Scherzinger R., Beesley T., Lee H.C., A systematic analysis of PCR contamination, *Journal of Forensic Sciences*. 44 (1999) 1042-1045.

Schwark T., Poetsch M., Preusse-Prange A., Kamphausen T., Von Wurmb-Schwark N., Phantoms in the mortuary—DNA transfer during autopsies, *Forensic Science International*. 216 (2012) 121–126.

Sczech E. *Persistence des traces biologiques sur des habits*, Travail de Master, Ecole des Sciences Criminelles, Université de Lausanne. Lausanne, 2011.

Shaw K., Sesardić I., Bristol N., Ames C., Dagnall K., Ellis C., Whittaker F., Daniel B., Comparison of the effects of sterilisation techniques on subsequent DNA profiling, *International Journal of Legal Medicine*. 122 (2008) 29–33.

Stouder S., Reubush K., Hobson D., Smith J., Trace evidence scrapings: A valuable source of DNA?, *Forensic Science Communications*. 3 (2001).

Sullivan H., Johnson K., Rowlands P., Allen D., New developments and challenges in the use of the UK DNA database: Addressing the issue of contaminated consumables, *Forensic Science International*. 146 (2004) S175–S176.

Szkuta B., Ballantyne K.N., van Oorschot R.A.H., Transfer and persistence of DNA on the hands and the influence of activities performed, *Forensic Science International: Genetics* 28 (2017) 10–20.

Szkuta B., Ballantyne K.N., Kokshoorn B., van Oorschot R.A.H., Transfer and persistence of non-self DNA on

---

## Evaluation des résultats ADN considérant des propositions au niveau de l'activité 2019

---

hands over time: Using empirical data to evaluate DNA evidence given activity level propositions, *Forensic Science International: Genetics* 33 (2018) 84–97.

Tamariz J., Voynarovska K., Prinz M., Caragine T., The application of ultraviolet irradiation to exogenous sources of DNA in plasticware and water for the amplification of low copy number DNA, *Journal of Forensic Sciences*. 51 (2006) 790–794.

Taroni F., Bozza S., Biedermann A., Garbolino P., Aitken C., Data analysis in forensic science; A Bayesian decision perspective, *Statistics in practice*, John Wiley & Sons (2010).

Taroni F., Biedermann A., Vuille J., Morling N., Whose DNA is this? How relevant a question? (a note for forensic scientists), *Forensic Science International: Genetics*. 7 (2013) 467–470.

Taroni F., Biedermann A., Bozza S., Bayesian networks for probabilistic inference and decision analysis in forensic science, *Statistics in practice*, John Wiley & Sons, 2<sup>nd</sup> edition (2014).

Taroni F., Bozza S., Biedermann A., Aitken C., Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio, *Law, Probability and Risk*. 15 (2016) 1-16.

Taroni F., Bozza S., Biedermann A., Aitken C., Rejoinder, *Law, Probability & Risk*. 15 (2016) 31-34.

Taroni F., Bozza S., Hicks T., Garbolino P., More on the question ‘When does absence of evidence constitute evidence of absence?’ How Bayesian confirmation theory can logically support the answer, *Forensic Science International*. 301 (2019) e59–e63.

Taylor D., Abarno D., Rowe E., Rask-Nielsen L., Observations of DNA transfer within an operational Forensic Biology Laboratory, *Forensic Science International: Genetics*. 23 (2016) 33–49.

Taylor D., Hicks T., Champod C., Using Sensitivity Analyses in Bayesian Networks to Highlight the Impact of Data Paucity and Direct Future Analyses: A Contribution to the Debate on Measuring and Reporting the Precision of Likelihood Ratios. *Science & Justice*, 56 (2016) 402-410.

Taylor D., Biedermann A., Samie L., Pun K.-M., Hicks T., Champod C., Helping to distinguish primary from secondary transfer events for trace DNA, *Forensic Science International: Genetics* 28 (2017) 155-177.

Taylor D., Kokshoorn B., Biedermann A., Evaluation of forensic genetics findings given activity level propositions: A review, *Forensic Science International: Genetics* 36 (2018) 34–49.

Taylor D., Samie L., Champod C., Using Bayesian networks to track DNA movement through complex transfer scenarios, *Forensic Science International: Genetics* 42 (2019) 69–80.

Topham G, Man wrongly convicted of sexual assault is freed after 17 years in jail. Appeal court overturns Victor Nealon's conviction after DNA evidence points to another perpetrator, *The Guardian*. December 13, 2013.

## Evaluation des résultats ADN considérant des propositions au niveau de l'activité 2019

---

Van Den Berge M., Ozcanhan G., Zijlstra S., Lindenberg A., Sijen T., Prevalence of human cell material: DNA and RNA profiling of public and private objects and after activity scenarios, *Forensic Science International: Genetics*. 21 (2016) 81–89.

Van Oorschot R.A.H., M.K. Jones M.K., DNA fingerprints from fingerprints., *Nature*. 387 (1997) 767.

Van Oorschot R.A.H., Phelan D., Furlong S., Scarfo G., Holding N., Cummins M., Are you collecting all the available DNA from touched objects?, *International Congress Series*. 1239 (2003) 803–807.

Van Oorschot R.A.H., Glavich G., Mitchell R.J., Persistence of DNA deposited by the original user on objects after subsequent use by a second person., *Forensic Science International: Genetics*. 8 (2014) 219–225.

Van Oorschot R.A.H., Szkuta B., Meakin G.E., Kokshoorn B., Goray M., DNA transfer in forensic science: a review, *Forensic Science International: Genetics*. 38 (2019) 140-166.

Verdon T.J., Mitchell R.J., Van Oorschot R.A., Evaluation of tapelifting as a collection method for touch DNA, *Forensic Science International: Genetics*. 8 (2014) 179–186.

Weller, R. v [2010] EWCA Crim 1085 (04 March 2010).

Wickenheiser R.A., Trace DNA: A review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact, *Journal of Forensic Sciences*. 47 (2002) 442–450.

Willis S. et al., ENFSI guideline for evaluative reporting in forensic science, European Network of Forensic Science Institutes, Dublin (2015). [http://enfsi.eu/wp-content/uploads/2016/09/m1\\_guideline.pdf](http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf) (Dernier accès le 25.10.2018).

## Annexes

---

### Annexe I: Articles publiés

I.1 Taylor et al., Helping to distinguish primary from secondary transfer events for trace DNA, *Forensic Science International: Genetics*. 28 (2017) 155–177.

Duncan Taylor<sup>1,2</sup>, Alex Biedermann<sup>3</sup>, Lydie Samie<sup>3</sup>, Ka-Man Pun<sup>3</sup>, Tacha Hicks<sup>4</sup>, Christophe Champod<sup>3</sup>

1. Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia

2. School of Biological Sciences, Flinders University, GPO Box 2100 Adelaide SA, Australia 5001

3. Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice, University of Lausanne, Lausanne-Dorigny, Switzerland

4. Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice and Fondation pour la formation continue UNIL-EPFL, University of Lausanne, Lausanne-Dorigny, Switzerland

#### **INTRODUCTION:**

DNA profiling evidence is commonplace in the courtroom for a variety of criminal offences. Powerful tools [1-7] exist that can help evaluate questions of whether or not the DNA of a person of interest (POI), or a combination of DNA from multiple POIs, is present in a particular DNA extract. Increasing so, with the advent of ever more sensitive DNA profiling systems and laboratory hardware, the value of such queries about DNA is being questioned. In fact, what is being questioned is not the reliability of the DNA profiling results, nor the evaluation of the DNA profile, but rather the significance of those findings in relation to how they support different activity-level propositions suggested by, for example, the competing assertions of the prosecution and defence. Such propositions are described as being activity level within the concept of the hierarchy of propositions [8]. One common question is whether the DNA that has been detected from an examined item was deposited by being directly handled (known as a primary transfer event) or whether there was an intermediary object that acted as a vector to transfer the DNA from the POI ultimately to the item in question (known as a secondary transfer event). Naturally, scenarios that explain the presence of DNA on an object need not be limited to primary and secondary transfers, and there has been literature that documents instances of tertiary and even quaternary transfer



events [9]. The mode of transfer by which DNA came to be on an item has profound implications on the way the DNA profiling results are considered in light of questioned activities. In order to assess the biological results given the alleged activities one needs to understand the factors that affect DNA transfer and persistence on differing target surfaces. Also required is knowledge of case specific details such as the amount of genetic material the POI is likely to shed, and the alleged activities (e.g., timing, type of contact with the objects).

To help address questions of transfer and persistence there have been numerous publications that consider transfer rates under varying conditions [9-14]. Often these studies replicate conditions of a specific case, or are very specific to the hardware and wetware used. This means that it can be difficult to apply their findings in a probabilistic manner to situations that are somewhat removed from those used in the study. We believe this may stem from the fact that researchers design studies and describe results without having a logical framework of interpretation in which to place them. Additionally, it is difficult to separate the factors of transfer to an object, persistence on the object and recovery in the laboratory and so they are often considered jointly, which again complicates the ability to apply the results more broadly to other cases.

Another complicating factor in the evaluation of transfer events is that there is a high degree of variability that exists in seemingly multiple factors, not the least of which is whether the individual is prone to shedding or retaining their DNA [15]. All of these difficulties were presented in a review of DNA transfer by Meakin *et al.* [11], who concluded that, by just the properties of the DNA profile obtained, no definitive conclusions could be made by an analyst as to whether it had resulted from primary or secondary transfer, the order in which individuals had touched an item or whether the DNA detected had been deposited by regular use or a one-off contact (amongst other similar findings). This finding has then been interpreted by many analysts as saying that there is no evaluative information within the DNA trace, implying that, given the findings at hand, any explanation is possible. In response to this interpretation of the Meakin *et al.* findings, Casey *et al.* [16] called for the evaluation of DNA profiling results, in light of questions of activity, to be strived for regardless of the difficulties involved (see also a response to this response from the original authors in [17]). This is a sentiment to which we agree and it has been influential in our decision to write this paper.

With increasing regularity Bayesian networks (*BN*) are being used to bring together various kind of datasets using Bayesian probability theory in order to help address questions at the source [18, 19] or activity level [20-23]. The graphical ability for *BN* to represent complex underlying calculations makes them ideal for addressing the issue of this paper i.e., the combination of biological results with the framework of circumstances that surrounds an activity in order to help address questions regarding the mechanism by which some DNA came to be on an item. We attempt to construct a *BN* in a manner that makes it adaptable to a wide range of situations. We do so by breaking apart considerations of transfer, persistence and recovery of DNA. In this research paper we have adopted a model that details many variables that are at play. We acknowledge that different models are possible, including ones with a less detailed account of the variables.

In the data collection section we review the findings of relevant literature to determine, firstly, which factors have been found to have an effect on trace DNA transfer and persistence and secondly what those effects are. In the Bayesian Network section we propose a structure for a *BN* that captures understanding and domain knowledge derived from published data, and then inform conditional probability tables with data wherever possible. In the application to different case examples section, we demonstrate how the *BN* developed and parameterised in the previous sections can be applied to several examples that the authors have encountered during testimony. Finally, in the Application to Controlled Case section, we study the performance of our *BN* on real results generated from known deposition events.

### **Preliminary considerations**

A great advantage of Bayesian networks is that they help advance thinking. A crucial step will be the definition of the variables: in forensic science this typically involves formulating the propositions and the results to be assessed. This may seem obvious, but it is not [24]. This is because results need to be communicated, therefore summarized to some extent. But, on the other hand, they have to be considered in sufficient detail, so that differences may be observed depending on the proposition. We know, for example, that one can observe transfer of DNA in different scenarios and that the quantity of DNA (or the relative quantity of each contributors in presence of mixtures) varies. This is also true of other types of trace material, such as glass or fibres that have been used for many decades to help discriminate between activity level propositions [25]. One key element that has been highlighted for such traces is that it is not transfer per se that is of interest to forensic scientists, but how different the results are given the alleged activities. If we take an example in glass, it is not the transfer of

---

glass that is key, but the recovery of *only one*<sup>1</sup> large group of fragments on a given garment (i.e., what is called the extrinsic characteristics). Indeed, the probability of recovering any glass in general may be very similar given the two activities, but the recovery of a large group will not. With the increase of sensitivity, absence and presence of DNA is not sufficient to discriminate between primary and secondary transfer. DNA results need therefore to be described more precisely in terms of quantity and/or quality, to show which extrinsic characteristics help discriminate alleged activities. This approach has been used, for example, in the Weller case [26], where scientists have considered the probability of their results (in terms of quantity of DNA and positioning) given the propositions describing competing activities.

Before describing the studies that are available to inform our knowledge on transfer, we would like to mention two last important points: first, answers to questions regarding activity level propositions are probabilistic in nature and that no experiments will tell us whether transfer was primary or secondary. Following this, whether transfer was primary or secondary is, ultimately, for the Court to consider given all the available information. Notwithstanding, because of their specialised knowledge, forensic DNA scientists can help the Court by evaluating their results (e.g., a recovered quantity of DNA leading to a profile of this quality) *given* the alleged activities (or given primary versus secondary transfer).

The above idea is often misunderstood [58] and it is worth expanding upon it. Any opinions provided on the more probable mechanism of transfer, given the DNA amounts, is a comment on the posterior probability, and as such, requires one or more prior probabilities. This is then must be the remit of the Court, who has access to non-scientific information that will be used to develop their prior belief. The BNs developed in this work always possess a parental node that has states used to signify the positions of prosecution and defence. These states require prior probabilities and for these we use a uniform distribution (i.e. all states are *a priori* equally probably). This of course will not be the position of the Court of the jury, but allows them to use the numerical value of the *LR* obtained from the BN as a 'belief updater'.

---

<sup>1</sup> By indicating 'only one', we also consider the absence of other glass that can be as important as the presence of material.

---

There are other parental nodes that require prior information (such as the DNA on hands or Extraction Efficiencies), however these are the remit of the analyst as they have a scientific basis and could not be expected to be informed by a lay person during a trial.

### **DATA COLLECTION:**

We present here a series of findings from investigations regarding trace DNA transfer, persistence and collection. For each publication we attempt to provide the raw data that we use to inform the *BN* developed in the next section. Not all data will be able to be used as many of the findings combine aspects of transfer, persistence and recovery in their ultimate results. Factors such as instrumentation used in laboratory for processing DNA also influenced experimental data. We will attempt to tease apart such interwoven aspects wherever possible, but concede that for experimentation that targets trace DNA this can be difficult to achieve.

We make a note here that the order in which we introduce these topics may not at first seem intuitive as it does not follow the order of the elements of the *BN*, nor the order of laboratory processes. We present the topics in the order in which they are required to model data, i.e. initially extraction/sampling efficiency are presented as they are required in the model for DNA on hands, all of which are then required for modelling transfer of DNA from hands to objects, and so on.

#### Presence of DNA already on object:

Lehmann *et al.* [27] found that, generally<sup>2</sup>, the presence of trace DNA on an object did not affect the deposition of further trace DNA and so there was no need to account for that in our *BN*. It was found, however, that the presence of other body fluids could affect the deposition, recovery and detection of trace DNA. However, we restrict our attention to situations where trace DNA only is assumed.

#### Extraction efficiency:

Extraction efficiency is defined as the amount of DNA on a sampling device that is released into a DNA extract and made available for PCR. This is known as 'absolute extraction

---

<sup>2</sup> If there is DNA from numerous persons, then this will affect our ability to detect the profile of interest, as noted by Lehmann *et al.* 'The presence of several different sources of background DNA created mixed profiles and had major negative influences on the detection of the target source of DNA'.

---

efficiency'. We use the information presented by Butts [28] who tested two extraction procedures (a salting out method and the Qiagen EZ1 Advanced XL extraction robot) on DNA, epithelial cells and blood. They found that the DNA loss was “*independent of extraction method or source of DNA*” and so we show the distribution of combined extraction efficiencies (from the summary graph given in [28]) from their work below in Figure 1. We have modelled the observed extraction efficiencies using a  $Beta(5,17)$  distribution. As seen in Figure 1, this fits the observed data reasonably well.

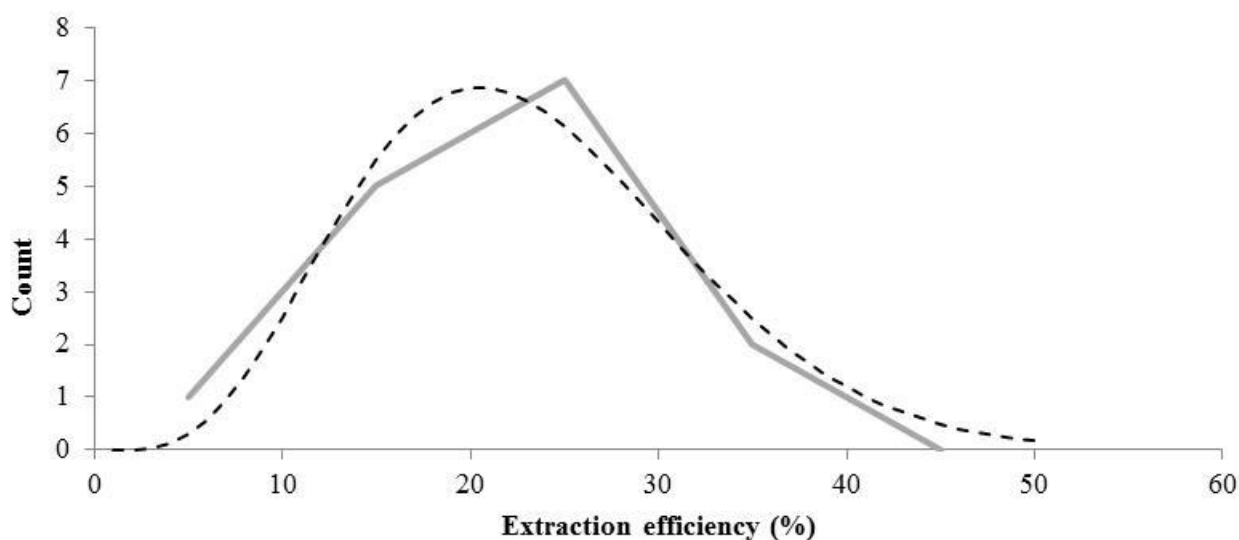


Figure 1: Distribution of extraction efficiencies determined by Butts [28] (grey) and fitted Beta distribution (dashed black line)

Differing extraction methodologies (not examined in [28]) may have varying efficiencies and we recognise that the findings we provide may need to be recreated for other extraction methodologies if the analyst wished the  $BN$  to reflect the properties of their laboratory processes.

Further improvements to the extraction efficiency node would include:

- Trialling extraction efficiencies at low DNA levels. The study of Butts [28] trialled DNA amounts from 24ng to 4800ng
- Trialling of different extraction techniques not tried in the Butts [28] study. Note that some information to this effect can be found in [29] and [30]

DNA sampling/collection efficiency:

We investigate two broad categories of sampling device, tapelifts and swabs. We recognise that both categories could be refined into a number of sub-categories that take into account the type of swab or the type of tape used. It is possible that the variation in sample to sample efficiency (or analyst to analyst differences) may overwhelm the difference in distributions of sampling efficiency from such fine-scale considerations. However, anyone using the BN given in Figure 8 could carry out sampling efficiency validation work to produce findings that are specifically suited to their laboratories process and performance. For the sampling efficiency we used the results of Verdon *et al.* [31]. We define the sampling efficiency as the amount of DNA present on an item that is recovered by the sampling device as detailed further below. Note that the sampling device then goes on to a DNA extraction (typically) and there is a secondary process we consider, the extraction efficiency, which we defined in the previous section. Within the Verdon study tapelifts (using Scotch® Magic™ and Scenesafe FAST™) and swabs (FABswab, Puratin, USA) were used to collect DNA from swatches of cotton, flannelette, Poly/cotton blend and polyester strapping that had contact DNA transferred through vigorous rubbing. The swatches were sampled (either using swabs or tapelifts) and some DNA extracted. The swatch (post collection) was then extracted directly (i.e. not tape-lifted or swabbed further, but instead placed directly into an extraction reaction) and again DNA extracted. The sampling efficiency could then be calculated as the amount of DNA obtained from the DNA extraction of the device to the total DNA extracted from the device plus the swatch. By representing the results as a ratio the effects of the initial amount of DNA deposited and the extraction efficiency are removed from consideration.

Verdon *et al.* [31] trialled swabs and tapelifts on both smooth (polyester strapping) and standard woven material (cotton, flannelette and Poly/cotton blend). In our BN we assume that swabs have been used on smooth surfaces and tapelifts on rougher, fabric, surfaces and so do not consider the cross-over of collection in our use of the Verdon *et al.* [31] findings. Verdon *et al.* [31] also found a significant difference between the two tapes trialled, and we choose to use the results from the Scotch® Magic™ tape as a more commonly used forensic tape. The findings of the tapelifting of fabrics (combining the findings of cotton, flannelette and Poly/cotton blend) in the Verdon *et al.* [31] study Figure 2, and fitting a Beta distribution by least squares yields an efficiency of  $Beta(1.9,16.6)$ . For swabbing we use the results of Verdon *et al.* [31] Figure 2 for the swabbing of strapping for which we use a  $Beta(25,20)$

distribution. These two efficiencies are shown in Figure 2 and in a similar manner as for the extraction efficiency modelling will be incorporated into the *BN*.

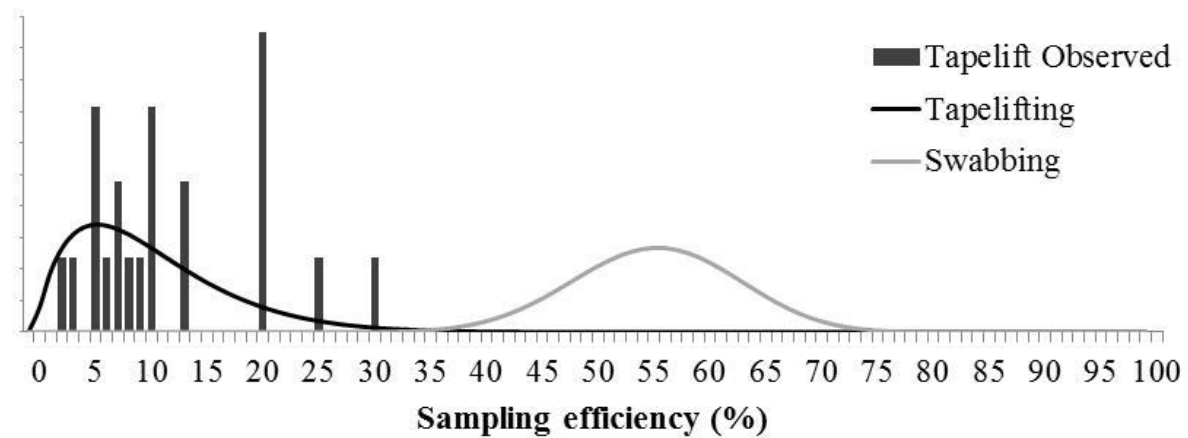


Figure 2: Sampling efficiency of tapelifting and swabbing from results of Verdon *et al.* [31] for tapelifting (black) and swabbing (grey). The histogram shows observed tapelift efficiencies. For swabbing there was only one average value given.

#### Persistence:

There is little data available on the persistence of trace DNA. There are a number of factors that are likely to affect persistence, such as the surface type, the length of time and the conditions the item is exposed to during the time. The best example of a trace DNA persistence study for contact DNA is the work by Raymond *et al.* [32]. In [32] known amounts of cellular (using buffy coat) and free DNA (using positive control DNA 9947A) was deposited on:

- An outdoor window frame
- A vinyl bag kept outdoors
- Glass slides kept in controlled laboratory conditions

The outdoor samples were in partly shaded areas over average temperature and humidity conditions of 24.1°C, 63% humidity (day) and 18°C, 71% humidity (night).

While the collection and extraction methodology will mean that absolute DNA amounts cannot be used, they are expected to remain an approximately constant factor throughout the experiment of Raymond *et al.* [32]. This allows relative amounts of recovery to be used for persistence. We recreate the results of Raymond *et al.* [32] Figure 1. However, we combine

the results of the outdoor bag and outdoor window frame experiments (by averaging) as well as averaging the trends across cellular and neat DNA. The reason behind this is two-fold:

- 1) The data from Raymond *et al.* [32] do not show a noticeable difference between these experiments
- 2) Trace DNA deposited onto an item is likely to consist of both cellular and free DNA [33]

For the same reason we average the cellular and neat DNA findings for the laboratory experiment. Finally, we display the results as a ratio relative to the maximum DNA amount observed (because clearly there must have been at least this amount of DNA available at time zero). All of this is shown in Figure 3.

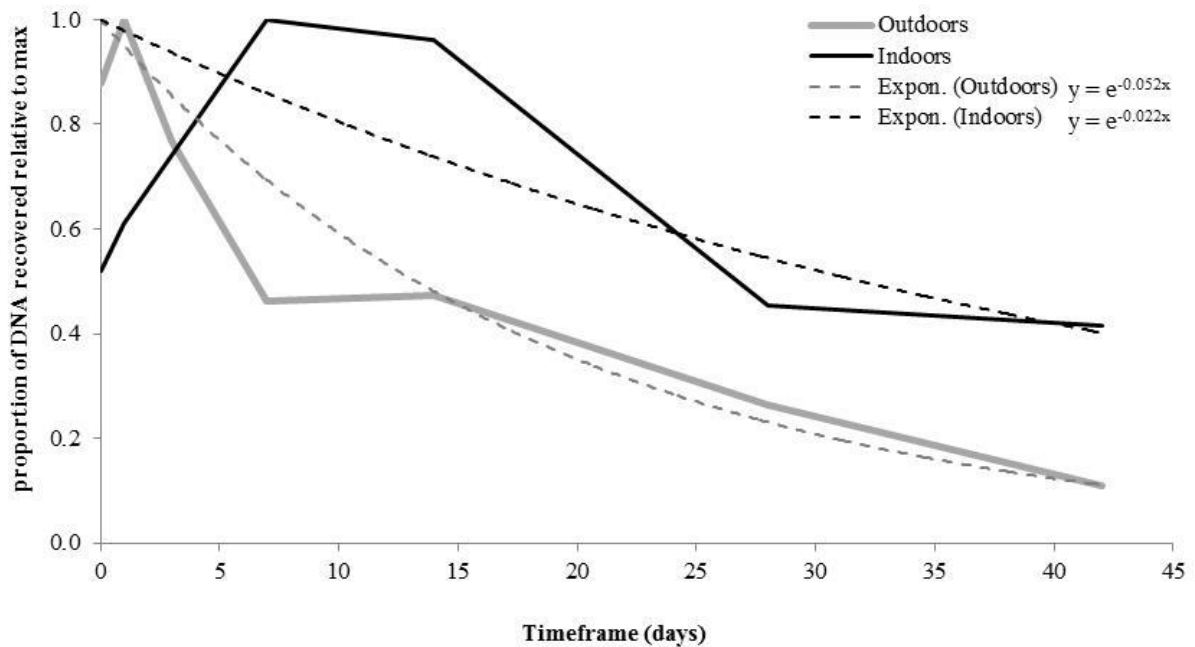


Figure 3: Data from Raymond *et al.* grouped into two categories, outdoors (grey) and indoors (black) with the trends modelled with an exponential curve (dashed lines)

Using the information from Figure 3 we then implement an exponential decay curve in the *BN* for the DNA reduction in samples that are kept in 'poor' or 'favourable' conditions over a number of days, '*t*'.

$$Decay = DNA \times e^{-\alpha t} \quad \text{where } \begin{cases} \alpha = 0.022 & \text{favourable} \\ \alpha = 0.052 & \text{poor} \end{cases}$$



Further improvements to the persistence node would include:

- Taking into account the nature of the surface type the DNA has been placed on when considering persistence
- Further investigation into different environmental conditions (e.g. rain, washed, full sun, etc.) on DNA persistence
- More data to confirm the DNA persistence rates found in Raymond *et al.* [32].
- Data on DNA persistence of DNA on objects after extended handling by other individuals (of which some work has been done in [34]), or from physical movements after initial deposition.

DNA on hands (shedder status):

Initially the *BN* will need to contain information regarding the amount of DNA available for transfer to an object, which is present on an individual's hand. This node encompasses the idea of good and poor shedding of DNA. The idea that individuals may deposit variable amounts of DNA is described in [35] and there has been some debate as to whether the level of shedding that has been attributed to 'good shedders' or 'bad shedders' is a transient property depending on day to day variation, the closeness to last cleaning [36] or that there are simply too many factors to consistently label someone as a 'good' or 'bad' shedder [33]. A recent work [37] suggests that the DNA available for deposition through contact is a mixture of skin cells, free DNA in sweat and sebum and a combination of other bodily fluids present on the individual's hands. Van Den Berge *et al.* [38] show that sebum and sweat contribute to an increase of the quantity of DNA on hands with a lower effect of the sweat compared to sebum. In their work Lacerenza *et al.* [37] swabbed 120 individuals' hands and submitted those swabs for DNA extraction and profiling and RNA extraction for use in body fluid identification. Investigation by Lacerenza *et al.* [37] of a number of factors found that the only significant factor was gender, where males had typically more DNA on their hands than females. The authors attributed this to a difference in general levels of hygiene between the genders. A work by Bontadelli [39] swabbed the hands of 50 individuals and found no difference between males and females.

By analysis of all these findings it seems likely that in reality there are not two distinct groups of individuals, some of which are prone to shedding and others that are not, but rather a distribution of shedding propensity, on which people will exist at different points. Certainly

there are individuals who consistently shed (or perhaps better put transfer) more DNA than others. Like many acts of grouping data in a binary fashion, the designation of 'good' and 'bad' shedder groups has, over time, lead to the belief in two distinct groups of people, when in reality it is simply that the binary classification is an oversimplification of an underlying continuous distribution. The DNA on hands node represents our uncertainty in the amount of DNA on individual's hands, and within this uncertainty exists the propensity of that individual to shed DNA. The node represents a distribution of DNA amounts for a 'random' individual meaning that if the case circumstances indicates some reason that the POI did not behave in a manner similar to a random individual (e.g. had a skin condition, which made them more prone to shedding DNA, such as the well-known case of *R v David Butler in Liverpool 2012*) then some modification of the distribution would need to be made for them. This could be accounted for in the BN by the use of a parent node to the DNA on hands node (e.g. a 'skin condition' node that would specify one of two possible distributions in the 'DNA on hands' node when instantiated).

The results of Bontadelli [39] and Lacerenza *et al.* [37] have both a sampling efficiency effect and a DNA extraction effect present in the data and the actual amount of DNA available on hands is likely to be higher than the quantification results they obtained. To account for these effects we carry out the following process using the Bontadelli [39] data using the software R [40] as follows:

- a) Randomly draw a DNA amount from a normal distribution that describes the distribution of log(ng of DNA) found by Bontadelli [39]
- b) Randomly draw an extraction efficiency from the Beta distribution described in the 'Extraction Efficiency' section and use this to adjust the DNA amount in a) to a DNA amount that was present on the swab head
- c) Randomly draw a sampling efficiency from the Beta distribution described for swabbing in the 'Sampling Efficiency' section and use this to adjust the DNA amount on the swab head in b) to a DNA amount that was present on the hand of the individual

Carrying out such a simulation for 100, 000 iterations produces the distribution shown in Figure 4, labelled 'Bontadelli (E)', which we model as normal distributions for use in the BN. Also shown in Figure 4 are the observed (O) distributions of DNA from the Bontadelli [39] and Lacerenza *et al.* [37] studies. All show a similar distribution.

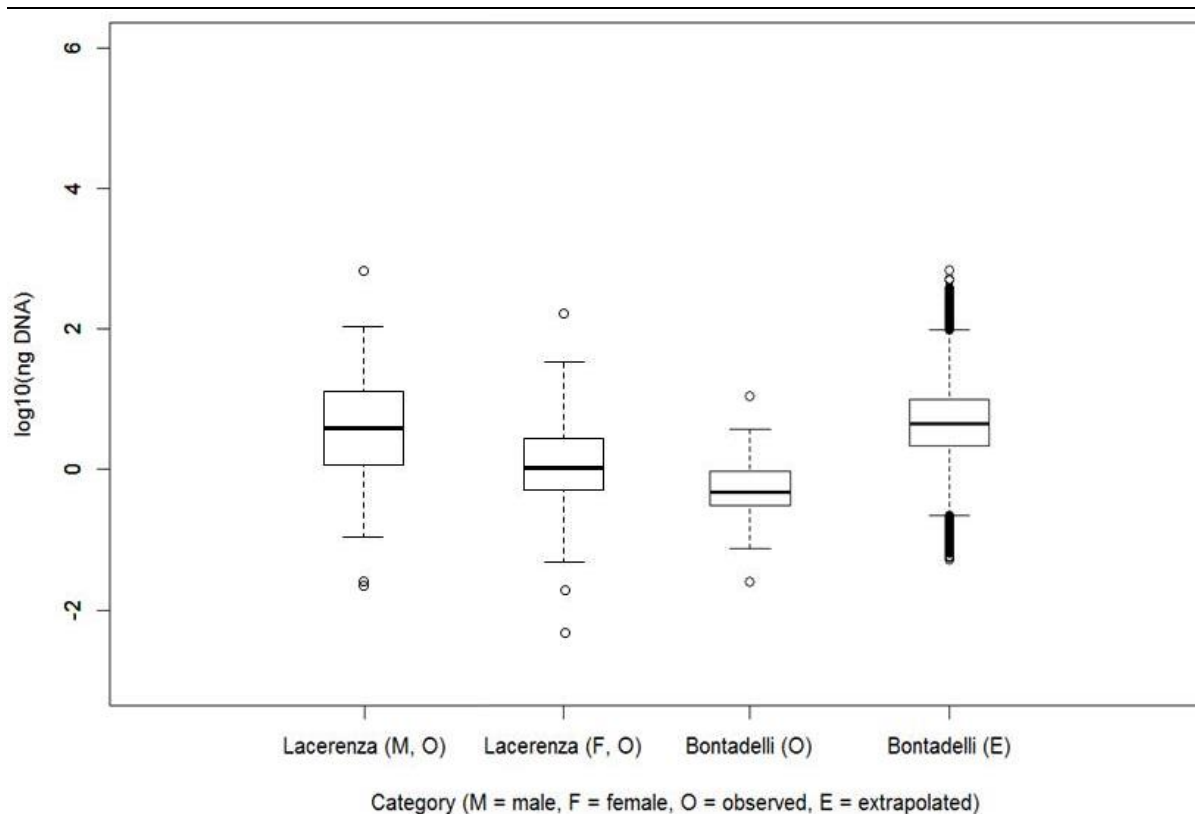


Figure 4: Observed (O) DNA amounts on hands from Bontadelli [39] and Lacerenza et al. [37] studies and adjusted values obtained by simulation (O) for DNA on hands, based on Bontadelli [39] data.

The log(ng of DNA) for DNA on hands from the 'Bontadelli (E)' distribution is modelled by  $N(0.66, 0.49)$ .

Further work in this area could include:

- Shedder consistency studies, i.e. whether an individual sheds DNA consistently in the upper or lower quantiles of the population shedding distribution. Studies could extend to a standard method for determining the approximate shedding propensity of an individual for use in primary vs secondary transfer considerations. Some work in this area has been done in [41].

#### Transfer from hand to object (Primary transfer):

There is much literature that presents findings of transfer from hand to object as the results of obtaining full, partial or no DNA profile [10, 42, 43]. While this information is indeed useful, for the current study (and in particular the primary transfer events depicted by the 'DNA

*transferred to object*' nodes) what is required is absolute DNA amounts. For this, we use the data obtained from Daly *et al.* [44]. In their work they asked 300 random volunteers to grasp an 8mL glass vial, a 7×7cm cotton cloth or an 8.5×1.7×3cm piece of wood for 60 seconds. They then tapelifted with Minitape (WA Products Ltd., UK) and extracted using Qiagen® QIAamp DNA mini kit.

The proportion of hand surface area that contacted the items is not known and so we assume this proportion to be one. We make this assumption so that the transfer data can be directly compared to the amount of DNA on direct hand swabs (which swab 100% of the hand). We also combine the results from the wood and cotton samples and combine them under the surface type category 'rough' and then use the glass results in the surface type category of 'smooth'. We fit gamma distributions to the observed data from Daly *et al.* [44] using least squares. These were  $\Gamma(0.64, 3.87)$  and  $\Gamma(0.33, 1.75)$  for rough and smooth surfaces respectively (graphs not shown). We then adjust the gamma distributions of DNA amounts observed to model the amounts that were present on the item taking into account sampling and extraction efficiency, in the same manner as we did in the 'DNA on hands' section of this paper, to obtain a distribution of DNA amounts transferred by the 300 volunteers to either smooth or rough surfaces as seen below in Figure 5.

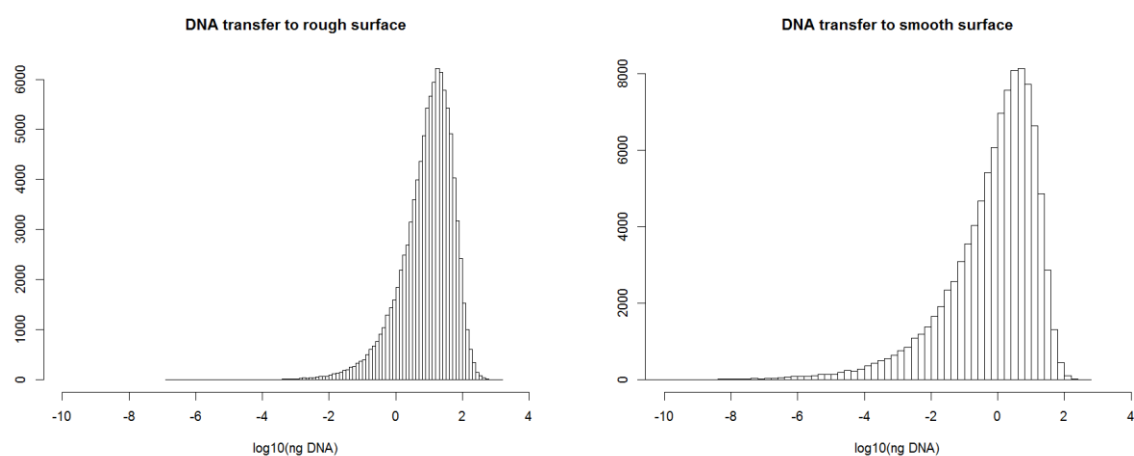


Figure 5: Simulated values for DNA amounts transferred to rough (left) or smooth (right) surfaces, based on the results of Daly *et al.* [44].

Having produced the two distributions, we are interested in the distribution that describes the decrease from the total amount of DNA present on an individual's hand (seen in the 'Bontadelli (E)' category in Figure 4) to the amount of DNA transferred (Figure 5).

In the Daly *et al.* [44] study there is no account of the type of contact that has been made with the object e.g. a glancing touch, pressure for a short period of time, friction, etc. However studies such as that conducted by Goray *et al.* [45] show that the type of contact is an important factor. In their study Goray *et al.* [45] trialled three different contact types:

Passive – described as a placing two substrates together for 60s

Pressure – described as the same as passive, but applying 1kg of weight

Friction – described as the same as pressure but moving the weight around for the 60s

We believe that the category of friction from the Goray *et al.* [45] study is best aligned with the experimental setup of Daly *et al.* [44]. The results given in the Goray *et al.* [45] study are given as mean percentage of DNA transfer. We take the results from Table 3 of that study and scale all findings between smooth (plastic) and rough (cotton) so that the friction category has a value of 1. We show the results of this data manipulation in Figure 6.

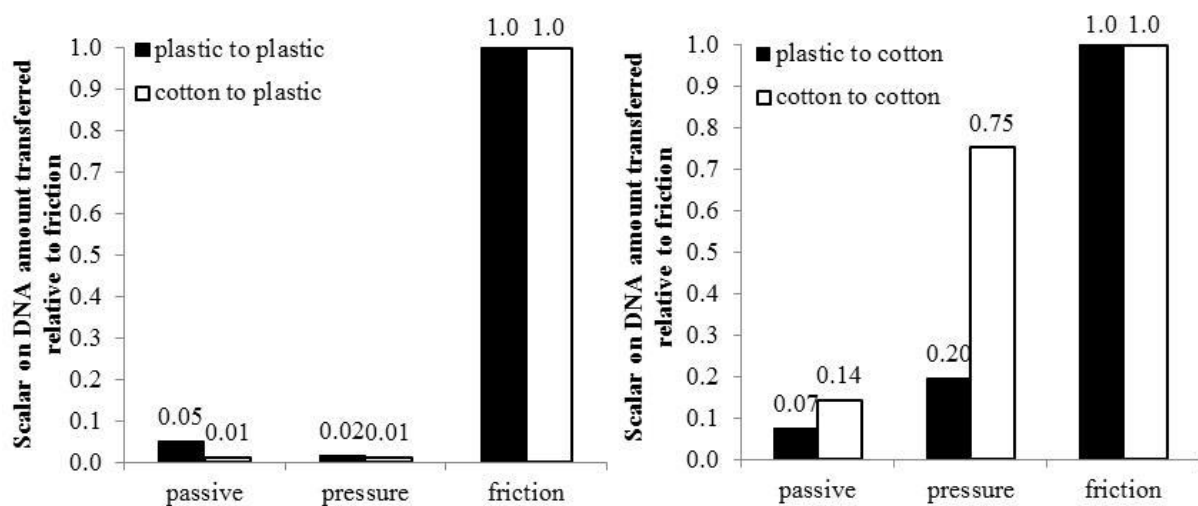


Figure 6: Effect of pressure type on amount of DNA transferred. We break the data into two groups; DNA transferred onto smooth objects (left) and DNA transferred onto rough object (right).

We model the proportion of DNA deposited by a simplified comparison of distributions.

Using software R we go through the following steps:

- a) Generate an array of 100,000 variables drawn from the extrapolated DNA amount on the presence on hands (seen in the 'Bontadelli (E)' category in Figure 4).
- b) Generate two arrays of 100,000 variable drawn at random from the distribution of values seen for deposition onto smooth and rough surfaces as shown in Figure 5.
- c) Order the arrays generated in a) and b) and generate two arrays of the proportion of DNA transferred from hands to object by dividing the values in the DNA of rough or smooth object array by the corresponding entries in the DNA on hands array. The result is two arrays of values between 0 and 1 which represent the proportion of DNA transferred from hands to smooth and rough objects<sup>3</sup>.
- d) Each of the values seen in Figure 6, has a level of data variability in the Goray *et al.* [45] study given as a standard deviation. We use these standard deviations (scaled down to align with the values seen in Figure 6) and draw values from the distributions of the reduction factor for passive, pressure and smooth from both rough and smooth surfaces. These values are then multiplied by the reduction values from c) to produce distributions for the reduction in DNA from hand to rough or smooth object for either a passive, pressured or frictional contact type.
- e) The resulting distributions for transfer type are seen in Figure 7. Beta distributions were fit using MLE to these distributions to represent the proportion of DNA transferred. For the distribution to cloth with a pressure or friction contact we fit a mixed beta distribution, which is given in Table 2.

With these distributions available, we have all the information required for the '*DNA transferred to object*' node of the *BN*.

---

<sup>3</sup> Due to the stochastic nature of the data and simulation there were a number of values within the 'proportion of DNA transferred' arrays that had values greater than 1. This is equivalent to greater than 100% of DNA from hands being transferred to an item and hence is nonsensical. In these instances the data was truncated at 1 to obtain the sensible values seen in Figure 7.

---

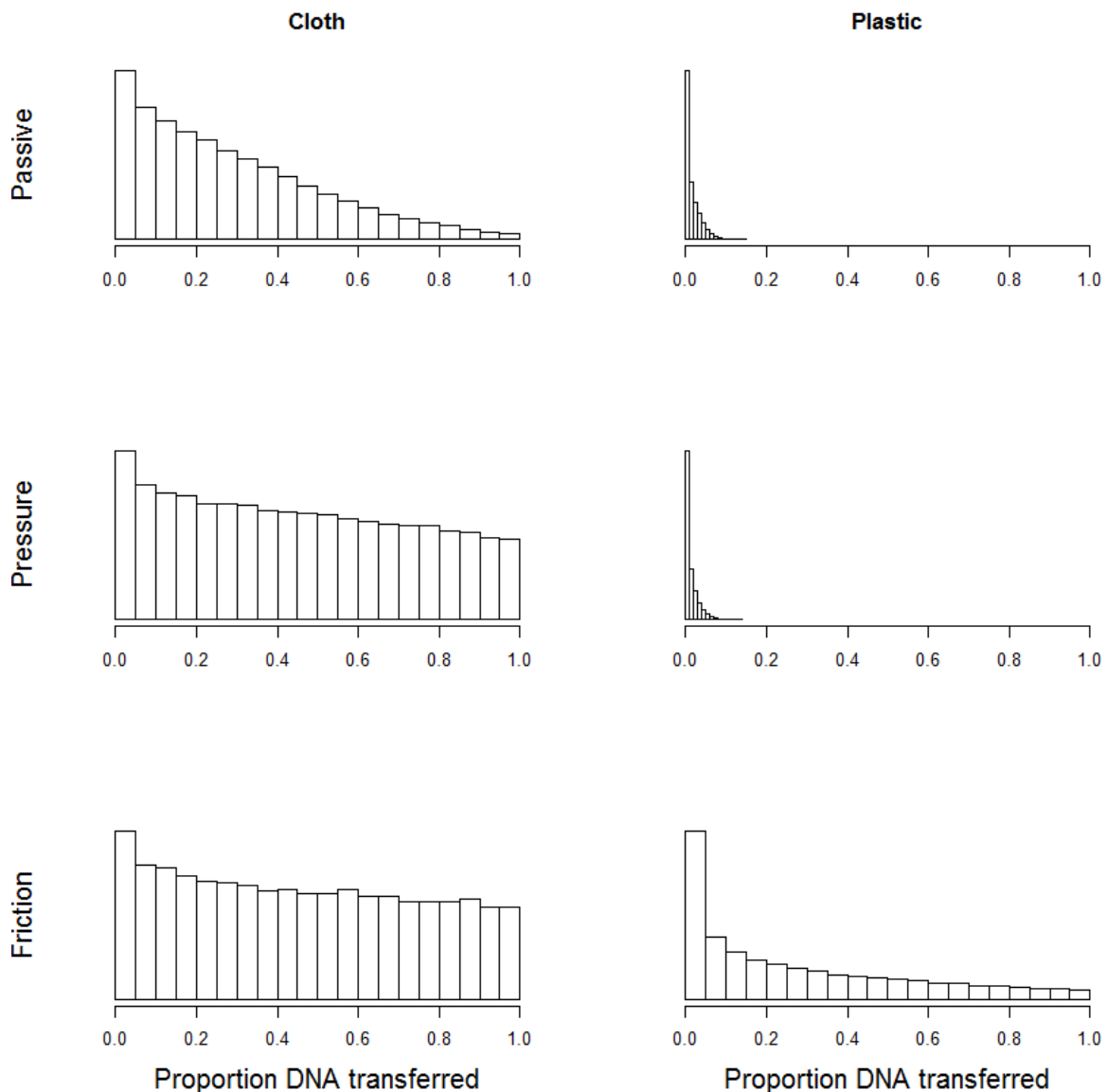


Figure 7: Proportion of DNA transferred from hands to smooth (plastic) or rough (cotton) objects when contact is passive, pressure or friction.

The age of the donor may also have an influence. The study of Poetsch [46] shows that there could be a correlation between the quantity of transfer DNA, the quality of the profiles coming from fingerprints and the age of the donor. On 209 child and adults, a full DNA profiles is obtained in 75% of cases with children under 11 (47 children), 9% with teenager between 12 and 20 (32 teenagers), 25% with adults between 21 and 60 (81 adults) and 8% with persons older than 60 (49 seniors). The factor is not taken into account in the BN if the relevant population is an individual between 21 and 60.

Further work that could be investigated in this area includes:

- A study of how the amount of time an object is held affects the amount of DNA transferred.
- To date we have found such information in [47] for hand to hand transfers mostly by length of handshake. Van Oorschot *et al.* [48] studied this question using polypropylene tubes held for varying lengths of time (5 s, 30 s, 3 min, 10 min) and found that the length of the contact did not influence the amount of DNA transferred. In contrast Saravo *et al* [49] showed that the quality of the profile is influenced by the length of contact (using steel cable).
- A study of the absolute amount of DNA transferred from hands to objects for different contact types e.g. light touch, pressure, friction (i.e. so that data from multiple studies does not need to be combined and extrapolated as we have done here).

Secondary transfer from object to object:

For this final section we again turn to the work of Daly *et al.* [45]. Again we use the results of Table 3 where dried contact DNA is transferred from object to object with varying primary and secondary substrates and different contact types. Table 3 from the Goray *et al.* [45] study gives percentage transfer, mean and standard deviations for all considered transfer scenarios. We apply these transfer distributions with the obvious restrictions that the transfer percentage is bound by 0 and 100%.

Further work that could be done in this area:

- Consideration of the amount of DNA transferred to object from a habitual use e.g. items in the home. Some work has been done in this area, such as [50].
- Transfer DNA amount for varying length of time of contact between primary and secondary substrates and for different types of activities.

### **BAYESIAN NETWORK:**

From the literature we have found the following factors to be important to:

The amount of DNA available for deposition

- Propensity of an individual to shed DNA [37]. We consider this node as describing the amount of DNA on an individual's hands available for transfer. It therefore



encompasses the idea of shedding ability of the individual, and we could consider aspects such as cleanliness, sweating, skin disease at this point.

- Amount of individual's hand that was in contact with object (no reference, this is based on common sense).

The transfer of contact DNA:

- Surface of object being touched [44, 45].
- Surface of object that DNA is currently on [45].
- Vigour and length of contact [45].

The persistence of DNA on an object:

- Time between deposition and sampling (or further deposition) [32].
- Condition the item is kept in between deposition and sampling (or further deposition) [32].
- The type of surface of the object.

The recovery of DNA:

- Sampling device used [31].

Note that we do not consider specific laboratory aspects such as the profiling system used, the number of PCR cycles or models or settings of laboratory hardware. We have made a deliberate choice to model DNA amount, which precedes these laboratory considerations and simplifies the BN. If a laboratory wished they could add nodes onto the BN that translate DNA amount to peak height.

By combination of these factors we formulate the *BN* shown in Figures 8 to 11. This *BN* is constructed as an object-oriented BN (OOBN) with the same sub-networks that are used at multiple points. This model has the advantage that it can be easily expanded to consider a range of transfer scenarios (something we demonstrate in this paper). Overall in Figure 8 we model from top to down the amount of transferred DNA from contact to recovery. In each column, we distinguish the quantity of DNA from the person of interest (POI), the quantity of DNA from the alternative offender (AO), if any, and the quantity of DNA present as background. The proposition nodes at the centre allow activating or not the various transfer options to be considered as a function of the choice of the prosecution or defence allegations. Typically, under the prosecution view, it will be alleged that the POI invoked a primary

transfer with the item under examination. Under the defence account, the POI may invoke a secondary transfer with the item alone (Hd1) or, that an alternative offender had a primary transfer (Hd2). On each column (POI or AO), the DNA can take the routes denoted as primary transfer or secondary transfer. The background DNA (on the right side) is not conditioned on the chosen propositions as it pertains to the item regardless of their states.

Part of the BN deals with the issue as to whether or not the obtained DNA profiles will match with either POI or AO. Before dealing with the results, we deal with the recovery of the DNA from the item. This structure represents the obtained results (at the bottom) in the form of the quantities of DNA arising from transfers of various types (primary or secondary) from POI, or not.

The use of object-oriented structures is shown by the use of the white blocks called TP (transfer and persistence), R (recovery) and M (matching DNA profiles).

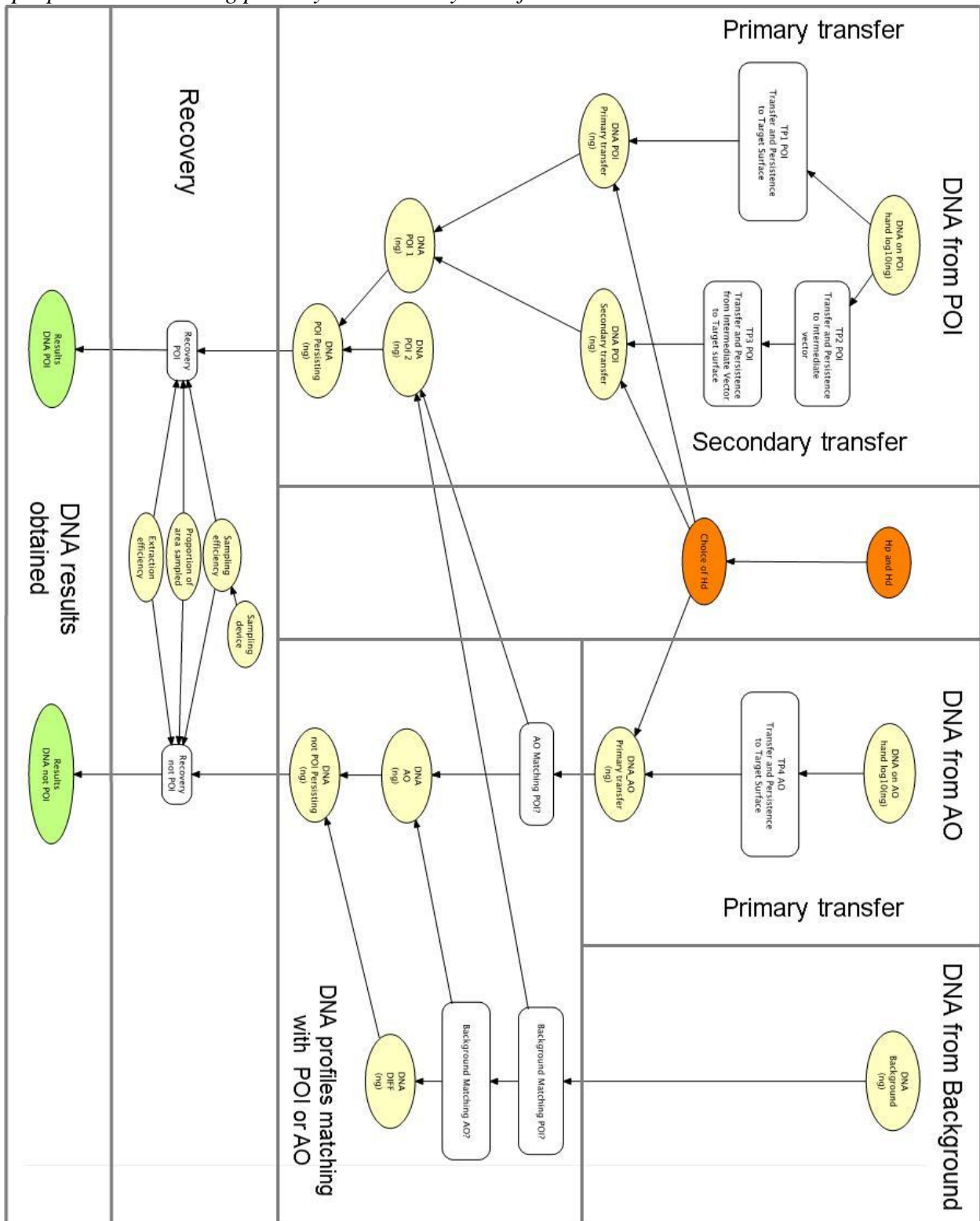
The block TP is shown in Figure 9. It takes a given amount of DNA as input (DNA IN) and progresses it through a transfer and persistency model to a resulting amount of DNA (DNA OUT). Within the block is highlighted the various factors that impact the transfer and persistency.

The block R (for recovery) shown in Figure 10 uses a DNA quantity as input and passes it through the steps of sampling (depending on the technique used) and extraction, leading to the final amount of resulting DNA.

The block M in Figure 11 is assigning the DNA as matching POI (or AO) versus different profiles (called DNA DIFF) as a function of the match probabilities (themselves depending on the quantity of DNA as input).

When considering DNA in term of a primary transfer, it will go through one TP block. When considering DNA in terms of a secondary transfer, two TP blocks are applied. The first deals with the transfer on the intermediate object and the second deals with the transfer from the intermediate to the item under examination. This flexible construction also allows us to model more complex scenarios (e.g. tertiary transfer), if necessary.

Figure 8: Bayesian network used to evaluate the findings with given activity level propositions involving primary vs secondary transfer event.



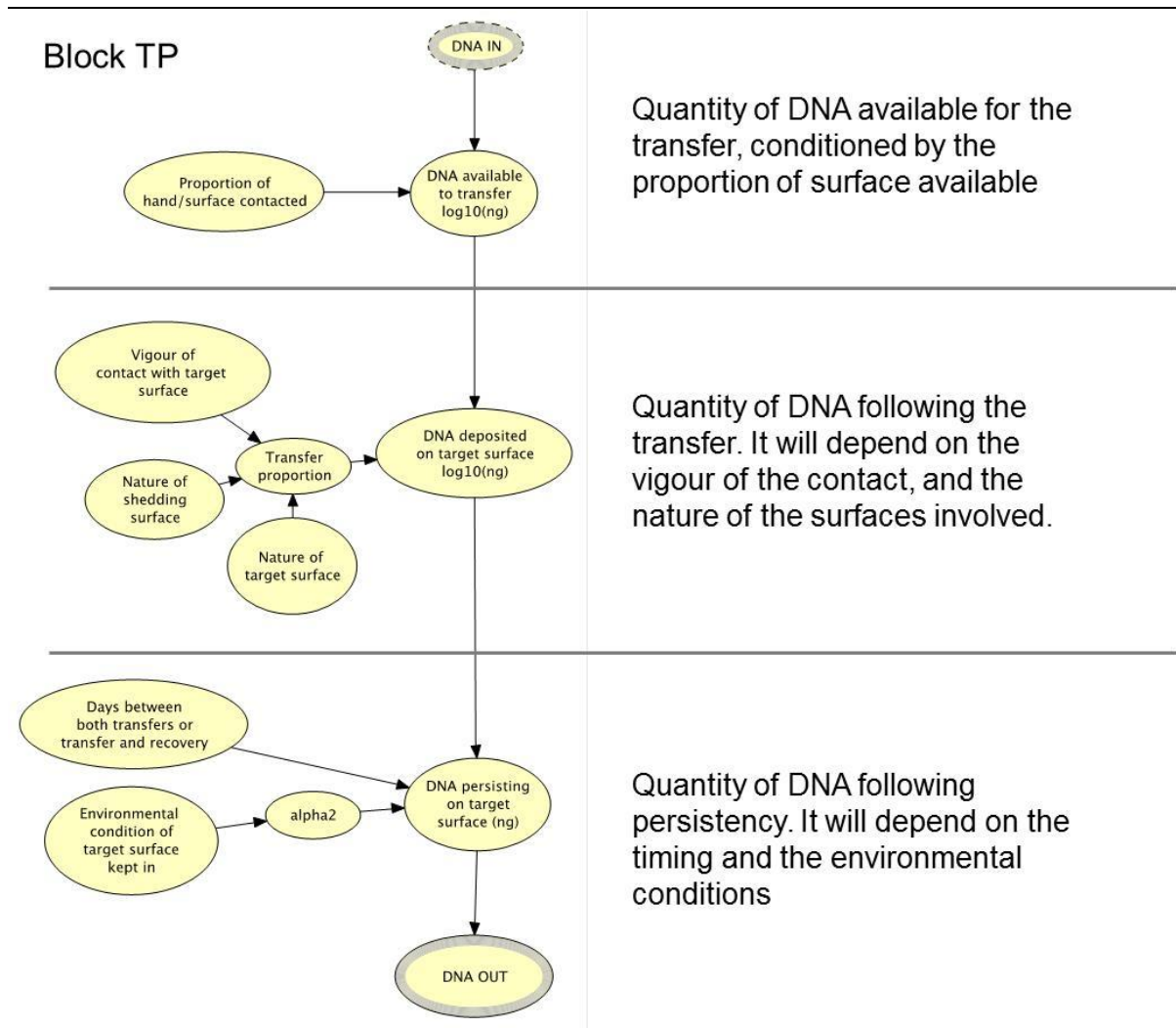


Figure 9: Bayesian network in block TP

Block R

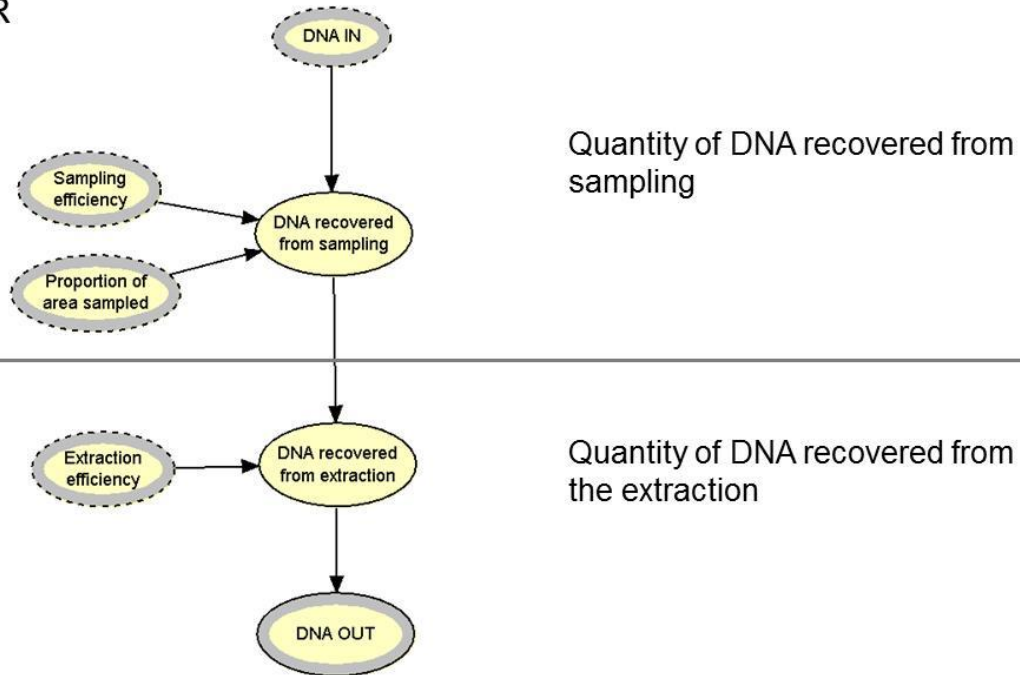


Figure 10: Bayesian network in block R

Block MATCH

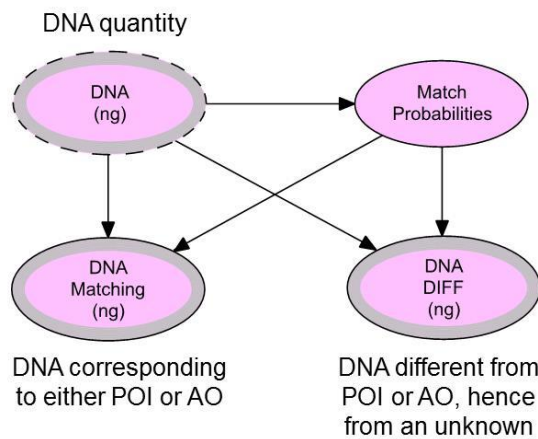


Figure 11: Bayesian network in block M

A combination of conditional probability tables and expressions have been used in the *BN* shown in Figures 8 to 11. Tables 1-4 (in appendix 1) summarise the node definitions, the node states and the manner in which probabilities are provided or generated. Some nodes repeat for each transfer step. The general terminology is given in Tables 1 to Table 4.

## **APPLICATION TO DIFFERENT CASE EXAMPLES:**

Having constructed and examined the BN that can help address secondary vs primary transfer we now apply these data to real casework examples. We provide three examples that have been encountered during testimony, giving a brief description of the alleged offence and the competing propositions. Details have been altered slightly from the real case so that we can demonstrate a range of situations.

### Case example 1:

A bus driver (the POI) is charged with indecent assault where Prosecution alleges that he touched the breasts of the victim over the top of her T-shirt. The victim's T-shirt was seized and sampled the following day and a tapelift from the outer front of the victim's shirt revealed the sole presence (0.15 ng) of the POI's DNA<sup>4</sup>. The POI claims that he put the seatbelt on the victim and his DNA transferred from his hands, to the seatbelt and then secondarily to the victim's T-shirt. The propositions are therefore:

- Hp: The POI touched the breasts of the victim on the outside of her T-shirt
- Hd1: The POI put the seatbelt on the victim and did not touch her breasts

Given the alternative proposition in this scenario, there is no indication of an alternative offender (AO), hence only Hd1 will be considered. The node "choice of Hd" has been set accordingly.

Any state within any node of the BN can be set as being true (with all other states within that node therefore being false). Information provided to a BN in this manner is called 'instantiation' (i.e. the user is instantiating the states of nodes) and once done the laws of probability can be used to propagate the information throughout the BN and update the posterior probabilities for states in non-instantiated nodes. Our instantiations (and rationale) of the nodes is given in Table 5.

---

<sup>4</sup> The findings were initially expressed as a likelihood ratio considering the probability of obtaining the DNA profiling results if the POI was the source of DNA rather than if an unknown male was the source of DNA. The *LR* in this instance was strongly in support of the first proposition over the second and it was conceded by both parties that the POI's DNA was present on the shirt of the victim. In subsequent scenarios when we talk about an individual's DNA being found on an item a similar course of events has taken place to come to that statement. We are not simply assigning identity as the sub-source level *LR* reaches some threshold.

---

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

<b>Node</b>	<b>Instantiated state</b>	<b>Reason/Explanation</b>
Proportion of hand /surface contacted	0.9 – 1	TP1: It is assumed that most of the hand of the POI would have contacted the shirt in the described assault.
	0.9 – 1	TP2: Under secondary transfer the same value will assuming that most of the hand of the POI would have contacted the seatbelt when he it on the victim.
	0.6 – 0.7	TP3: Typically only inner surface of the seatbelt would be contacting the shirt, which is 0.5, however there may be slight opportunity for the outer surface to contact the shirt so we choose 0.6 – 0.7
Nature of target surface	rough	The surface of the victim's woven shirt is rough
Vigour of contact with target surface	friction	TP1: The alleged assault describes a rubbing motion over the top of the victim's shirt. TP2 and TP3: There is typically gripping and pulling motion when putting on a seatbelt, which can be assimilated to friction.
Nature of shedding surface	hand	TP1 and TP2: The hand is the primary source of DNA
	rough	TP3: The surface of the seatbelt is rough
Days between both transfers or transfer and recovery	1	TP1 and TP3: 1 day, the item was examined the following day.
	0	TP2: The seatbelt was immediately in contact with the victim's shirt
Environmental conditions	favourable	TP1 and TP3: Kept indoors in a paper bag TP2: Inside bus (note that as the time node associated with this condition node is set to 0, either condition would give the same result when instantiated)
Proportion of area sampled	0.9 – 1	The entire front of the shirt was sampled
Sampling device	tapelift	A tapelift was used to sample the shirt
Results DNA POI	0.1-0.2	The DNA amount obtained from laboratory analysis
Results DNA not POI	0-0.01	The DNA amount obtained from laboratory analysis

Table 5: Choice for node instantiations as seen in Figure 9

The result was 0.15ng of the DNA for the POI without any other DNA contribution (hence DNA not POI is set to 0). It gives a *LR* of 4. The *LR* for different observed quantities of the POIs DNA in scenario 1 shown in Figure 12.

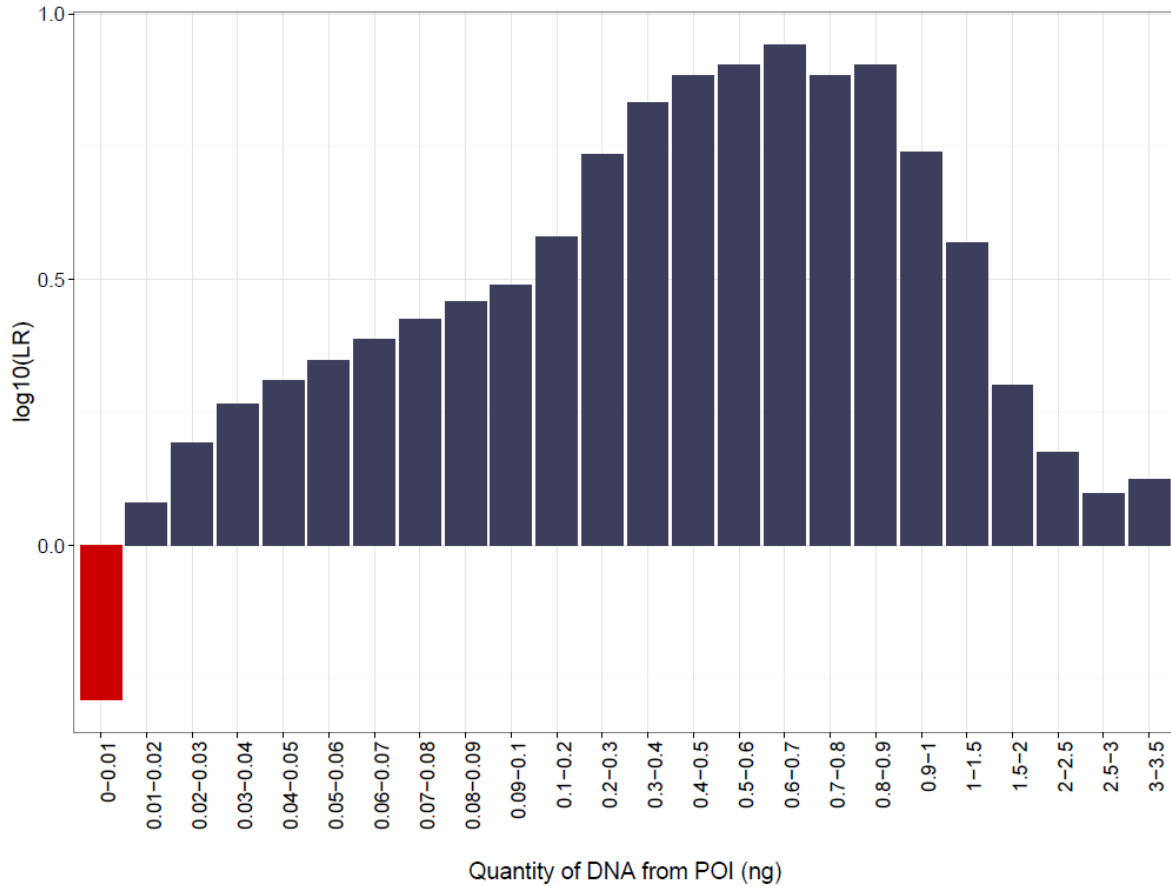


Figure 12: Ratio of probability for primary vs secondary transfer obtained for the instantiated network considering different levels of detected POI DNA.

What is interesting is that as the observed amount of the POI's DNA is above 0.01ng, the support for a primary transfer over a secondary transfer is now above 1 and increases, until it reaches approximately 10 when DNA amounts are 0.6-0.7 ng. As DNA amount increases beyond 0.7 the level of support for a primary transfer decreases compared to a secondary transfer. This initially appeared counterintuitive, but an examination of the 'DNA on hands' node reveals that such a finding means that the individual is a very good DNA shedder. This then obscures the difference between a primary and secondary transfer event. If the amount of DNA on an individual's hands is instantiated as low, then the effect is increasing support for primary over secondary transfer as the amount of POI's DNA detected increases.



The other point to notice is that in this case example, the maximum value the *LR* reaches is approximately 10. This demonstrates – in this case example – the support we can assign to propositions of transfer type with our current knowledge. It should also be noted that increasing our knowledge may not necessarily yield higher levels of support. If the variability of transfer events is high (even after taking into account additional factors) then further knowledge and experimentation will only serve to reinforce that fact.

Also, there are other factors that could be taken into account with further modelling and information. For example, it may be that the suspect is the regular driver of the bus and it could therefore be expected that a level of his DNA is present on the seatbelt prior to the time of the alleged crime. This too could be modelled through a node that considers the amount of DNA present on an item through habitual use. Or a more direct study could test seatbelts in buses and compare them to the reference DNA of the regular driver, although this would be a more difficult task for many laboratories to perform.

#### Case example 2:

A POI was accused of stabbing the victim with a plastic handled knife. A swab of the handle of the knife produced a single sourced DNA profile that was conceded as originating from the POI. Prosecution claim that that it was the POI who used the knife to stab the victim. The POI claims that he was at a party shortly before the incident, where he shook hands with a male friend. His DNA could have been on the knife handle because he transferred his DNA to the hands of his friend, who then used the knife to stab the victim. In doing so the POI's friend transferred the POI's DNA onto the knife handle. The propositions are therefore:

- Hp            The POI stabbed the victim with the knife
- Hd2          The POI shook hands with a friend who stabbed the victim with the knife

In this case, the alternative proposition assumes an alternative offender (AO), hence Hd2 will be considered. The node “choice of Hd” has been set accordingly.

The same structure of the BN as shown in Figure 8 can be used, but instantiated differently. We do not provide the same explanation regarding node instantiations as we did for scenario 1 in Table 5. The choices themselves are not as important in these demonstrations as the construction and function of the *BN* itself. One can imagine what assumptions may be made regarding surface types and types of contact in regular casework, and if this information is

---

unknown then it is always possible to either leave the node(s) uninstantiated or to trial the effect of instantiating with different values. In this way the scientist can determine how important that piece of information is to the robustness of the *LR* provided. In this scenario we have a hand to hand transfer and have not obtained data for such situations. We consider that a surface type 'hand' is the same that a surface type 'rough'. However, DNA transfer and persistence on hands may not act the same as the data for 'rough' or 'smooth' surface types and ideally further considerations for the surface type 'hands' should be considered within the *BN*. We note that the DNA persistence on hands may be different due to bacterial degradation or normal wear and tear on hands from everyday activities.

What is interesting is that as the observed amount of the POI's DNA increases above 0.02ng, the support for a primary transfer over a secondary transfer is now above 1 and increases, until it reaches approximately 14 when DNA amounts are 0.8-0.9 ng. As DNA amount increases beyond 0.9 the level of support for a primary transfer decreases compared to a secondary transfer. Above 7ng, the findings start to support a secondary transfer over a primary transfer can be observed. (see Figure 13). This is a combination of the *BN* moving to a position where the POI is considered a high shedder and the AO a low shedder and the level of background DNA being higher with a coincidental matching alleles. Instantiating these factors so that they cannot be the case i.e. restricting the shedder status of the POI to an average value, removing the consideration of background DNA and specifying that the AO DNA profile is not matching with the POI sees the *LR* with propositions as stated above steadily increase with increasing POI DNA amounts.

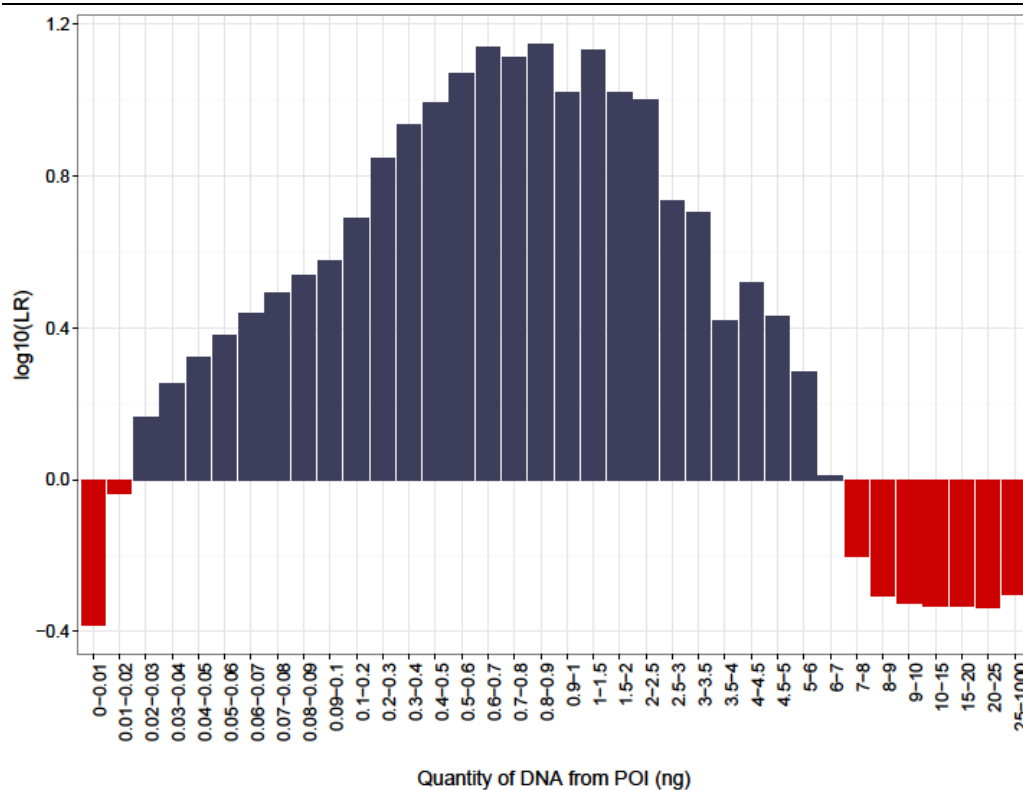


Figure 13: Ratio of probability for primary vs secondary transfer obtained for the instantiated network, but considering different levels of detected POI DNA, but always keeping the detected levels of the unknown individual's DNA as 0 – 0.01ng.

We could also create a similar BN for situations like case example 2, but where there are two POIs, one of which stabbed the victim, both of their DNA is detected on the item, and both having submitted reference DNA samples. This could be achieved by the addition of a secondary transfer route on for the AO (who would be considered the second POI) into the BN in Figure 8 with the 'Choice of Hd' node specifying either a primary transfer for POI1 and a secondary transfer for POI2 or vice versa. We do not provide a Figure showing such a BN.

### Scenario 3:

An unregistered firearm was found on a couch in the house owned by the POI. Police seized the firearm and a swab of the firearm stock revealed a mixed DNA profile originating from two individuals of which the POI was conceded as being the minor source. The prosecution claim that the POI handled the firearm. The defence claim that the POI had not known about the firearm and someone must have put it on his bed and then moved it to the couch very recently. The POI also claims that he saw something sitting on his bed earlier that day and

thinks it may have been the firearm, but he didn't investigate or touch it. The POI therefore states that the presence of his DNA on the firearm is due to it transferring from the POI to the couch, or bed, and then to the firearm. The propositions are therefore:

H<sub>p</sub> The POI recently handled the firearm

H<sub>d2</sub> The POI has never handled the firearm, but someone must have moved it

Here we have the deficiency of knowledge regarding the amount of DNA on a regularly used couch and bed. There is also the possibility of an accumulative effect, i.e. if someone rubs their hands on the same item multiple times on different occasions, does the DNA keep accumulating or does it reach a saturation point? At present this information is not known and so we make a number of assumptions that we explain below. These are further areas of research that would be beneficial for the forensic community to carry out.

Again we do not specify the reasoning behind each instantiation in the *BN*. The intention here is to demonstrate the adaptability and power of the *BN* to handle a variety of situations.

The extended *BN* for scenario 3 can be seen in Figure 14. We can see now that this is similar to the initial *BN* as seen in Figure 8, however we have the two routes for secondary transfer from the POI, one in relation to the bed, the other in relation to the couch. They meet in a node that adds the DNA amounts from secondary transfer together. In doing this we assume that if DNA has been transferred from both secondary pathways then the effect is pure accumulation and none of loss of DNA. This is another area which requires some research to be conducted.

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

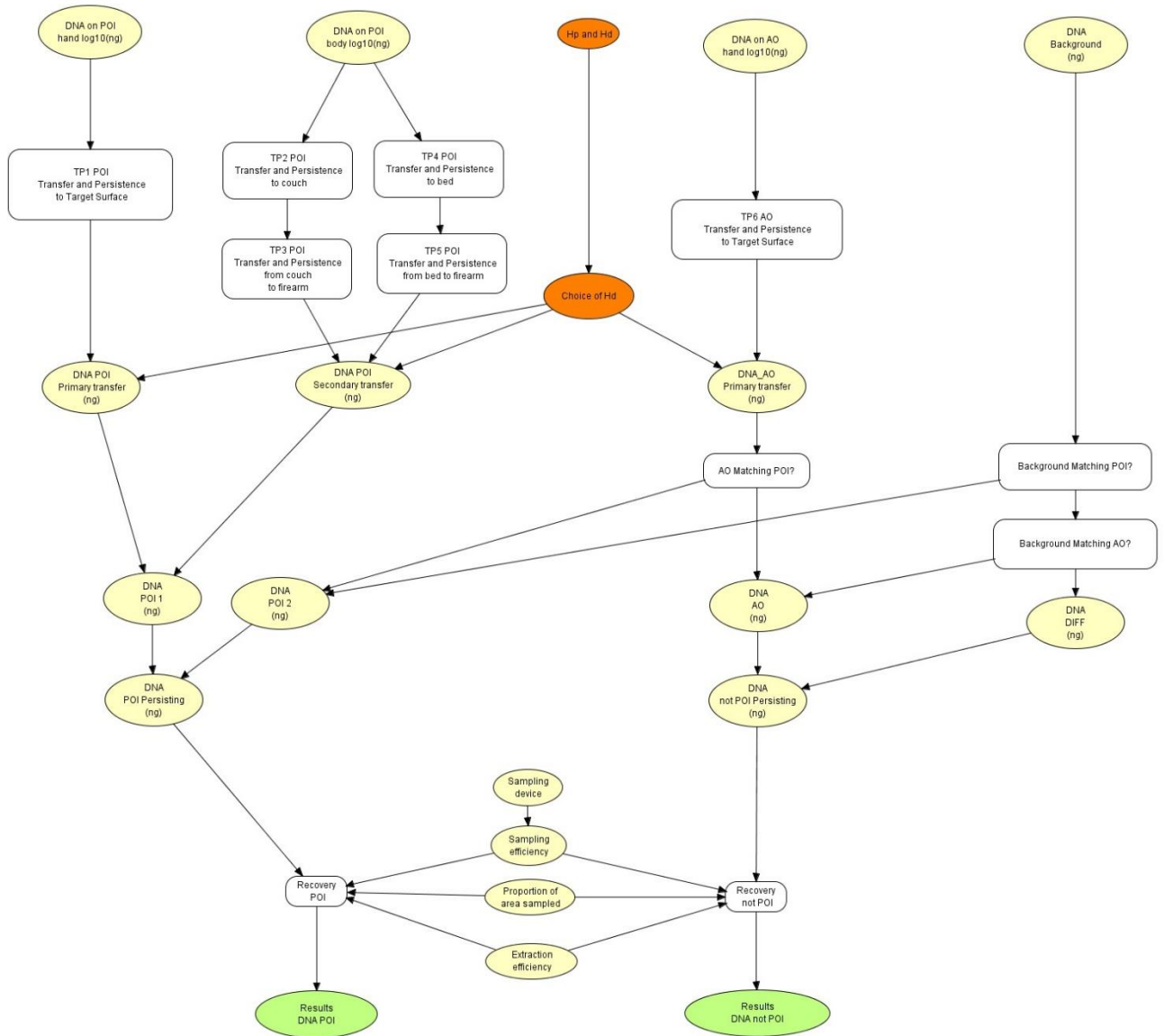
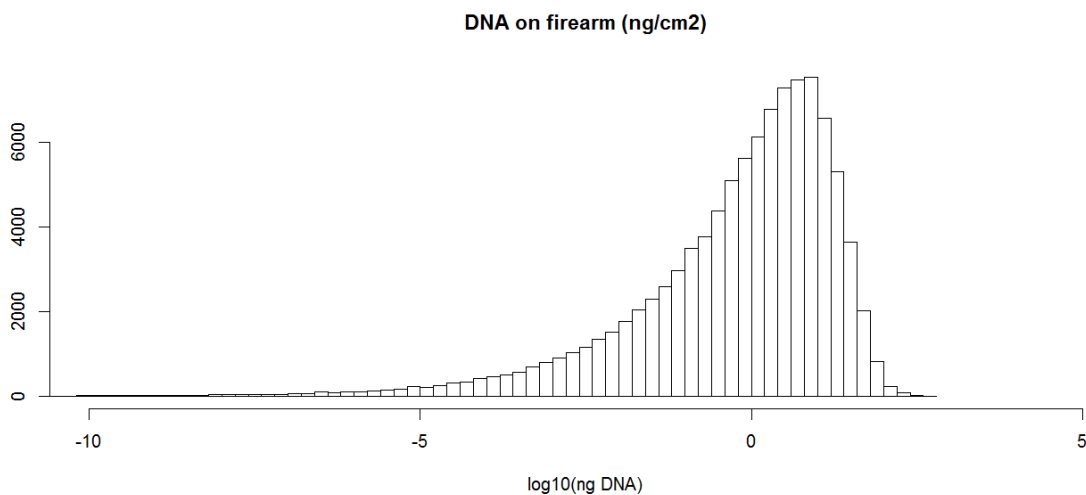


Figure 14: Bayesian network used to evaluate the findings with given activity level propositions involving primary vs secondary transfer event for scenario 3

For the background node we require the amounts of DNA found on firearm from individuals other than the primary user(s). The reason for this is that the primary user(s) will have transferred their DNA via a primary contact and we are interested in background levels (deposited through some unknown mechanism, and not by a known contact). This information is not readily provided by any literature sources we could find. The study of McKenna [51] found that a DNA profile was not observed in 26% of firearms swabs. A study by FSSA [52] examined DNA amounts from 300 firearm swabs obtained in casework. While ideally for our purpose this data would be compared to the owner or regular user of the firearm so they could be screened out of the DNA obtained, we use the distribution of DNA amounts observed from [52]. We then extrapolate back to DNA amounts found on the firearms in the same manner as previously described to obtain the distribution of DNA originally on the firearm, which we show in Figure 15. We apply the distribution values seen in Figure 15 directly into the Background DNA node.



*Figure 15: DNA amount (grey) obtained from swabs of firearms and fitted distribution (black)*

In this case example we assume that under the prosecution proposition that the contact by the POI to the firearm is a brief one-off contact (perhaps as an allegation of a recent purchase). If we were to assume an extended habitual use then different data would be required.

Using the *BN* seen in Figure 14 yields  $LR = 8$  when the amount of the POI's DNA detected is above approximately 1-1.5ng. The POI's DNA has come to be on the item either from a direct contact with his hands or from a secondary transfer from the POI's bed or couch transferred onto the firearm when it was placed there. The amount of DNA expected from such a transfer is less than what is possible from primary transfer. Figure 16 shows the LR over a range of DNA amount for POI and either relatively high or low DNA amounts from the non-POI.

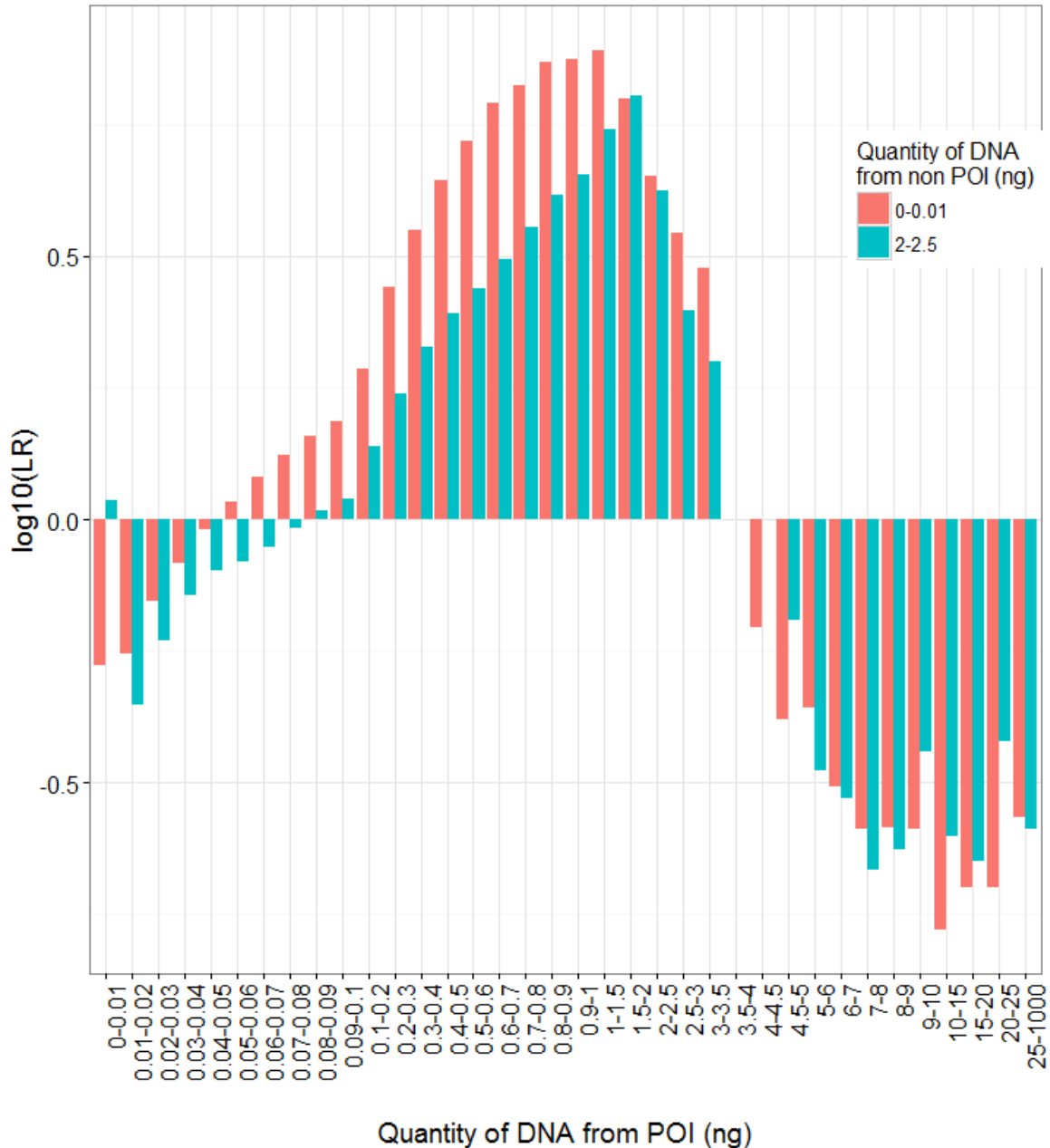


Figure 16: LR over range of detected DNA for POI and relatively high and low DNA amounts for non-POI.

Again, as the DNA amount for the POI increases from zero to 0.5ng the support for primary over secondary transfer increases, then as DNA increase further the support for primary transfer decreases until approximately 4ng, when the findings start to support secondary over primary transfer. At higher DNA levels the support for secondary over primary transfer is a product of the fact that the modelling of background DNA, coupled with the accumulation from multiple sources (couch and bed) means that higher amounts of DNA are more indicative of secondary transfer in this scenario.

This example shows the importance of a good set of data for modelling background DNA and levels of DNA expected on items from habitual use. The number of samples and methods in which they have been collected for our example, suggest that further work would need to be required in order to address the findings in consideration of the propositions in scenario 3. The closeness of the primary transfer probability to the background DNA probability in Figure 15 suggests that much of the data captured in the FSSA study are likely to be resulting from primary contacts.

As a point of interest, if the scenario were changed to one which stated that the unknown male rubbed the gun on the bed in a deliberate attempt to transfer DNA, then the support for secondary transfer over a primary transfer persists for quite high levels of observed POI DNA. Note that in the evaluation of all the evidence the court is likely to have quite different prior beliefs on whether someone else brought a firearm into the house, compared to that person then wishing to deliberately 'frame' the suspect. This shift in prior beliefs may well outweigh the differences obtained from the activity level considerations of the DNA findings, but of course it is not up to the scientist to base their decisions on such considerations.

## **APPLICATION TO CONTROLLED CASES**

We have explained the construction of an OOBN that can consider competing transfer mechanisms and demonstrated its use in several scenarios. We wish now to test the ability of the BN to distinguish known primary and secondary transfer events as described in scenario 2 of the stabbing case. We have used the work of Samie et al. [43], which importantly, was not used in any of the modeling to assign conditional probabilities in the BN. This allows the results of Samie et al. [43] - where the authors study primary DNA transfer to knife handles

---



during the action of stabbing- to act as a test set. For the purpose of this research, a series of secondary transfer experiments was this time performed. 12 experiments were carried out. In order to mimic stabbing under the second scenario (secondary transfer), the same four individuals (two males and two females) and same type of knives were used. The persons' hands were washed at 8 am. Around 11 am or 3 pm, one volunteer (POI, acting as the innocent suspect) was asked to shake hands with the volunteer who would act as the stabber (alternative offender, AO). The person then carried out normal activities in their office environment (i.e., having lunch or coffee with their colleagues, speaking with them etc.). Thirty minutes later, the AO was asked to 'stab' a cardboard box with a knife. Right afterwards, traces were collected using the double swab method. The following day, the experiment was repeated with another volunteer acting as the innocent POI (Figure 17). Each volunteer took the role of POI associated with each of the three stabbers. In order to limit background DNA (i.e., DNA present for unknown reasons), the knife was cleaned between each experiment (by using bleach, ethanol and leaving the items under UV light for 30 min). To monitor background, a negative control was taken from the knife after cleaning and before the experiment. Results were all negative (no DNA profile). To collect, extract and amplify the DNA, we have used the same method as described in Samie et al. [41]. However, here, DNA was quantified using the Investigator® Quantiplex (Qiagen) kit following standard protocols and the amplicons were analysed using a 3500 Genetic Analyser ABI (Applied Biosystem) and GeneMapper®IDX Software.

Then, the DNA profiles were interpreted using STRmix™ v2.3.05 [1, 53, 54, 55]. This forensic software has been developed to resolve mixed DNA profiles based on a continuous approach. The programme uses peak height information and statistically accounts for the possibility of degraded DNA and stochastic variation such as stutter, allelic drop-in and drop-out. It provides information regarding the mixture proportion and the weight that is given to the possible genotypes of the contributors. The weight is used to express how well a proposed genotype explains the profile. These information, combined with the total of DNA quantity obtained, allowed to inform the node "Results DNA POI" and "Results DNA non POI". The number of contributors was determined based on the number of the peaks detected at each locus, peak height balance information and how the experiments were planned (i.e., we expected the DNA of two persons in different proportions).

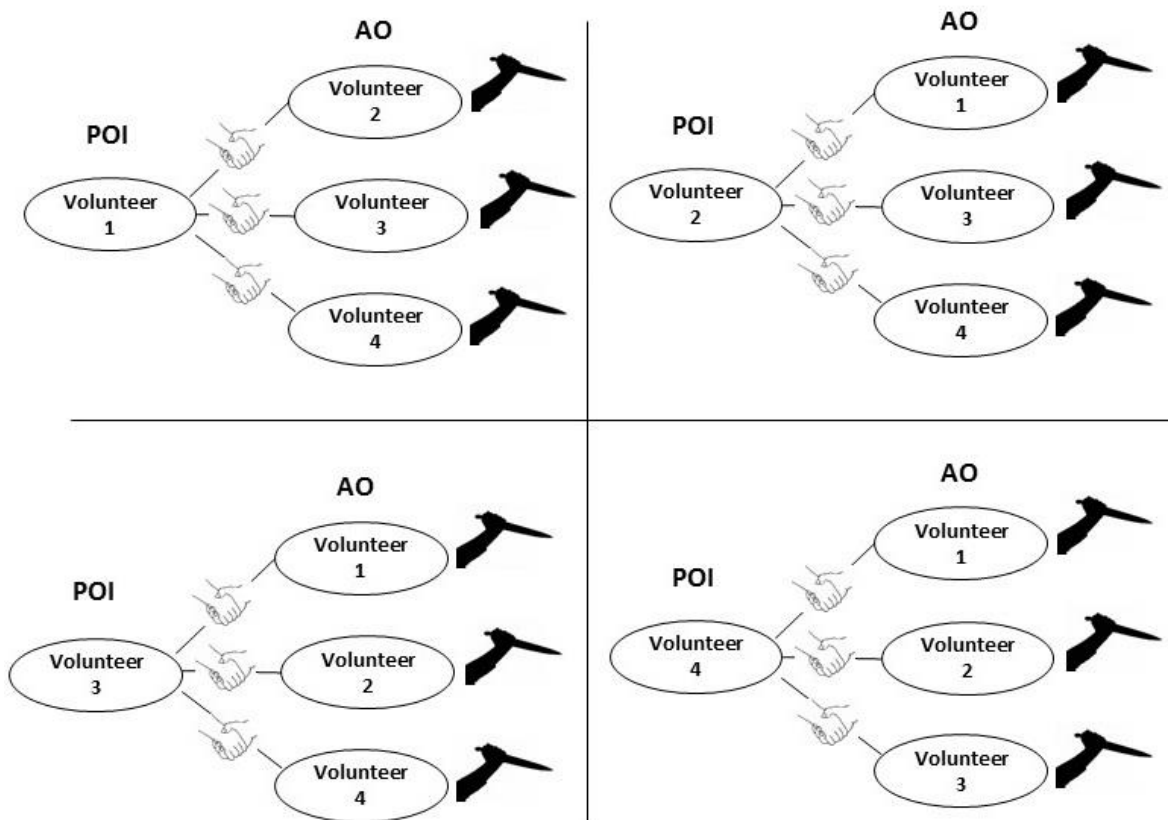


Figure 17: Figure showing the experimental design adopted for experiments on primary and secondary transfer.

$LR$  where  $H_p$  = primary transfer and  $H_d$  = secondary transfer

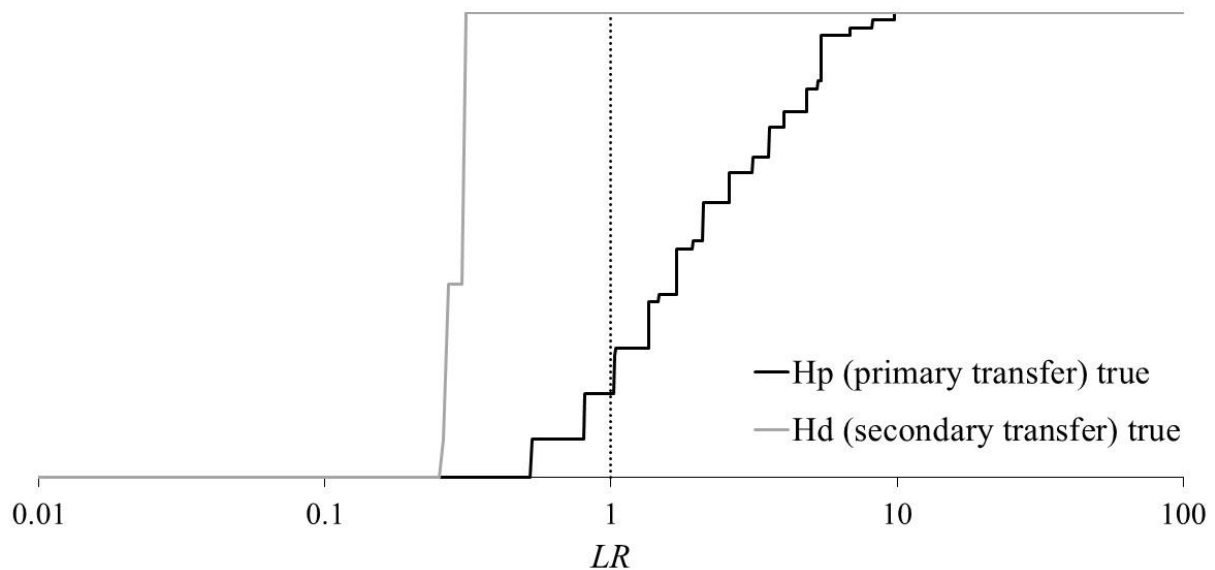


Figure 18: Tippett curve showing the ability of the developed BN to distinguish primary and secondary transfer events.

Figure 18 shows the results of considering each of the transfer events as primary. Given the limited dataset available, there appears that there is some ability for the system to distinguish primary from secondary transfer events, even without the knowledge of the shedder status of those involved.

## CONCLUSION:

We show here the construction of *BNs* built up from building blocks that address different features of contact DNA transfer, persistence and recovery. By splitting the transfer, persistence and recovery into separate nodes we make the *BN* high configurable to a number of situations, which we demonstrate through three quite different scenarios. The scenarios also highlight the importance of having the applicable data to inform conditional probabilities that underlie each node.

In theory the complexity of the scenarios that *BN* can consider are endless but in reality there will be a diminishing return as more complexity is added. There are likely to be key information that has the greatest effect on the posterior probabilities of the propositional node and these can be explored with sensitivity analyses (something which we have not

demonstrated in this paper, but intend to pursue to determine which factors the LR is most sensitive to, and also to help direct further research areas [56]).

There are a number of ways in which *BN* such as those we have developed here could be used other than to evaluate findings given competing transfer mechanisms. If the mechanism is known, then the *BN* could be used to investigate whether an individual is a high or low shedder. Alternatively, the *BN* could be used in a case assessment and interpretation framework (see [57] for an explanation of case assessment and interpretation), where items are triaged depending on the level of power that will exist to support one proposition over the other when transfer mechanism is disputed.

We show, in the 'application to controlled cases' section that the *BN* appears to have some ability to distinguish primary and secondary transfer under the conditions tested. This is despite not having information regarding the shedder status of the individual, which has been classically thought to be a limiting factor in the ability to evaluate evidence given competing transfer scenarios. Additional ground truth tests would be beneficial and would likely highlight situations where additional complexity, or additional modelling is required to inform the *BN*. In appendix 2 we suggest a number of studies we have already identified that would provide beneficial data to inform us on the important factors in transfer and persistence of trace DNA. As further information regarding the factors affecting DNA findings is obtained it is the authors' hope that they can be incorporated into evaluations in a logical and helpful manner.

The work here also brings up an interesting point regarding the modelling of 'background DNA'. This was particularly important in the assignment of probabilities in scenario 3. All DNA must have come to be on an item through some mechanism. Background DNA is defined as DNA that is expected to be present on an item, but is not related to the activity in question. Typically we consider that background DNA has not come from one of the POIs, but (as with any DNA source) may adventitiously possess the same alleles. When background DNA is modelled it is common practise to consider background DNA on an item as that which has not come from the primary handler. However, in a *BN* that it used to model multi-step transfer mechanisms the more correct modelling of background DNA would be DNA that has come from one more transfer step than the most complex mechanism being modelled i.e.

---

if one scenario being modelled is a tertiary transfer then background DNA would be DNA that has come from a quaternary transfer event. All other sources of DNA that have come from primary, secondary or tertiary transfers should be modelled in full. This is not an easy task to achieve, firstly due to the complexity of the BN that would arise, but secondly in the modelling of background DNA amounts at differing levels of transfer complexity.

There is always a difficulty when modelling data from the literature to find experimental designs that match the case scenario exactly. Indeed, some case scenarios simply cannot be exactly replicated in experimental work due to ethical or legal reasons. A common question arising from this is therefore, whether *BN* such as those constructed here can be applied to casework at all (see [58] for a discussion of this point and others in the same vein). Such a line of argument fails to recognise how information and casework circumstances are evaluated by the court. If the primary dispute when considering DNA evidence is one of a transfer mechanism, then the presence or absence of DNA is not in dispute. We then must ask, who is best placed to answer questions of transfer mechanism, which will inevitably require knowledge of transfer, persistence and recovery of DNA and the levels of background DNA in the environment. We suggest that it would be unrealistic to expect the average juror, judge or lawyer to possess such knowledge and that the highly specialised considerations are best explored by the scientist. Having made this decision, the question needs to be asked, how will the scientist take into account such a wide range of considerations and where will they draw their knowledge from. We would argue that the most logical and transparent manner in which this can be done is by setting out all the factors requiring consideration within a *BN* and populating probabilities using the most applicable data available. This will inform the scientist, which in turn can educate the court in the most robust manner possible. Having said all of this, we do not advocate the use of data which are clearly not suitable for assigning probabilities; but more often the case when some specific piece of information is not known, a reasonably close substitute can be used. As long as the conditions to which the data pertain (i.e. experimental settings) are judged appropriately close to be considered as an acceptable substitute for the case at hand then the resulting *BN* still represents the best evaluation of findings available. When this practice is adopted the use of the substitute data should be clearly pointed out in the report. One example of how this can be applied in the *BN* shown in our work is for the proportion of the hand that has contacted an item. Clearly, this information will not be available to the analyst. However, depending on the item an educated value can be

---

used, e.g. if a swab of the trigger of a firearm was taken then only the upper part of one finger is likely to have contacted the item and this can be portrayed in the values instantiated in the "Proportion of Hand contacted" node. If there is a high degree of uncertainty regarding the proportion of area contacted, then a distribution of prior probabilities across the states in the node can be applied to reflect this.

No-one is served by the scientist simply refusing to provide an opinion due to an absence of some small area of data that perfectly aligns with case circumstances. When the absence of knowledge is great (such as is scenario 3) then this will become apparent to the scientist and under these circumstances it is appropriate for them to state that the results cannot be evaluated robustly given the two competing propositions (see [56] for a discussion on this).

**ACKNOWLEDGEMENTS:**

Points of view in this document are those of the authors and do not necessarily represent the official position or policies of their organisations.

**REFERENCES:**

- [1] Taylor D, Bright J-A, Buckleton J. The interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics*. 2013;7:516-28.
- [2] Haned H. Forensim: An open-source initiative for the evaluation of statistical methods in forensic genetics. *Forensic Science International: Genetics*. 2011;5:265-8.
- [3] Lohmueller K, Rudin N. Calculating the weight of evidence in low-template forensic DNA casework. *Journal of Forensic Sciences*. 2013;58:s234-59.
- [4] Steele CD, Balding DJ. Statistical Evaluation of Forensic DNA Profile Evidence. *Annual Review of Statistics and Its Application*. 2014;1:361-84.
- [5] Puch-Solis R, Rodgers L, Mazumder A, Pope S, Evett I, Curran J, et al. Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters. *Forensic Science International: Genetics*. 2013;7:555-63.
- [6] Balding DJ, Buckleton J. Interpreting low template DNA profiles. *Forensic Science International: Genetics*. 2009;4:1-10.
- [7] Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, et al. Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences*. 2011;56:1430-47.
- [8] Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA. A hierarchy of propositions: Deciding which level to address in casework. *Science and Justice*. 1998;38:231-40.
- [9] Fonnelløp AE, Egeland T, Gill P. Secondary and subsequent DNA transfer during criminal investigations. *Forensic Science International: Genetics*. 2015;17:155-62.
- [10] Goray M, Oorschot RAHv. The complexities of DNA transfer during a social setting. *Legal Medicine*. 2015;17:82-91.
- [11] Meakin G, Jamieson A. DNA transfer: Review and implications for casework. *Forensic Science International: Genetics*. 2013;7:434-43.
- [12] Van Oorschot RA, Jones M. Retrieval of DNA from touched objects. *The 14th International Australia and New Zealand Forensic Science Society Symposium for Forensic Sciences*. Adelaide, Australia 1998.
- [13] Van Oorschot RAH, Ballantyne KN, Mitchell RJ. Forensic trace DNA: A review. *Investigative Genetics* 2010;1.
- [14] Van Oorschot RAH, Goray M, Eken E, Mitchell RJ. Impact of relevant variables on the transfer of biological substances. *Forensic Science International: Genetics Supplement Series*. In Press, Corrected Proof.
- [15] Farmen RK, Jaghø R, Cortez P, Frøyland ES. Assessment of individual shedder status and implication for secondary transfer. *Forensic Science International: Genetics Supplementary Series*. 2008;1:415-7.

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

---

- [16] Casey D, Clayson N, Jones S, Leisj J, Boyce M, Fraser I, et al. A response to Meakin and Jamieson DNA transfer: Review and implication for casework. *Forensic Science International: Genetics*. 2016;21:117-8.
- [17] Meakin G, Jamieson A. A response to a response to Meakin and Jamieson DNA transfer: Review and implications for casework. *Forensic Science International: Genetics*. 2016;22:e5-e6.
- [18] Taylor D, Abarno D, Champod C, Hicks T. Evaluating forensic biology results given source level propositions. *Forensic Science International: Genetics*. 2016;21:54-67.
- [19] Wolff TRD, Kal AJ, Berger CEH, Kokshoorn B. A probabilistic approach to body fluid typing interpretation: an exploratory study on forensic saliva testing. *Law, Probability and Risk*. 2015;14:323-39.
- [20] Evett IW, Gill PD, Jackson G, Whitaker J, Champod C. Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences*. 2002;47:520 - 30.
- [21] Gittelsohn S, Biedermann A, Bozza S, F. T. Bayesian Networks and the Value of the Evidence for the Forensic Two-Trace Transfer Problem. *Journal of Forensic Sciences* 2012;57:1199-216.
- [22] Biedermann A, Bozza S, Taroni F. Probabilistic evidential assessment of gunshot residue particle evidence (Part I): Likelihood ratio calculation and case pre-assessment using Bayesian networks. *Forensic Science International*. 2009;191:24-35.
- [23] Wieten R, Zoete JD, Blankers B, Kokshoorn B. The interpretation of traces found on adhesive tapes. *Law, Probability and Risk*. 2015;14:305-22.
- [24] Hicks T, Biedermann A, Koeijer JAd, Taroni F, Champod C, Evett IW. The importance of distinguishing information from evidence/observations when formulating propositions. *Science & Justice*. 2015;55:520-5.
- [25] Curran JM, Hicks TN, Buckleton J. *Forensic Interpretation of Glass Evidence*. London: CRC Press; 2000.
- [26] R v Peter Weller [2010] EWCA Crim 1085.
- [27] Lehmann VJ, Mitchell RJ, Ballantyne KN, Oorchot RAHv. Following the transfer of DNA: How does the presence of background DNA affect the transfer and detection of a target source of DNA? *Forensic Science International: Genetics*. 2015;19:68-75.
- [28] Butts E. Exploring DNA Extraction Efficiency Forensics@NIST 2012 Meeting, Gaithersburg. 2012;[http://www.nist.gov/oles/3\\_Butts-DNA-extraction-2.pdf](http://www.nist.gov/oles/3_Butts-DNA-extraction-2.pdf).
- [29] van Oorschot RAH, Phelan DG, Furlong S, Scarfo GM, Holding NL, Cummins MJ. Are you collecting all the available DNA from touched objects? *International Congress Series*. 2003;1239:803-7.
-



- [30] de Jong B, Antonie M, Hardon A, van den Bergen D, Kneppers A. Successful validation of a fully automated sample lysis workstation. *Forensic Science International: Genetics Supplement Series*. 2013;4:e93-e4.
- [31] Verdon TJ, Mitchell RJ, Oorschot RAHV. Evaluation of tapelifting as a collection method for contact DNA. *Forensic Science International: Genetics*. 2014;8:179-86.
- [32] Raymond JJ, Oorschot RAHV, Gunn PR, Walsh SJ, Roux C. Trace evidence characteristics of DNA: A preliminary investigation of the persistence of DNA at crime scenes. *Forensic Science International: Genetics*. 2009;4:26-33.
- [33] Quinones I, Daniel B. Cell free DNA as a component of forensic evidence recovered from touched surfaces. *Forensic Science International: Genetics*. 2012;6:26-30.
- [34] Buckingham AK, Harvey ML, van Oorschot RAH. The origin of unknown source DNA from touched objects. *Forensic Science International: Genetics*. 2016;25:26-33.
- [35] Lowe A, Murray C, Whitaker J, Tully G, Gill P. The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. *Forensic Science International*. 2002;129:25-34.
- [36] Phipps M, Petricevic S. The tendency of individuals to transfer DNA to handled items. *forensic Science International*. 2007;168:162-8.
- [37] Lacerenza D, Aneli S, Omedei M, Gino S, Pasino S, Berchiolla P, et al. A molecular exploration of human DNA/RNA co-extracted from the palmar surface of the hands and fingers. *Forensic Science International: Genetics*. 2016.
- [38] Berge Mvd, Ozcanhan G, Zijlstra S, Lindenbergh A, Sijen T. Prevalence of human cell material: DNA and RNA profiling of public and private objects after activity scenarios. *Forensic Science International: Genetics*. 2016;21:81-9.
- [39] Bontadelli L. Study of DNA shedder quality. School of Criminal Justice of Lausanne, University of Lausanne. Lausanne; 2009.
- [40] Plummer M. Bayesian graphical models using MCMC. *RJAGS*. 2012.
- [41] Goray M, Fowler S, Szkuta B, Oorschot RAHV. An analysis of self and non-self DNA in multiple handprints deposited by the same individuals over time. *Forensic Science International: Genetics*. 2016;23:190-6.
- [42] Meakin G, Butcher EV, Van Oorschot RAH, Morgan RM. The deposition and persistence of indirectly-transferred DNA on regularly-used knives. *Forensic Science International: Genetics Supplement Series* 5. 2015:e498–e500.
- [43] Samie L, Hicks T, Castella V, Taroni F. Stabbing simulations and DNA transfer. *Forensic Science International: Genetics*. 2016;22:73–80.
-

- [44] Daly DJ, Murphy C, McDermott SD. The transfer of touch DNA from hands to glass, fabric and wood. *Forensic Science International: Genetics*. 2013;6:41-6.
- [45] Goray M, Mitchell RJ, Oorschot RAHv. Investigation of secondary transfer of skin cells under controlled conditions. *Legal Medicine*. 2010;12:117-20.
- [46] Poetsch M, Bajanowski T, Kamphausen T. Influence of an individual's age on the amount and interpretability of DNA left on touched items. *International Journal of Legal Medicine*. 2013;127:1093–6.
- [47] Ladd C, Adamowicz MS, Bourke MT, Scherzinger CA, Lee HC. A systematic analysis of secondary DNA transfer. *Journal of Forensic Sciences*. 1999;44:1270 - 2.
- [48] Van Oorschot RA, Jones M. DNA fingerprints from fingerprints. *Nature*. 1997;387:767.
- [49] Saravo L, Spitaleri S, Piscitello D, Travali S. DNA typing from steel cable. *International Congress Series*. 2004;1261:473–5.
- [50] Breathnach M, Williams L, McKenna L, Moore E. Probability of detection of DNA deposited by habitual wearer and/or the second individual who touched the garment. *Forensic Science International: Genetics*. 2016;20:53-60.
- [51] McKenna L. Understanding DNA results within the case context: importance of the alternative proposition. *Frontiers in Genetics*. 2013;4:1-4.
- [52] Sly N, Fietz P, Windram K, Windram R. The success of trace DNA recovery using STRmix. The 22nd International ANZFSS Symposium. Adelaide, South Australia 2014.
- [53] Bright JA, Taylor D, Curran JM, Buckleton JS. Developing allelic and stutter peak height models for a continuous method of DNA interpretation, *Forensic Science International: Genetics*. 2013; 7:296–304.
- [54] Bright JA, Taylor D, Curran JM, Buckleton JS. Degradation of forensic DNA profiles, *Australian Journal of Forensic Sciences*. 2013; 45(4):445–449.
- [55] Bright JA, Taylor D, Curran JM, Buckleton JS. Searching mixed DNA profiles directly against profile databases, *Forensic Science International: Genetics*, 2014; 9:102–110.
- [56] Taylor D, Hicks T, Champod C. Using sensitivity analyses in Bayesian Networks to highlight the impact of data paucity and direct future analyses: a contribution to the debate on measuring and reporting the precision of likelihood ratios. *Science & Justice*. 2016;56:402-10.
- [57] Cook R, Evett IW, Jackson G, Jones PP, Lambert JA. A model for case assessment and interpretation. *Science and Justice*. 1998;38:151-6.
- [58] Biedermann A, Champod C, Jackson G, Gill P, Taylor D, Butler J, et al. Evaluation of forensic DNA traces when propositions of interest relate to activities: analysis and discussion of recurrent concerns. *Frontiers in Genetics*. 2016;(accepted).

**Appendix 1: Tables that describe elements of the BN in Figure 8 to 11**

Node	Categories	Values
Hp and Hd	Hp Hd	uniform prior
Choice of Hd	Hp: Primary transfer Hd1: Secondary transfer without AO Hd2: Secondary transfer with AO	uniform prior or adapted as a function of the alleged activities put forward by the defence.
DNA on POI hand log <sub>10</sub> (ng)	interval node; From -inf to -1.5 From -1.5 to 3 in steps of 0.1 3 to 4 in steps of 0.5	N(0.66, 0.49)
DNA on AO hand log <sub>10</sub> (ng)	interval node; From -inf to -1.5 From -1.5 to 3 in steps of 0.1 3 to 4 in steps of 0.5	N(0.66, 0.49)
DNA POI Primary transfer (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	$10^{\text{DNA OUT}}$ from TP if Hp is true 0 otherwise
DNA POI Secondary transfer (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	$10^{\text{DNA OUT}}$ from TP if Hd1 or Hd2 is true, 0 otherwise
DNA AO Primary transfer (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	$10^{\text{DNA OUT}}$ from TP if Hd2 is true 0 otherwise
DNA Background (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5	uniform prior  Note the probabilities provided in this node will need to be tailored to the item and circumstances of the case. A uniform distribution is not a realistic description of background DNA on most items.

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

	25 to 1000 1000 to inf	
DNA POI 1 (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	DNA POI primary transfer + DNA POI secondary transfer
DNA POI 2 (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	AO matching POI + Background matching POI
DNA AO (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	AO DIFF POI + Background matching AO
DNA DIFF (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	Background DIFF AO
DNA POI persisting (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	DNA POI 1+ DNA POI 2
DNA not POI Persisting (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1	DNA AO + DNA DIFF

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

	<p style="text-align: center;">10 to 25 in steps of 5 25 to 1000 1000 to inf</p>	
Results DNA POI	<p style="text-align: center;">interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf</p>	DNA OUT from R
Results DNA not POI	<p style="text-align: center;">interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf</p>	DNA OUT from R
Proportion of area sampled	interval node; values from 0 to 1 in steps of 0.1	Uniform Prior
Sampling device	Tapelift Swab	Uniform Prior
Sampling efficiency	interval node; from 0 to 1 in steps of 0.05	B(1.9,16.6) If Sampling device is Tapelift B(25,20) If Sampling device is Swab
Extraction efficiency	interval node; from 0 to 1 in steps of 0.05	Beta(5,17)

*Tables 1: Expressions and probabilities for nodes that underlie the main BN in Figure 8*

Node	Categories	Values
DNA IN	interval node; from -inf to -1.5 then from -1.5 to 3 in steps of 0.1 then from 3 to 4 in steps of 0.5	Input node
Proportion of hand/surface contacted	interval node; from 0 to 1 in steps of 0.1	uniform prior
DNA available to transfer log <sub>10</sub> (ng)	interval node; From -inf to -1.5 From -1.5 to 3 in steps of 0.1 3 to 4 in steps of 0.5	DNA IN + log <sub>10</sub> (Proportion)
Vigour of contact with target surface	Passive Pressure Friction	Uniform Prior
Nature of shedding surface	Hand Smooth Rough	Uniform Prior
Nature of	Hand	Uniform Prior

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

target surface	Smooth Rough	
Transfer proportion	interval node; from 0 to 1 in steps of 0.05	B(0.89,2.25) if Surface=Rough/Hand & Vigour= Passive B(0.49,24.11) if Surface= Smooth & Vigour= Passive  0.33B(0.77,1.45) + 0.67B(1,1) if Surface= Rough/Hand & Vigour= Pressure B(0.47,30.85) if Surface= Smooth & Vigour= Pressure  0.2B(0.7,1.64) + 0.8B(1,1) if Surface= Rough/Hand & Vigour= Friction B(0.45,1.13) if Surface= Smooth & Vigour= Friction
DNA deposited on target surface log <sub>10</sub> (ng)	interval node; From -inf to -1.5 From -1.5 to 3 in steps of 0.1 3 to 4 in steps of 0.5	DNA available to transfer + log <sub>10</sub> (Transfer proportion)
Days between both transfers or transfer and recovery	interval node; 0, 0.5 then 1 to 31 in steps of 1	Uniform Prior
Environmental condition of target surface kept in	Favourable  Poor	Uniform Prior
alpha 2	-0.052 -0.022	1 if Environmental is poor, 0 otherwise 1 if Environmental is favourable, 0 otherwise
DNA persisting on target surface (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	$10^{\text{DNA deposited}} e^{\alpha 2 * \text{Days}}$
DNA OUT	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	log <sub>10</sub> (DNA persisting)

Tables 2: Expressions and probabilities for nodes that underlie the main BN in Figure 9

Node	Categories	Values
DNA IN	interval node; 0 to 0.1 in steps of 0.01	Input node

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

	0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	
Proportion of area sampled	interval node; values from 0 to 1 in steps of 0.1	Input node
Sampling efficiency	interval node; from 0 to 1 in steps of 0.05	Input node
DNA recovered from sampling	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	DNA IN * Proportion* Sampling efficiency
Extraction efficiency	interval node; from 0 to 1 in steps of 0.05	Input node
DNA recovered from extraction	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	DNA recovered from sampling * Extraction efficiency
DNA OUT	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	DNA recovered from extraction

Tables 3: Expressions and probabilities for nodes that underlie the main BN in Figure 10

Node	Categories	Values
DNA (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1	Input node

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

	10 to 25 in steps of 5 25 to 1000 1000 to inf	
Match Probabilities	True False	True with probability between 0.3 and 10E-9 depending on DNA profile
DNA Matching (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	DNA (ng) if Match Probabilities true, 0 otherwise
DNA DIFF (ng)	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf	DNA (ng) if Match Probabilities false, 0 otherwise

*Tables 4: Expressions and probabilities for nodes that underlie the main BN in Figure 11*



## **Appendix 2: Further studies which would assist in answering questions of transfer mechanism**

It is important that the studies reflect as much casework circumstances as possible: studies that try and maximise transfer cannot be used for casework purposes. Data collection and data treatment should also reflect casework procedures. Here, we have used data where quantity of DNA was available. What would be ideal is to report the quantity of DNA relative to each contributor. This can be estimated by combining the quantity and the relative contribution of the persons to the mixture (such as in [41]). Through this work we have identified a number of studies which would assist in evaluation of evidence in light of propositions that suggest differing DNA transfer mechanisms. We collectively provide the list below:

1. Trialling extraction efficiencies at low DNA levels. The study of Butts [28] trialled DNA amounts from 24ng to 4800ng.
2. Trialling of different extraction techniques not tried in the Butts [28] study. Note that some information to this effect can be found in [29].
3. Taking into account the nature of the surface type the DNA has been placed on when considering DNA persistence.
4. Investigation into different environmental conditions (e.g. rain, washed, full sun, etc) on DNA persistence.
5. More data to confirm the DNA persistence rates found in Raymond *et al.* [32].
6. Data on DNA persistence of DNA on objects after extended handling by other individuals (of which some work has been done in [34]), or from physical movements after initial deposition.
7. Data on DNA persistence in standard exhibit packaging
8. Shedder consistency studies, i.e. whether an individual sheds DNA consistently in the upper or lower quantiles of the population shedding distribution and whether this is noticeable through events such as washing or sweating. Studies could extend to a standard method for determining the approximate shedding propensity of an individual for use in primary vs secondary transfer considerations.
9. A study of how the amount of time an object is held affects the amount of DNA transferred. To date we have found such information in [47] for hand to hand transfers only by length of handshake.

10. A study of the absolute amount of DNA transferred from hands to objects for different contact types e.g. light touch, pressure, friction (i.e. so that data from multiple studies does not need to be combined and extrapolated as we have done here).
11. Consideration of the amount of DNA transferred to object from a habitual use e.g. items in the home. Some work has been done in this area, such as [50].
12. Transfer DNA amount for varying length of time of contact between primary and secondary substrates and for different contact types.
13. A study into the level of transfer from hand-to-hand for various contact types (e.g. handshake, high-five, clasping, struggling) for different times of contact and the persistence of DNA on hands through various timeframes and activities.
14. The levels of background DNA (i.e., not from known users) on various items.
15. The levels of primary user's DNA on regular use items such as furniture or objects around the home or office.
16. A study into the accumulation of contact DNA on items resulting from multiple contacts from the same person.
17. All studies used in the construction of the *BN* in this paper concentrate on DNA amount. The rate of degradation across the profile may also have some power to distinguish factors involved in transfer, persistence and recovery (particularly persistence). There are a number of experiments, that would be worthy of study in this area. For those experiments that have already been carried out (many of which we reference) the data already exists, it just requires analysis of a different measured variable.

I.2 Samie et al., The efficiency of DNA extraction kit and the efficiency of recovery techniques to release DNA using flow cytometry, *Science & Justice*. 59 (2019) 405-410

Lydie Samie<sup>1</sup>, Christophe Champod<sup>1</sup>, Valérie Glutz<sup>2</sup>, Miguel Garcia<sup>2</sup>, Vincent Castella<sup>3</sup>, Franco Taroni<sup>1</sup>

1. Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice, University of Lausanne, Switzerland

2. Plateforme technologique de cytométrie en flux, Ecole Polytechnique Fédérale de Lausanne, Switzerland

3. Forensic Genetics Unit, University Center of Legal Medicine Lausanne and Geneva, Lausanne, Switzerland

## **Introduction**

In forensic investigations, low levels of DNA are often recovered from touched surfaces. As recommended by the ENFSI Guideline for Evaluative Reporting in Forensic Science [1], the evaluation of these DNA traces should be carried out using activity-level propositions which involves a relative assessment of the expected quantities of recovered DNA under the alleged activity depending on the propositions of interest. In order to do so, the quantity of the recovered DNA plays an important role and the efficiency of DNA extraction kit is one of the variables that should be considered [2]. Without knowledge of the extraction efficiency of the kit used by the laboratory, a meaningful evaluation of the findings would not be possible for DNA expertise or research. This study aims at showing how the efficiency of DNA extraction kits and the yield of release of cells from swabs can be measured.

Only a few studies dealt with the efficiency of extraction kits for traces of low levels of DNA [3, 4]. In Browlow *et al.* [3] the obtained measure of extraction efficiency jointly considered the type of surface and the efficiency of the swab used to collect and then release the cells and DNA; however, the sole efficiency of the extraction kit alone remains unknown because DNA traces were deposited on a surface. In Wood *et al.* [4], the efficiency of recovery techniques was evaluated from recovery up to the release of cells and DNA. While this considers the ability of the DNA swabs to release cells and DNA, which is a variable that affects the overall efficiency of the DNA extraction process, the efficiency of the extraction kit itself remains unknown since it combines the extraction efficiency of the kit and that of the release of cells

---

and DNA. To measure its specific efficiency of extraction, one needs to know the initial quantity of DNA to be extracted. Flow cytometry is cited by Butts [5] as the most appropriate method to select a low number of cells to be used as the starting material for the measure of the extraction yield. In this research, we used flow cytometry to prepare constant number of cells that will be directly submitted to the extraction procedures or deposited on swabs.

Extraction kits are used by different persons from different laboratories, operating manually or using automated platforms, which influences the extraction efficiency. The impact of the laboratory is reported as well.

This study has three objectives. The first is to measure the extraction efficiency of two commercial DNA extraction kits (Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit, and QIAshredder-QIAamp DNA Mini kit from Qiagen with Microcon® 30 spin column) used to extract and purify low quantities of DNA based on initial quantities of DNA obtained using flow cytometry. The second is to study the impact of the laboratory on the yield offered by the best performing kit (Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit from Qiagen). The last is to report on the efficiency of a swab (FLOQSwab™ from COPAN) to release cells and show how to obtain it by the combined usage of a swab and an extraction kit (QIAshredder-QIAamp DNA Mini kit from Qiagen with Microcon® 30 spin column).

## **Methodology**

### *Type and number of cells*

The method adopted here starts from a given and known number of cells obtained by cell cytometry. The cells were selected using the P658282Z3001 FACS Aria IIu cytometer with FACSDiva 8.0.1 version application.

The type of cells chosen for this study is adult keratinocytes, which are typical of skin cells. Epidermal keratinocytes cell culture (Human Epidermal Keratinocytes – Neonatal) from Lonza was performed according to manufacturer's instructions. In order to avoid cell differentiation, cells were passed before they reached 80% of confluence and we minimized the doubling population. Cells were sorted after two population doublings. Propidium Iodide staining was used to sort the nucleated, living, cells.

To select the number of cells representing a quantity of DNA obtained when touching a surface, different numbers of cells were tested. First, four samples of 50, 100, 500 and 5000

---

cells were prepared respectively twice, then directly introduced into a microtube of 1.5mL containing 180µL of a tissue lysis buffer (ATL buffer from Qiagen). Cell concentration was around 1million/ml and generates a flow rate of 900 events/sec. Given this concentration, the “Single-cell” as the mode of precision used was chosen.

The extractions of these eight samples were performed using the combination of two kits: QIAshredder and QIAamp DNA Mini kit from Qiagen, concentrated to a final volume of 25µL with Microcon® 30 spin column. To simplify, these kits will be denoted as QIAshredder-QIAamp DNA Mini kit. The quantities of results obtained on the four numbers of cells are given in Table 1.

**Table 1: Table representing the average extracted quantity of respectively 50, 100, 500, 5000 cells**

Number of cells obtained by cell cytometry [cell]	50	100	500	5000
The average quantity of DNA obtained using the QIAshredder-QIAamp DNA Mini kit [pg]	125	250	1200	15000

One-hundred cells have been selected for the experiments as it led to an amount of around 250 pg of DNA, which corresponds to the average amount of DNA obtained in a previous study focusing on DNA traces, obtained when touching a surface [6].

#### *Extraction efficiency of the kits*

For each kit, extractions were made based on an initial preparation of 100 cells. Cell concentration was low, generating a flow rate of around 20-40 events/sec. The “Purity” precision mode was selected in order to increase the probability where a cell of interest could be sorted.

The cells were directly introduced into each of the baskets containing 60µl of Phosphate buffered saline (PBS) of pH 7.4, allowing the cells to be kept intact. The kits were used following manufacturer’s instructions. Quantifications were performed directly following the DNA extraction using the Investigator® Quantiplex kit from Qiagen on Rotor-Gene® Q according to the manufacturer’s protocols. 30 extractions were performed using the QIAshredder-QIAamp DNA Mini kit, following the body fluid protocol, concentrated to a

final volume of 25µL with Microcon® 30 spin column, whereas 22 extractions were made with the Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit from QIAGEN with a final volume of 60µL without using microcon® 30 spin column, due to laboratory constraints. The difference between the two kits is the use of Spin basket for the Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit from QIAGEN instead of QIAshredder column and Microcon® 30 spin column.

### *Effect of the laboratory*

The kit which was proven to be the best performing kit is the Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit from Qiagen. To study the impact of the laboratory's performance on the yield offered by this kit, the extractions were performed manually by two operators in two different laboratories (Figure 1). One-hundred cells were selected, using the "Purity" precision mode, then directly introduced into each of the 44 Lyse&Spin baskets containing 60µl of Phosphate buffered saline (PBS) of pH 7.4, allowing the cells to be kept intact. Twenty-two extractions were made by each operator, with a final volume of 60µL. All the quantifications were performed together in the same run at the same time following the DNA extraction which was made two days after the flow cytometry.

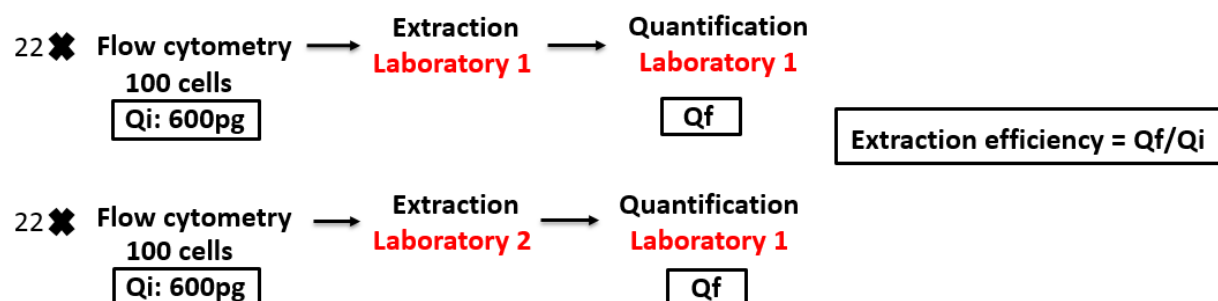


Figure 1: Illustration of the method used to study the impact of the laboratory on the yield offered by Investigator® Lyse&Spin basket-QIAamp DNA Mini kit. Qi is the initial quantity of DNA to be extracted, whereas Qf is the final extracted quantity of DNA.

*Release of cells by the DNA swab*

Figure 2 describes the method used to study the efficiency of the FLOQSwab™ to release cells. The measure of the extraction efficiency for the QIAshredder-QIAamp DNA Mini kit has been already measured (see *Extraction efficiency of the kits*). In the following experiment, we will measure the joint yield (swab cells release and DNA extraction).

To measure it, 20000 cells were introduced into a microtube of 1,5mL containing 1.1 mL of PBS, to avoid the destruction of the plasma membranes. Because of the technical impossibility to directly deposit cells on the swab, the microtube was mixed by vortexing and 35µL (636 cells) was pipetted on each 30 FLOQSwab™. To take into account the possible loss of cells being retained by the swab, the selected number of cells is higher than the number (100) used to study the extraction efficiency.

Swabs were dried during the afternoon before performing the DNA extraction using the QIAshredder-QIAamp DNA Mini kit. A concentrated final volume of 25µL was obtained at the end of the extractions using Microcon® 30 spin column. These 30 samples allowed for obtaining a joint measure of efficiency to release cells combined with the efficiency of the DNA extraction kit.

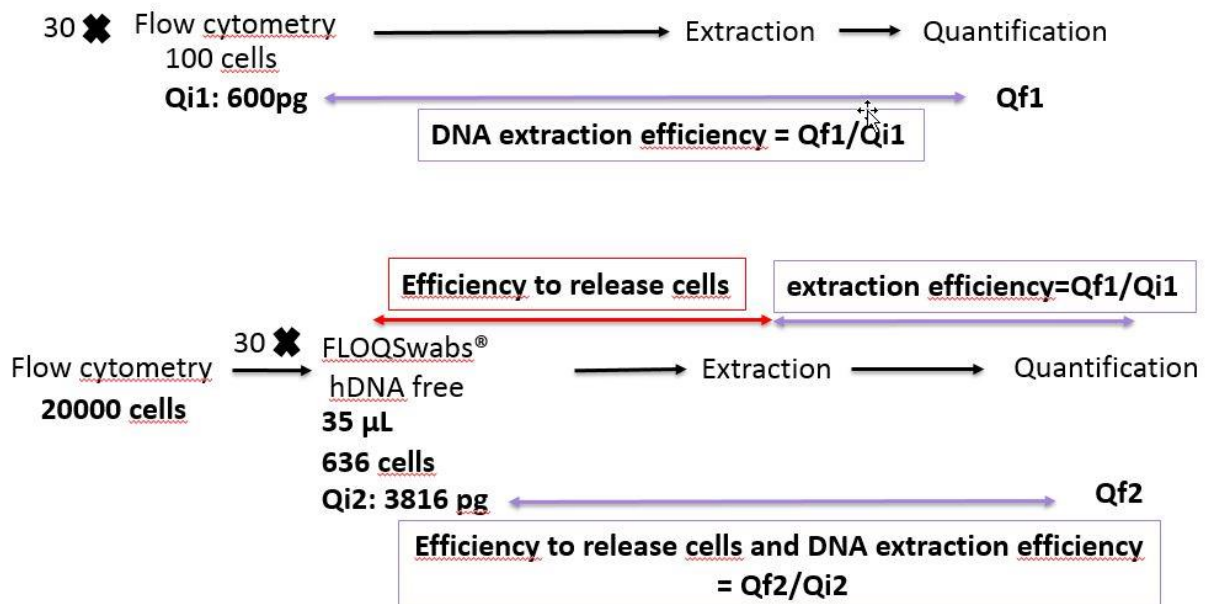


Figure 2: Illustration of the method used to obtain the extraction efficiency of the kit and a joint measure of efficiency to release cells combined with the efficiency of the DNA extraction kit (in purple) in order to obtain the efficiency of the sampling device to release cells (in red).

### *Calculating efficiency*

The efficiency is measured by the ratio between the initial quantity of DNA (approximated in pg) and the final quantity of DNA (measured in pg after quantification). The initial quantity of DNA is related to the weight associated with 100 cells obtained by flow cytometry. There is an average of 6pg per cell [7] based on the following formula:

Average DNA quantity per cell = Average number of base pair per cell  $\times 2 \times$  average molecular weight of one base /  $N_A$

Hence: Average DNA quantity per cell =  $3 \times 10^9 \times 2 \times 660_{(g/mol)} / (6,022 \times 10^{23}_{(mol^{-1})})$

Using an average of 6pg of DNA per cell, the initial quantity of DNA was set to 600pg. The final quantity of DNA is the product of the concentration obtained after quantification and the volume left at the end of the extraction.

For the swab measure of release, the initial quantity of DNA is known: 636 cells were initially deposited on the FLOQSwab™ from COPAN. The quantity of cells released by the swab corresponds to the quantity of cells available for next extraction step (Figure). This quantity is unknown, but will be measured indirectly after the measure of the extracted quantity of DNA with the QIAshredder-QIAamp DNA Mini kit from Qiagen. The results obtained previously on the extraction kit alone will be used to infer the swab cells release performance. This is illustrated in Figure 3 below.

The choice of the Beta distributions is motivated by the nature of the measured variable (a proportion). Beta distributions are ideally suited to model distributions between 0 and 1 (or 0% to 100%).

Mean and standard deviation of the distribution of the DNA extraction efficiency of the kit itself are known. Mean and standard deviation of the joint efficiency to release cells and extract DNA are also known following the above measurements.





Figure 3: Illustration of the extraction efficiency, of the efficiency to release cells and of the efficiency to release cells then extract DNA, with the parameters associated with each distribution that is known (in purple) or unknown (in red).

By assuming that both extraction and release contribute jointly to the final product, it is easy to find parameters  $c$  and  $d$  of the beta distribution representing the efficiency of the swab to release cells. Dufresne [8] gives the equations of the moments for the product of two Beta distributions. The parameters of a Beta distribution can be defined based on the mean and the variance of the distribution [9]. Solving an equation with two unknowns, we obtain these parameters “ $c$ ” and “ $d$ ” as follows:

$$c = \frac{X^2 - XY}{XY + Y}$$

$$d = \frac{X - Y}{XY + Y}$$

With:

$$X = \frac{\text{mean3}/\text{mean1}}{1 - (\text{mean3}/\text{mean1})}$$

And

$$Y = \frac{\text{Sd3}}{\text{mean3} * \text{mean1}} * \frac{(a + b + 1)}{a + 1}$$

## Results

### *Efficiency of the extraction kits*

Figure 4 presents the DNA extraction efficiency obtained on the 22 and 30, respectively, samples following the extraction using each extraction kit:

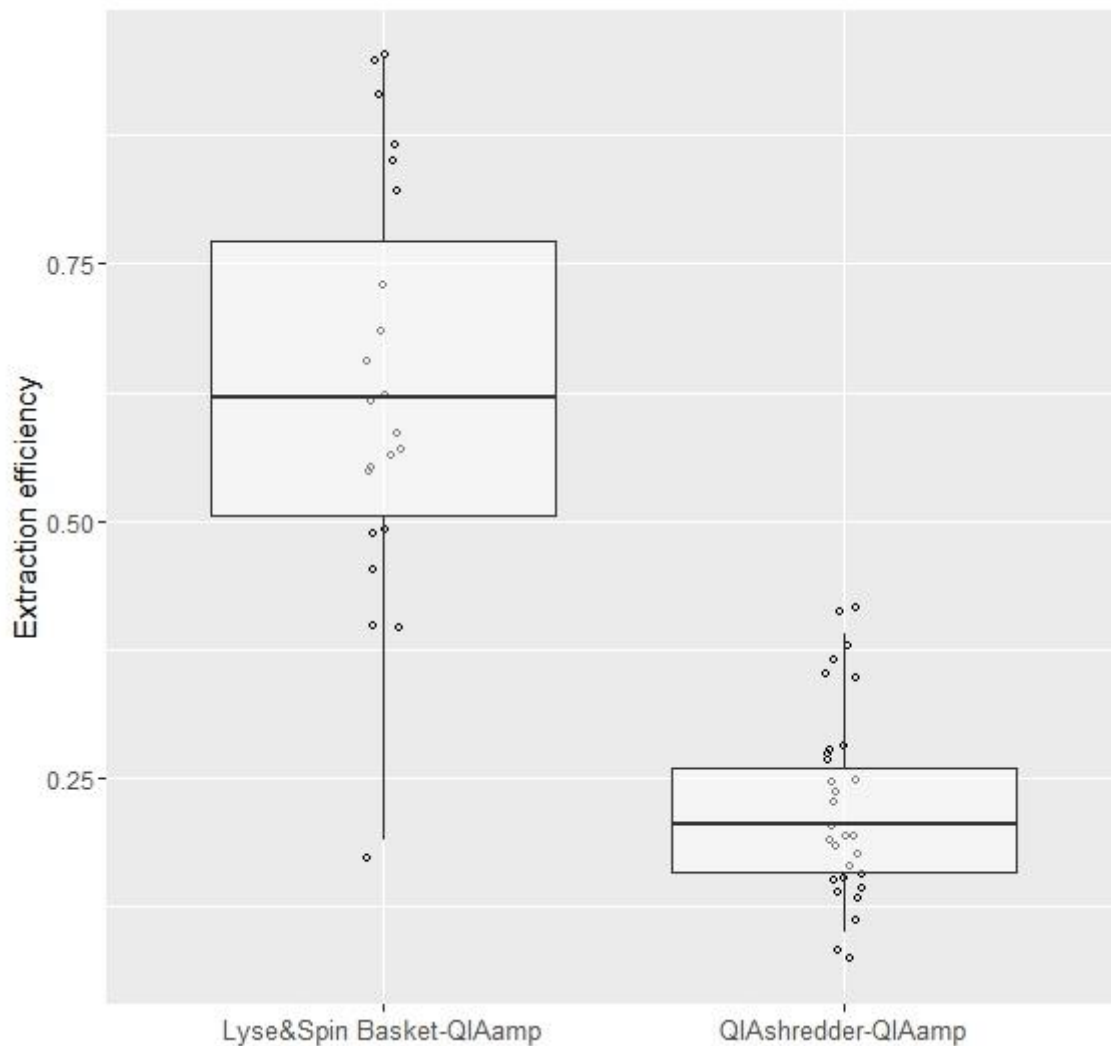


Figure 4: Extraction efficiency of the Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit (Lyse Lyse&Spin Basket-QIAamp) and QIAshredder-QIAamp DNA Mini kit (QIAshredder-QIAamp).

An average of 63% and 23% of the DNA is recovered respectively with Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit and QIAshredder-QIAamp DNA Mini kit (Table 2). We can observe that the Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit is more efficient. Further, it shows the importance of considering the extraction kit used when

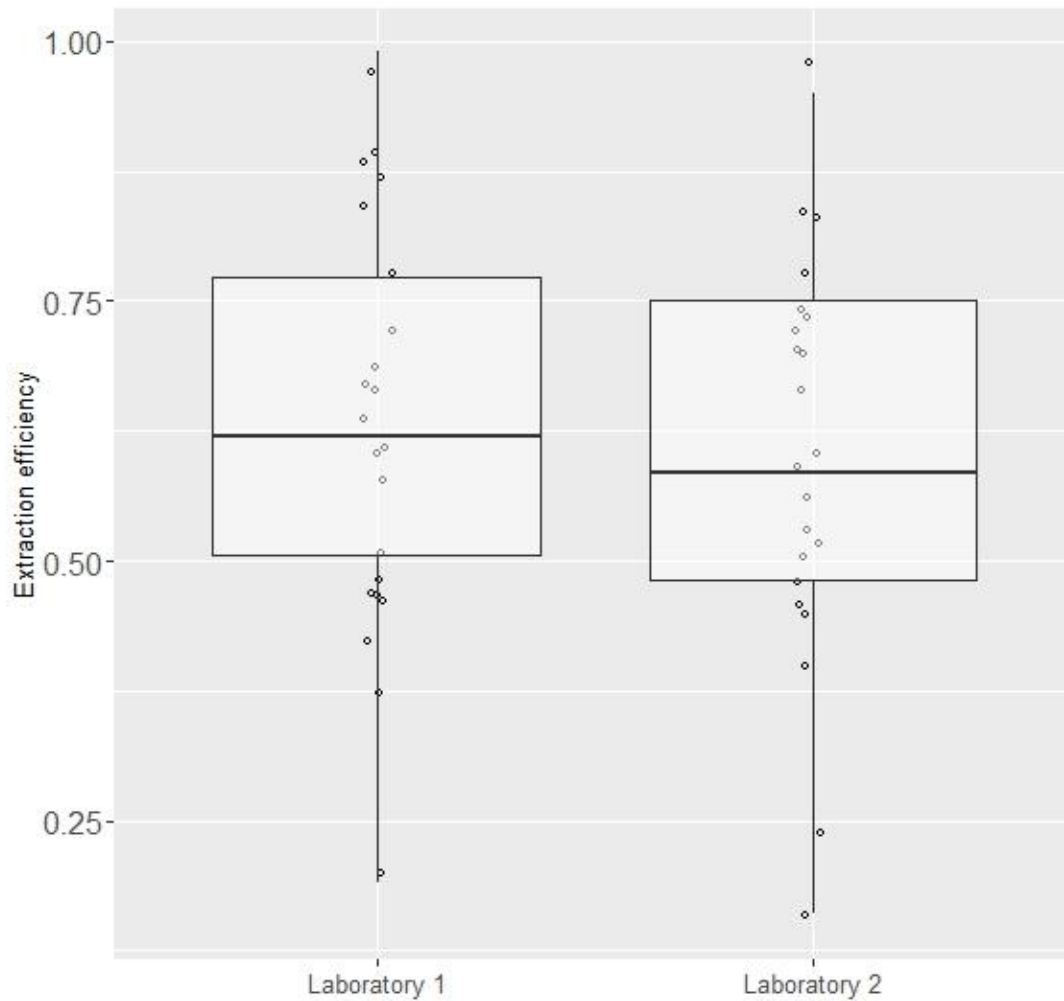
assessing a given amount of recovered DNA in an attempt to infer the initial quantity of DNA available.

**Table 2: Summary statistics of the extraction efficiencies obtained using both kits following the analysis of 30 samples respectively.**

Extraction kit	Min	0.05 percentile	Median	Mean	0.95 percentile	Max
Lyse&Spin Basket- QIAamp DNA Mini kit	0.19	0.41	0.62	0.63	0.92	0.99
QIAshredder-QIAamp DNA Mini kit	0.10	0.11	0.20	0.23	0.39	0.43

*Impact of the laboratory*

Figure 5 shows the DNA extraction efficiency of the 22 samples using Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit performed by each of the two laboratories.



**Figure 5 : Extraction efficiency of the Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit performed by two laboratories.**

For the first laboratory, an average of 63% of the recovered DNA is observed. The efficiency is an average of 59% for the second laboratory (Table 3). The difference between the two means is not significant. The Bayes factor supports the hypothesis that there is no difference between the two means [10].

**Table 3: Summary statistics of the extraction efficiencies obtained using the Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit performed by each of the two Laboratory. Laboratory 1 carried out the analysis on 30 samples. Laboratory 2 worked on 22 samples.**

Laboratory	Min	0.05 percentile	Median	Mean	0.95 percentile	Max
Laboratory 1	0.19	0.41	0.62	0.63	0.92	0.99
Laboratory 2	0.16	0.22	0.59	0.59	0.83	0.95

Taken jointly, it means that, for the Lyse&Spin and QIAamp DNA mini Kit, about 61% of DNA was recovered with no difference between the yields obtained by two different laboratories.

*The efficiency of cells release from swabs*

The extraction kit used here is the QIAshredder-QIAamp DNA Mini kit for which the extraction efficiency has been reported in the section *Efficiency of the extraction kits*. We recall that for this kit, only about 23% of the initial quantity of DNA was recovered.

The efficiency results associated with the cell release and DNA extraction with the kit are shown in Figure 6, jointly with the results on the DNA extraction kit only. It represents 30 samples deposited on 30 FLOQSwab™ and subsequently extracted with the kit.

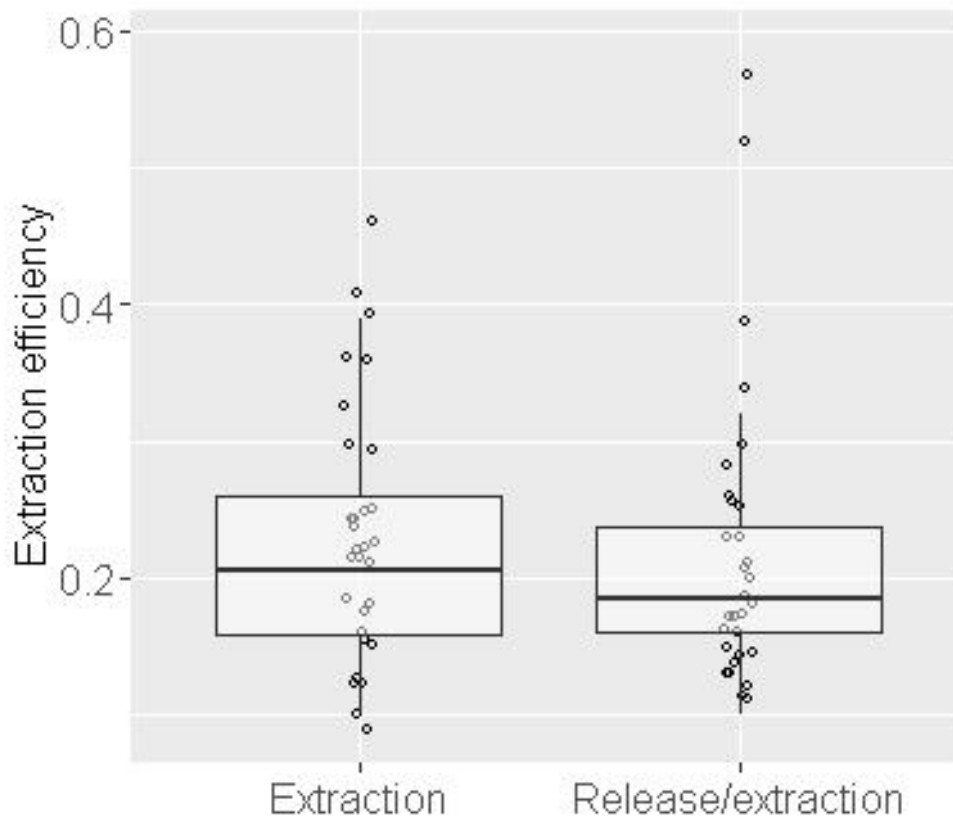


Figure 6 : Boxplot of the DNA extraction efficiency of QIAshredder-QIAamp DNA mini kit (left) with the boxplot of the efficiency associated with the cell release by the FLOQSwab™ and DNA extraction with the kit (right).

About 22% of the initial quantity of DNA is recovered after the deposition on the FLOQSwab™ and the extraction using the QIAshredder-QIAamp DNA Mini kit. The detailed data summary (Table 4) is below and compared the data obtained from the extraction kit alone.

Table 4: Summary statistics of the extraction efficiency of the kit alone and of the efficiency associated with the cell release by the FLOQSwab™ combined with the DNA extraction using the kit. In total 30 samples were analysed under both conditions.

Efficiency	Min	0.05 percentile	Median	Mean	0.95 percentile	Max
Extraction kit alone	0.10	0.11	0.20	0.23	0.39	0.43
Release/Extraction	0.10	0.11	0.18	0.22	0.46	0.59

The average efficiency to extract DNA is close to the efficiency to release cells and to extract DNA. It means that the cell release efficiency is close to 100%. How we estimate the cell release efficiency is presented next.

Knowing the mean and the standard deviation of both distributions representing the DNA extraction efficiency and the efficiency to release cells taking into account the DNA extraction efficiency of QIAshredder-QIAamp DNA Mini kit, the parameter “c” and “d” of the beta distribution  $Be(c, d)$  representing the efficiency of the swab release only can be calculated. A Beta distribution  $Be(32.26, 0.98)$  was obtained.

To obtain simulated data for the efficiency of the swab to release cells, 1000 values were randomly sampled from this Beta distribution  $Be(32.26, 0.98)$ . Each value is a theoretical result of the efficiency – between 0 and 100% – to release cells by the swab.

We can show that the FLOQSwab™ allows releasing about 97% of the cells on average. Summary statistics of the simulations are given below (Table 5 & Figure).

**Table 5: Summary statistics of the efficiency of the FLOQSwab™ to release cells, based on 1000 simulated values taken from a Beta(32.26, 0.98).**

Min	0.05 percentile	Median	Mean	0.95 percentile	Max
0.82	0.92	0.98	0.97	1	1

The distribution representing these 1000 random samples is given in Figure 7.

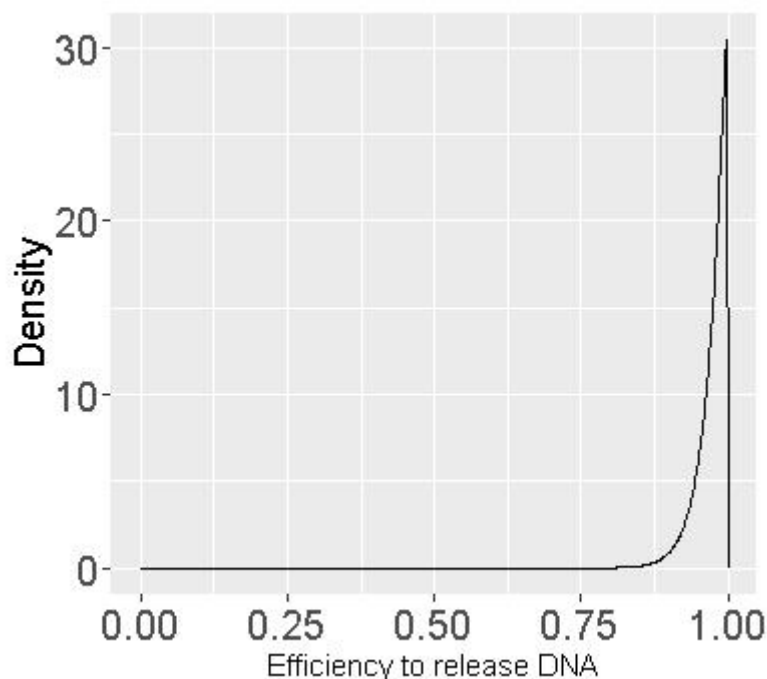


Figure 7 : Beta probability distribution of 1000 simulated values taken from a Beta(32.26, 0.98) representing the efficiency of the FLOQSwab™ to release cells.

## Discussion

This study had three objectives.

- To measure the extraction efficiency of two commercial DNA extraction kits (Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit, and QIAshredder-QIAamp DNA Mini kit from Qiagen),
- To study the impact of the laboratory on the yield offered by the best performing kit (Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit),
- To report on the efficiency of a swab (FLOQSwab™ from COPAN) to release cells and to show how to obtain it.

In the first part of the study, four DNA extractions were made using QIAshredder-QIAamp DNA Mini kit showing an average efficiency of 41% (Table 1) against 23% (Table 2) with the 30 samples. Further, a large variation (Figure 4 & Table 2) from 10% to 43% in the efficiency can be observed. These two observations show that a large number of experiments (greater than four) need to be done.



We report here a large difference of efficiency between both tested kits, despite the fact that the kits are quite similar regarding the laboratory protocols. The difference between the two kits is the use of Spin basket and no Microcon® 30 spin column for the Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit from Qiagen instead of the use of QIAshredder and microcon® 30 spin column for the QIAshredder-QIAamp DNA Mini kit. This observation can be a warning regarding the evaluation considering proposition at the activity level if specific data of the extraction kit should be used. In order to do this assumption, the impact of this different set of data on the result of evaluation should be studied. A lab can perform experiments on efficiencies with respect to its own method. If a lab is relying on data obtained using another kit, the impact on the result of the evaluation (on the likelihood ratio) of these other data, compared to the specific data of the laboratory, should be studied.

The large difference of efficiency between both tested kit could be explained by the different number of the DNA pipetting. QIAshredder-QIAamp DNA mini kit (QIAamp DNA Mini kit combined with QIAshredder and using the Microcon® 30 column) requires three DNA pipetting operations, including the pipetting into the microcon® 30 column, whereas the Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit need only one. At each pipetting of the total volume, a loss of DNA could occur with DNA being retained on the wall of the microtube or of the tips both made of polypropylene. Indeed, Gaillard [11] shows that adsorption of DNA to polypropylene tubes can occur. The large difference of efficiency between both tested kits could also be explained by the different number of spins used to retain DNA. Indeed, some DNA fragment could pass through the spin [12] instead of being retained. QIAshredder-QIAamp DNA Mini kit has more spins and microcon® 30 column than the other kit.

We have observed no significant difference between the DNA extraction efficiencies with the same kit used by two laboratories. This observation suggests that the effect of the laboratory is small compared to the variation due by the kit itself. However, given the limited number of laboratories involved (2), we ought to take this conclusion with the necessary caution.

We have also noticed that the maximum of the efficiency to release cells and to extract DNA is greater than the maximum of DNA extraction efficiency only. If the ratio of these two maximum values were done, an efficiency of swab to release cells greater than 1 would be obtained. However, this observation is possible, knowing that experiments are independent

---

and knowing the large variation between efficiencies. Therefore, taking the ratio of the two efficiencies values seems not ideal. All data allowing determining both extraction efficiency and efficiency to release cell and extract DNA should be used to estimate the efficiency of swab to release cells, as shown in Part 2 (Methodology- *Calculating efficiency*).

We have shown a large variation in efficiencies for a same kit in the same operator. This could be explained by the kit itself, but also by the flow cytometry. We suggest that the error introduced by flow cytometry is negligible. The calibration and quality controls performed on the instrument have shown that a variation on the cell number between 5 and 10% can occur, depending of the cell type and the cell concentration. It means that with a target number cells of 100, 90 to 110 cells will be selected. Therefore, the initial quantity of DNA may be slightly estimated. This effect is considered negligible compared to the ratio between initial quantity of DNA and final quantity of DNA. Because of this large variation, a distribution of efficiency values (and not a single point estimate such as the mean) should be taken into account when evaluating cases considering propositions at the activity level.

This study shows how flow cytometry can be a very effective tool to conduct DNA extraction and cell release efficiency research.

In Wood *et al.* [4], an extraction efficiency around 81% was reported, using QIAamp® DNA Investigator Kit (QIAGEN). This is higher than those reported in this paper: 23% and 63%, using respectively, QIAshredder-QIAamp DNA Mini kit and Investigator® Lyse&Spin Basket-QIAamp DNA Mini kit. However, when using QIAamp® DNA Investigator Kit (QIAGEN), EtOH is added in the first step of extraction protocol. This step may increase the recovery of DNA. Besides, the direct comparison between them has its limits. Indeed in Wood *et al.* [4], acellular DNA was used whereas keratinocyte cells were used in this study. DNA traces, obtained when touching a surface may be the results of a mix between acellular DNA, and cells [13]. Therefore, the extraction efficiency obtained in Wood *et al.* [4] or in this study may underestimate the extraction efficiency for DNA traces, obtained when touching a surface. Indeed, Propidium Iodide staining was used to sort the nucleated, living, keratinocytes cells. In that case, only porous cells are selected.

Wood *et al.* [4] obtained a lower efficiency of DNA release for nylon-flocked swabs (COPAN's FLOQSwabs™) that could also be due to the use of acellular DNA instead of

cells. Free DNA and cell membranes could interact differently with the microfibers of the swab.

Regarding the ability of the swab to release cells, unfortunately, a fixed number of cells cannot be directly deposited on the swab. A volume of the cell suspension containing a known concentration of cells is pipetted onto the swab. A loss of cells and DNA could occur via the pipetting, but the adsorption of cells and DNA to polypropylene tubes is limited by taking a partial volume of 35  $\mu$ L of a total volume mixed by vortexing. The efficiency of the swab to release cells could be underestimated. In addition, the chosen initial number of cells allowed obtaining quantity of DNA larger than the one obtained for touch DNA traces. In that case, the efficiency to release cell could be overestimated.

The nylon-flocked swabs (COPAN's FLOQSwabs™) have a higher efficiency to release cells than the two cotton swabs, Dryswab™ and Applimed SA [14]. However, samples of diluted blood were used in Rocque et al. [14] instead of a fixed number of keratinocytes.

To obtain the final quantity of DNA, a quantification needs to be performed. To perform this quantification, a loss of DNA could occur. However, the loss due to the use of a different quantification kit is supposed to be negligible (limited number of pipetting). Regarding the quantification, the quantity of DNA depends on the kit of quantification and the instrument of quantification. For consistency in this study, a single operator performed the quantification using the same kit and the same instrument in order to focus only on the impact of the laboratory on the extraction efficiency.

## **Conclusion**

Knowledge of the extraction efficiency of the kit used by the laboratory has a bearing on the assessment of the expected quantities of DNA that could be the result of different types of activities. It will impact the evaluation of the DNA results considering propositions at the activity level, especially when the case involves a low level of DNA. We developed a method to measure the efficiency of DNA extractions kits and the release efficiency of DNA swabs can be measured using flow cytometry. Flow cytometry allows obtaining a fixed number of cells. Therefore, the initial quantity of DNA, before performing an extraction, is known and controlled. It proves to be a very efficient technique to obtain adequate estimates of DNA extraction kit efficiency.

We measured the extraction efficiency of two commercial DNA extraction kits, Investigator® Lyse&Spin Basket-QIAamp DNA Mini Kit, and QIAshredder-QIAamp DNA Mini Kit used to extract and purify low quantities of DNA.

Results have shown that for the Lyse&Spin and QIAshredder-QIAamp DNA Mini Kit, about 61% of DNA is recovered with no difference between the extracts obtained by two different laboratories. For the QIAshredder-QIAamp DNA Mini Kit, only about 23% of the initial quantity of DNA is recovered.

Furthermore, we measured the efficiency of a swab, the FLOQSwab™ from COPAN, to release cells and have shown that the FLOQSwab™ releases about 97% of the cells.

### **Acknowledgements**

We wish to thank Patrick Basset for his help and the two persons for their participation to the laboratory part of this study, Magali Stoudmann and Delphine Nivard. We would like to express our gratitude to Henry Swofford for the insightful comments.

## Bibliography

- [1] S. Willis et al., ENFSI guideline for evaluative reporting in forensic science, European Network of Forensic Science Institutes, Dublin (2015). <http://enfsi.eu/documents/forensic-guidelines/> (accessed April 28, 2018).
- [2] D. Taylor, A. Biedermann, L. Samie, K.-M. Pun, T. Hicks, C. Champod, Helping to distinguish primary from secondary transfer events for trace DNA, *Forensic Sci. Int. Genet.* 28 (2017) 155–177.
- [3] R.J. Brownlow, K.E. Dagnall, C.E. Ames, A comparison of DNA collection and retrieval from two swab types (cotton and nylon flocked swab) when processed using three QIAGEN extraction methods, *J. Forensic Sci.* 57 (2012) 713–717.
- [4] I. Wood, S. Park, J. Tooke, O. Smith, R.M. Morgan, G.E. Meakin, Efficiencies of recovery and extraction of trace DNA from non-porous surfaces, *Forensic Sci. Int. Genet. Suppl. Ser.* 6 (2017) 153–155.
- [5] E. Butts, Exploring DNA extraction efficiency, Forensics@NIST 2012 Meeting, Gaithersburg, 2012 [https://www.nist.gov/sites/default/files/documents/oles/3\\_Butts-DNA-extraction-2.pdf](https://www.nist.gov/sites/default/files/documents/oles/3_Butts-DNA-extraction-2.pdf) (accessed May 14, 2018).
- [6] L. Samie, T. Hicks, V. Castella, F. Taroni, Stabbing simulations and DNA transfer, *Forensic Sci. Int. Genet.* 22 (2016) 73–80.
- [7] J. Russo, F. Sheriff, T.J. Pogash, T. Nguyen, J. Santucci-Pereira, I.H. Russo, *Techniques and Methodological Approaches in Breast Cancer Research*, Springer, 2014.
- [8] D. Dufresne, The Beta Product Distribution with Complex Parameters, *Commun. Stat. —Theory Methods.* 39:5 (2010) 837–854.
- [9] N. Johnson, S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions Vol. 2*, 2nd ed., Wiley, 1995.
- [10] F. Taroni; S. Bozza, A. Biedermann, P. Garbolino, C. Aitken, *Data analysis in forensic science: a Bayesian decision perspective*, John Wiley and Sons: *Statistics in Practice*, 2010.
- [11] C. Gaillard, Avoiding adsorption of DNA to polypropylene tubes and denaturation of short DNA fragments, *Tech. Tips Online.* 3 (1998) 63–65.
- [12] H. Tian, A.F.R. Hu, J.P. Landers, Evaluation of Silica Resins for Direct and Efficient Extraction of DNA from Complex Biological Matrices in a Miniaturized Format, *Anal. Biochem.* 283 (2000) 175–191.
- [13] S. Zoppis, B. Muciaccia, A. D'Alessio, E. Ziparo, C. Vecchiotti, A. Filippini, DNA fingerprinting secondary transfer from different skin areas: Morphological and genetic studies, *Forensic Sci. Int. Genet.* 11 (2014) 137–143.
- [14] M Rocque, S Leake, MP Milon, V Castella, The tightness of the cotton swabs meshing influences the chances of getting conclusive DNA profiles, *J. Forensic Res.* (2014) 5-234.

## Annexe II: Articles soumis à publication

II.1 Samie et al., The use of Bayesian Networks and simulation methods to identify the variables impacting the LR in stabbing cases evaluated given activity level propositions (2019) (Soumis pour publication)

Lydie Samie<sup>1,2</sup>, Christophe Champod<sup>1</sup>, Duncan Taylor<sup>3,4</sup>, Franco Taroni<sup>1</sup>

1. School of Criminal Justice, Faculty of Law, Criminal Justice and Public Administration, University of Lausanne, Lausanne-Dorigny, Switzerland
2. Forensic Genetics Unit, University Center of Legal Medicine Lausanne and Geneva, Lausanne, Switzerland.
3. Forensic Science SA, PO Box 2790, Adelaide, SA, 5000, Australia
4. College of Science and Engineering, Flinders University, Adelaide, Australia

### **Introduction**

Technical analytical developments have made it possible to analyse very low amounts of DNA. The drawback of this progress is that evaluation of the results has become increasingly complex when activity level propositions are taken into account. Indeed, low quantity DNA traces can be the result of a secondary transfer or even a tertiary transfer. Variables related to transfer, persistence and recovery are becoming increasingly important considerations in evidence evaluation, as the relevant question arising in Court is no longer “who is the source of this DNA?” but rather “how did this DNA get there?” [1-2]. These considerations were at the basis of the ENFSI guideline for evidence reporting [3] and the advice given to DNA reporting officers to systematically report trace DNA evidence considering activity level propositions adopting a likelihood ratio (LR) approach. This article is focused on how to interpret traces with small amounts of DNA, typically found following the touching of an object. This is not a simple task as it requires the probabilistic consideration of multiples variables such as transfer, persistence, recovery and background level of DNA. Many experts face practical difficulties when assessing these variables due to the perception that every case has a unique set of circumstances and numerical assignments are critically dependent on the specificities of the case. Biedermann *et al.* [4] explained further why forensic scientists may struggle with these assessments. A common argument put forward is: “because each case has its own feature, the use of numerical values from experimental studies performed under

controlled (laboratory) conditions cannot be used for evaluation in real life case". Besides, experiments can only cover a limited number of options for any particular variable, so it is difficult to envisage experiments taking into account all possible variations. However, although each case may have different data for these variables, this does not mean that the LR would be affected by all possible variations of these variables. Identifying the variables that impact on the LR will help forensic scientists to focus on a limited number of variables of interest in order to limit time and cost of the required data acquisition. The key task that will be explored in this paper is to identify the variables that have a significant impact on the weight to be assigned to the DNA findings.

The objective of this paper is to present a methodology to support forensic scientists in the evaluation of their results given activity level propositions. This study will show how to identify the variables impacting on the LR taking advantage of simulation methods. Such a strategy will allow scientists to focus the attention and efforts on the variables of interest.

Taylor *et al.* [5] built a Bayesian Network (BN) [6] allowing the transfer mechanisms to be considered. They offer a way to compute the LR associated with the DNA findings considering activity level propositions. However, in order to be used in casework, the BN needs to be informed with adequate data. The number of variables in the BN from [5] is large and a case-specific data acquisition to inform the parameters on all of them would be out of reach. This paper provides a method on how to identify the key variables that truly impact on the LR by performing simulations based on the BN. It is an extension of the simulation approach adopted by Taylor *et al.* [7] on body fluids attribution. Because the BN construction in [5] is the basis of the present contribution, the reader is advised to refer to it as we will not fully describe here its construction. In this paper, we will limit the description of the BN to the few modifications that we have introduced, the BN parametrisation and the simulations techniques that we will use.

To illustrate this method, the scenario of a stabbing attack with a knife is used. It can be summarized as follows: A victim is found dead in his flat following a stabbing incident carried out by an offender using a knife. The exact day of the stabbing is uncertain. At the crime scene, a knife is recovered and believed to be the attacker's weapon. The knife had been properly secured. It is believed by the investigators, in line with pathologist's assessment, that the stabbing occurred about 2 days ( $\pm 0.5$  day) before the discovery of the crime scene.

Based on the elements of the investigation, not related to DNA evidence, a person of interest (POI) is arrested and suspected to be the offender who stabbed the victim. Two days after the reporting of the incident, a DNA swab is taken from the unstained smooth plastic handle of the knife with a view to detect potential trace DNA left by the offender.

The prosecution's proposition (denoted  $H_p$ ) is that the POI was the person who used the knife to stab the victim.

In this paper, to gain some generalisation, two options are studied to reflect upon the defence point of view (denoted  $H_{d1}$  and  $H_{d2}$  respectively):

- In the first defence proposition,  $H_{d1}$ , the possibility of a secondary transfer is explicitly considered. The POI claims that, he shook hands with an unknown person, probably few hours before the stabbing, and it is that unknown person who is the real offender (who will be called alternative offender AO).
- In the second defence proposition,  $H_{d2}$ , it is alleged that the POI didn't stab the victim, but someone else unrelated to him did it (AO). In addition, the POI denies any prior encounter with either the victim, the knife or the AO. This represents a situation in which the POI is claiming no direct link with the offence.

Exploring these two defence's propositions will allow to show that, depending on the propositions of interest, the variables that have the strongest impact on the LR might be different.

The above-described circumstances provide the elements of what we will call the "initial case". Again, with a view to explore further that this specific set of circumstances, we will develop two additional cases. The first will adopt circumstances favouring the transfer of the POI's DNA (a "high transfer case"). The second adopts circumstances that are less favourable to the transfer of POI's DNA (an "low transfer case"). These sets of circumstances are further described in section 2.4.



## **Methodology**

The methodology adopted for this research is decomposed into several stages. First, we constructed a Bayesian network (BN) allowing a scientist to assess DNA findings associated with the stabbing attack scenario (including the possibility to change the defence proposition). The construction is mainly based on Taylor *et al.* [5] but has been adapted in order to carry out simulations. Then the conditional probability tables (CPTs) for each node of the network have been informed based on the data available from the literature. The parameters have been modelled in a way to (a) allow a Bayesian update in the light of new data and (b) to reflect when applicable the fact that the amount of data may be sparse or limited.

As such the constructed BN can be used to compute a likelihood ratio for a given case as done in [5]. However, we would like to go further by exploring impact and limitations of data on the LR values. A similar approach was adopted by [7]. This will be done using simulations resampling from the underpinning distributions used to inform the CPTs. The sampling will be carefully chosen by exploring one variable after the other. That is done in order to identify which variable (or node in the BN) has the most significant impact on the variations observed on the LRs. The isolation of the variables that have the most bearing on the variation of the LR from one simulation to the other is important to inform a data acquisition strategy. As it will transpire later, the developed BN has multiple variables and forensic laboratories will certainly not be able to systematically investigate the underpinning data for all of them. We will show that this simulation methodology allows us to successfully identify the key variables that should be the focus of future data acquisition. Such methods have already been applied successfully in areas such as DNA and fibres [8, 9]. Here, it is proposed to adopt such techniques to establish a baseline to inform future data acquisition campaigns.

In the next section, we will firstly present the development of the Bayesian network on the basis of what we call the “initial case”. It will be used as the starting point from which all simulations will be carried out. Then, the method used to perform the simulations, from this “initial case”, will be presented.

### *The Bayesian network, the underpinning data to inform the CPTs and variable instantiations for the “initial” case*

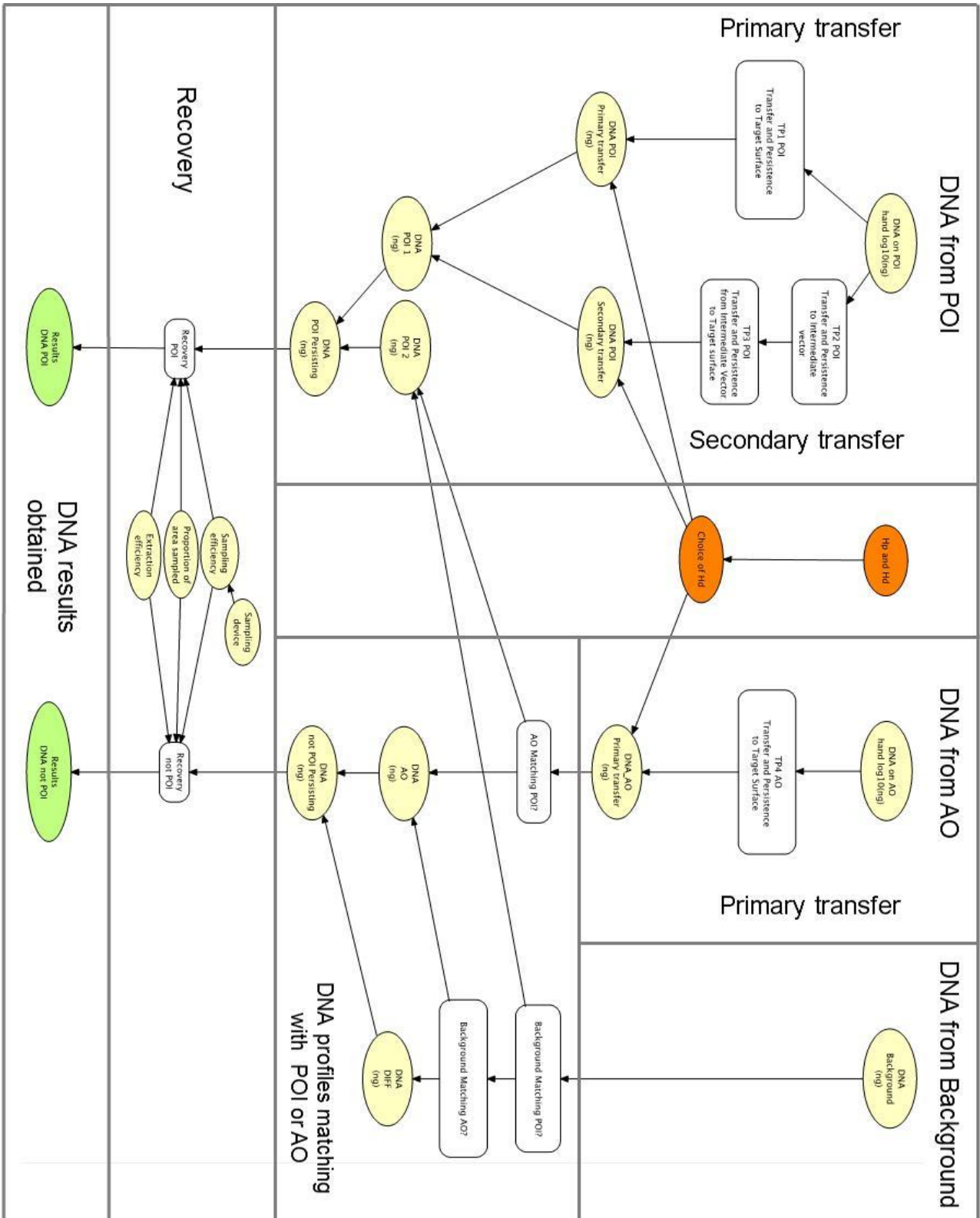
In this section, we present the BN, its variables, their states and the data used to inform the CPTs. In addition, we will use this BN with specific variable instantiations representing the circumstances of a case. We have called it the “initial case”. Initially, it is given this set of

---

circumstances that we will explore how the variables of the BN impact on the LR values. The other two cases (high transfer and low transfer) will be dealt with in a second step of the study.

The Object-Oriented Bayesian Network (OOBN), illustrated in Figure 8 of Taylor *et al.* [5] is reproduced here in Figure 1 and will be used in our study. It was adapted to be easily used in simulations.

Figure 1: Bayesian network used to evaluate the findings under activity level propositions involving primary vs secondary transfer events adapted from [5].



Each TP block contains variables involved in transfer and persistence event, more specifically:

- Block TP1 contains variables involved in the transfer and persistence of POI's DNA from POI's hand to the knife handle that occurs during the stabbing.
- Block TP2 contains variables involved in the transfer and persistence of POI's DNA from POI's hand to AO's hand that occurs during the handshake.
- Block TP3 contains variables involved in the transfer and persistence of POI's DNA from AO's hand to the knife handle that occurs during the stabbing.
- Block TP4 contains variables involved in the transfer and persistence of AO's DNA from AO's hand to the knife handle that occurs during the stabbing.

Variables included in each block called TP (for Transfer Persistence), M (for matching) and R (for recovery) are described in Table 1. The BN has been developed using Hugin Researcher (version 8.6, [www.hugin.com](http://www.hugin.com)).

Block	Variable
Block TP	Proportion of Contact
	Transfer proportion
	Nature of surface (for both target and shedding surface)
	Type of Contact
	Environment
	Days
	Decay factor of the loss of DNA (node named 'alpha2'), function of Environment.
Block M	Match Probability
Block R	Sampling efficiency
	The proportion of area sampled
	Extraction efficiency

Table 1: Variables taken into account in the blocks of the BN described in Figure.

Regarding the states of the variables, they can be instantiated depending on the case information available.

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

Table 2 details the states for each variable according to the case circumstances of the “initial case” and the type of sampling used by the laboratory.

<b>Variable</b>	<b>Possible states</b>	<b>Instantiated state for “initial case”</b>	<b>Explanation for the choice of the instantiated state</b>
Nature of target surface	Hand	Smooth	TP1-TP3-TP4: The surface of the knife handle is smooth.
	Smooth		
	Rough	Hand	TP2: It's a handshake
Nature of shedding surface	Hand	Hand	TP1-TP2-TP3-TP4: The hand is the primary source of DNA
	Smooth		
	Rough		
Vigour of contact	Passive	Friction	TP1-TP3-TP4: It is assumed that type of contact is friction when you stab a person.
	Pressure	Pressure	TP2: It is assumed that the type of contact is pressure when you shake hands.
	Friction		
Environment	Favourable	Favourable	TP1-TP3-TP4: The target surface is the knife. It is assumed that the knife being kept in a paper bag by the crime scene investigators shortly after it was seized. The environment (the paper bag) is considered favourable in the sense that DNA will be preserved.
		Poor	TP2: The intermediate surface is a hand. It is assumed that this surface can be considered as an unfavourable environment because of the high risk of contact with other surfaces, resulting in a loss of DNA.
	Poor	Poor	
Sampling device	Tapelift	Swab	A swab was used to take samples from the knife
	Swab		

Table 2: Variables with their associated states and the instantiated state corresponding to the circumstances of the “initial case”.

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

Table 3 presents the non-instantiated variables and the data that will be used to inform their CPTs.

Variable	Data informing the variable
Days	Data 1
Proportion of area sampled	Data 2
Transfer proportion	Data 3
Sampling efficiency	
Extraction efficiency	
DNA quantity on hands	
Background	

Table 3: Non-instantiated variables and corresponding data used to inform their CPTs.

The data that will be used to inform their CPTs are referred to as Data 1 to Data 3 are detailed hereinafter.

The outcomes of the DNA analysis are set in two variables that gives the amount of DNA in ng, respectively for the POI and for not POI (Table 4).

Variable	Possible states (ng)
Results DNA POI	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf
Results DNA not POI	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1 10 to 25 in steps of 5 25 to 1000 1000 to inf

Table 4 : Possible results for the amount of DNA in ng respectively for POI and not POI.

#### Representing our lack of knowledge in the BN

Within the nodes in the BN there is uncertainty as to what the state of nature was in any particular instance. This can be thought of as comprising two distinct parts, the uncertainty

surrounding the state that applies to a specific case, and the uncertainty surrounding the suitability of our data to model the world. The node 'DNA on hands' is a good example to further explain this idea.

In the stabbing scenario being considered there is uncertainty surrounding the amount of DNA on the hands of the POI (or AO) and we deal with this in the BN by treating the 'DNA on hands' node as a distribution that is meant to reflect the amount of DNA that the general population will possess on their hands. This distribution reflects our prior belief on the amount of DNA that the suspect (or AO) had on their hands at the time of the offence. Within the BN architecture the POI and AO have separate 'DNA on hands' nodes to reflect the fact that they are different people and can have different amounts of DNA on their hands (the posterior distribution of which would be obtained after instantiation of case information). In order to model the distribution of DNA on hands in the population, we must take a sample from the population and measure the amount of DNA on peoples' hands. The second component of our uncertainty is then how representative our sample is of the general population. It is only this second component of uncertainty that will benefit from additional research and sampling i.e. will potentially reduce the sensitivity of the LR to the data underlying these nodes. Consider a scenario where the results have very little power to distinguish between the propositions (and hence a LR close to one would be obtained). If this is due to the shape of the distributions used to model the population (and not how representative the sample is of the general population), then no amount of additional sampling will help to increase discrimination.

In our sampling schemes (which we list below) there are different aspects of case information and experimental data that we focus on during our simulations that represent both aspects of uncertainty.

#### Data 1: Days

To account for the uncertainty on the number of days between the stabbing and the crime scene attendance, and the number of days (or hours) between the stabbing and the alleged handshake, the variable "Days" is modelled by a Gamma distribution  $Ga(\alpha, \beta)$  that has been discretized on a range of possibilities (from 0 to 31 days). The choice for a Gamma distribution allowed to easily model events from 0 to infinity. The parameters  $\alpha$  and  $\beta$  are calculated using [10] based on a mean and a variance (set by the circumstances) (Table 5). The variance is set to 0.5 in order to account for an uncertainty of about two days maximum.

Note that in [5] we modelled the persistence of DNA (or the reduction of the amount of DNA over time) using an exponential decay curve set by a variable called 'alpha2'. The parameter of the decay curve depends on the state of the variable "Environment" (either 'poor' or 'favourable') and is based on [Raymond et al, 2009]. The parameter 'alpha2' is set to 0.022 if the environment is favourable or to 0.052 if the environment is poor.

Variable	Discretized states	Mean and variance	Explanation	Ga( $\alpha$ , $\beta$ ) modelling the variable
Days	interval node; 0, 0.5 then 1 to 31 in steps of 1	$\mu=2$ $\sigma^2=0.5$	TP1-TP3-TP4: The item was examined around two days after the offence.	Ga(8, 4)
		$\mu=0.5$ $\sigma^2=0.5$	TP2: It is assumed that the handshake was made less than 12 hours but more than 2 hours before the stabbing.	Ga(0.5, 1)

Table 5: Variable "Days" with its states and parameters (mean and variance and associated Gamma distributions) used to inform the CPTs.

### Data 2: Proportion of area sampled

This variable is modelled using a Beta distribution  $Be(\alpha, \beta)$ , whose parameters are estimated from a mean and variance for the proportion. It represents the proportion the touch surface that is sampled (using a swab or a tapelift). It allows to account for the fact that the whole (100%) of the touched surface may not be sampled. Note that it does account for the uptake efficiency of the swab or tapelift. The variability on the latter is accounted directly in the variable "Transfer proportion". The parameters  $\alpha$ ,  $\beta$  are computed from mean and variance according to [11]. Table 6 summarizes these parameters. As for "Days", this distribution is used here to reflect an uncertainty regarding the circumstances of the case. If we had no doubt regarding them, it would be unnecessary to carry out such simulations.

Variable	Discretized states	Mean and variance of data	Explanation	Be( $\alpha$ , $\beta$ ) based on mean and variance
Proportion of area sampled	interval node; values from 0 to 1 in steps of 0.1	$\mu=0.95$ $\sigma^2=0.001$	We have assumed that almost the entire knife handle is sampled, representing about 95% of its surface.	Be(44.175, 2.325)

Table 6: Variable "Proportion of area sampled" with the explained mean and variance associated to the data used to inform the CPTs, with the parameters of the beta distribution based on the mean and variance.

### Data 3: Transfer proportion, Sampling efficiency, Extraction efficiency, DNA quantity on hands and Background

We have adopted a full Bayesian strategy to inform the parameters associated with these variables. It means that we have initially set a prior probability distribution for the variables.



Then, based on data from the scientific literature, we have updated these distributions, leading to posterior distributions that will be used to inform the CPTs associated with them.

The prior distributions set for each variable are presented in Table 7.

For the variables “Transfer proportion”, “Sampling efficiency” and “Extraction efficiency”, a so-called flat prior  $Be(1,1)$  has been chosen to start from a uninformed situation between 0 and 1.

The variable “DNA quantity on hands” is transformed in  $\log_{10}$  (from  $-\infty$  to  $+\infty$ ) with a prior distribution based on a mean quantity of 2ng of DNA and a large variance (variance of 1). The mean quantity of 2ng is what appears to be a reasonable amount of DNA. A large variance was chosen to account for the paucity of data available at this stage. Note that the above method with these parameters led to only positive values.

The variable “Background” is modelled using a Gamma distribution whose parameters have been estimated [10] from a mean quantity of background DNA (0.5 ng) and an associated variance (variance of 5). The mean of 0.5 ng is viewed as a reasonable upper bound quantity. The large variance has been chosen to be large reflecting the paucity of data available. Ten prior observations have been drawn from that Gamma distribution to act as our prior data counts. We will then update these counts with the data obtained from the literature. 10 was chosen to reflect the paucity of the sample sizes available in the literature. Indeed, we couldn't claim more in the prior counts than what is actually published.

The data used to update the above prior distribution for the variable “Transfer proportion” are based on [5] where the authors estimated the parameters of the Beta distribution modelling “Transfer proportion” based on simulations from the original data from Daly [12], Bontadelli [13] and Goray [14]. To reflect the fact that a laboratory will conduct only a limited number of experiments to inform that variable, we randomly selected 51 data from the beta distribution obtained in [5] to act as the dataset that we will use to update our prior distribution. 51 is the number of experiments done by Daly [12].

For the variable for the “Sampling Efficiency”, we will distinguish the “swab” from the “Tapelift”. In [5], the parameters  $Be(25, 20)$  of the beta distribution representing the data for

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

“Swab” condition were based on [15]. As before, we will randomly draw a limited sample (21) from that distribution to carry out the Bayesian update of our prior distribution. Indeed, 21 experiments were done in [15]. For the “Tapelift” condition, we have used directly the data from [15].

For the variables “Extraction efficiency”, “DNA quantity on hands” and “Background”, we have used directly the data from the literature (Table 7).

Beta, normal and Gamma distributions were updated using standard Bayesian methods [16].

Variable	Discretized states	Prior distribution	Data Source	Number of data points	Posterior distribution
Transfer proportion	interval node; from 0 to 1 in steps of 0.05	Be(1,1)	51 random sample from the Beta distribution defined in [5]	51	Hand(rough)/passive: Be(1.89,3.83)  Hand(rough)/pressure: Be(1.99,2.78)  Hand(rough)/friction: Be(2.02, 2.98)  Smooth/passive: Be(1.55,32.65)  Smooth/pressure: Be(1.73,33.43)  Smooth/friction: Be(1.43, 2.10)
Sampling efficiency	interval node; from 0 to 1 in steps of 0.05	Be(1,1)	State “Tapelift”:[15]  State “Swab”: 21 random sample from beta distribution from [5]	21	Swab: Be(14.36,11.79)  Tapelift: Be(3.15,17.11)
Extraction	interval node;	Be(1,1)	[17]	15	Be(5.79,18.43)

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

efficiency	from 0 to 1 in steps of 0.05				
Log10(DNA quantity on hands)	interval node; -inf to -1.5 -1.5 to 3.5 in steps of 0.1 3.5 to inf	N(0.3,1)	[13]	50	N(0.764,0.004)
Background	interval node; 0 to 0.1 in steps of 0.01 0.1 to 1 in steps of 0.1 1 to 5 in steps of 0.5 5 to 10 in steps of 1	10 data randomly selected from the distribution Ga(0.08,0.16)	[18]	301	Ga(0.6, 30.15)

*Table 7: Variables, discretized states, prior distribution, data source to update them and posterior distribution used to inform their CPTs.*

*Methods used to perform simulations from the “initial case”*

In the previous section, we have presented the BN that captures the “initial case”. As such it can be used to compute a LR for any DNA outcome in terms of quantity of DNA corresponding to the POI and non-corresponding to the POI. For a given outcome, we will obtain one LR that encapsulates the knowledge that has been used to inform the CPTs.

Using the Bayesian update mechanism presented before (under Data 3), we have used specific data sets from published studies. The numbers of data points used in these studies are rather sparse and can be seen as small subsets from larger and unknown populations. Through simulations, we would like to show the impact (if any) on the LR of such limited samples. To do that, we have resampled with replacement the data points that have been used to carry out the Bayesian update. The simulation method is detailed later in this section. At each simulation then, the above-described posterior distributions are recomputed, the BN CPTs are updated and LRs obtained for any DNA outcome. We carried out that task 100 times. Hence for a given DNA outcome, we have now 100 LRs. These LRs will have a range that can be characterized [e.g. IQR, min to max]. Note that these 100 LRs represents 100 slightly different scenarios.

A first set of results out of these simulations will show these ranges (one per possible DNA outcome and for each of the defence propositions). They reflect how the LR's vary based on the state of knowledge and understanding for sets of data points used to inform the CPTs.

The second question we will address through simulation is to identify which variables impact the most on the observed ranges of LR. This identification of impacting variables will allow prioritisation of further data acquisitions. That approach stems from the realisation that, given the complexity of the problem, we cannot expect systematic acquisition of large datasets for all the variables identified in the BN. To make that selection, we will carry out resampling on a variable by variable basis (keeping all the other variables constant). For each set of simulations (100 simulations per variable), we will measure the ranges (for each DNA outcome and considering each of the defence propositions). These simulations on a variable per variable basis allow pinpointing of the variables that have the most impact on the LR's. These shall then constitute the focus for further data acquisition, because additional datasets have the potential to reduce the observed ranges of LR's and improve on their robustness.

The simulations were performed in Rstudio (Version 1.1.463) [19] with R (version 3.5) [20] combined with RHugin (version 8.4) [21]. The library RHugin specifically allows to liaise directly with the Hugin inference engine. It allows then to load the BN as in Hugin and to interact directly with it without resorting to the Hugin GUI. It means that all the captured dependencies between the variables in the BN are duly maintained and are part of the computation.

Table 8 presents the variables that will be used for the simulation exercise (either jointly or separately) with an indication of the method of simulation that will be applied. These three simulation methods are described below.

Variable	Simulation method
Days	1
Proportion of area sampled	
Transfer proportion	2
Sampling efficiency	
Extraction efficiency	
DNA quantity on hands	
Background	3
Match probability	

Table 8: Each variable associated with a method of simulation.

*Simulation method 1:* For each simulation, the mean for each variable (“Days” and “Proportion of area sampled”) is resampled from a Normal distribution with a mean set as per

the initial case but allowing a variance, respectively of 0.5 and 0.01 around it. For instance, the mean for the number of days in the TP1 block is 2. For each simulation then, the mean will be obtained by randomly selecting from a sample from a  $N(2,0.5)$ , until the mean is positive (For "Proportion of area sampled", the mean is resampled until a value between 0 and 1 is obtained). From that mean, and keeping the variance constant, we estimate, as before, the parameters of the Gamma distribution (for "Days"). The values for the variance were chosen because, in the authors' opinion, they adequately reflect the amount of uncertainty surrounding the timeframes.

*Simulation method 2:* For each of these variables that were subjected to the Bayesian update, the simulations are carried out by resampling (with replacement) from original data presented in the previous section.

*Simulation method 3:* The block M of the BN, presented in Figure 2, shows a dependence between the node "Match probability" and the node "Quantity of DNA" as in [5].

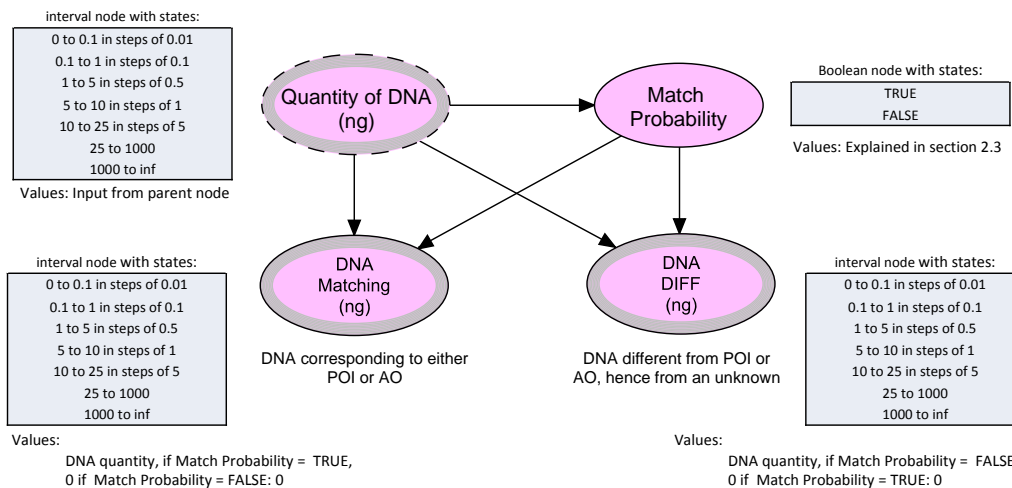


Figure 2: Block M with its four nodes and associated states and values.

In that block, only the node "Match Probability" requires input probabilities (the node "Quantity of DNA" being set by its parent). Its state is TRUE when the DNA profile is matching, FALSE otherwise. The probability of being TRUE is set by the match probability which in turn depends on the quantity of DNA. Linked to the quantity of DNA is the number alleles that the profile will show. To compute the match probability, we have accounted for that relationship between the quantity of DNA and the number of alleles.

Hence, we will first establish the relation between quantity of DNA and the number alleles and then propose a way to compute match probability as a function of the number of alleles.

The number of alleles corresponding to a given quantity of DNA is modelled based on the empirical observations [22] made between the quantity of DNA and the minimum and maximum numbers of detected alleles (Table 9). A Gamma distribution is used to represent the numbers of detected alleles as a flexible modelling distribution for counts between 0 and infinity. The parameters of the Gamma distribution are informed on the mean and variance obtained from [22]. These Gamma distributions will be used at each simulation to generate a given number of alleles corresponding to a given quantity of DNA (sampled from its own distribution).

Quantity (ng)	Minimum to maximum numbers of detected alleles	Mean and variance	Ga( $\kappa$ , $\theta$ ) modelling the Number of alleles
0-0.01	0 to 1	$\mu=1$ ; $\sigma^2=0.1$	Ga(10, 10)
0.01-0.02	0 to 2	$\mu=1$ ; $\sigma^2=0.4$	Ga(2.5, 2.5)
0.02-0.03	1 to 3	$\mu=2$ ; $\sigma^2=0.4$	Ga(10, 5)
0.03-0.04	2 to 4	$\mu=3$ ; $\sigma^2=0.4$	Ga(22.5, 7.5)
0.04-0.05	3 to 5	$\mu=4$ ; $\sigma^2=0.4$	Ga(40, 10)
0.05-0.06	4 to 6	$\mu=5$ ; $\sigma^2=0.4$	Ga(62.5, 12.5)
0.06-0.08	5 to 7	$\mu=6$ ; $\sigma^2=0.4$	Ga(90, 15)
0.08-0.1	6 to 8	$\mu=7$ ; $\sigma^2=0.4$	Ga(122, 17.5)
0.1-0.2	9 to 22	$\mu=15$ ; $\sigma^2=1.5$	Ga(150, 10)
0.2-0.3	20 to 30	$\mu=25$ ; $\sigma^2=1.4$	Ga(446, 17.9)
0.3-0.4	29 to 32	$\mu=31$ ; $\sigma^2=0.5$	Ga(1922, 62)
>0.4	32	the number of allele is 32	the number of allele is 32

Table 9: Range of quantities of DNA (ng) with the associated minimum and maximum numbers of alleles observed empirically. For each number of alleles (min-max), a mean and variance is set to reflect these ranges and the Gamma distributions parameters are obtained.

To obtain the match probability for a given number of alleles (associated with a given quantity of DNA), we proceeded as follows:

- (1) 5 million full DNA profiles (32 alleles in total) are randomly generated based on allelic frequencies of the NGMSelect kit [23] for the Swiss Caucasian population.
- (2) From the above profiles, 27046 partial DNA profiles are created using the allelic degradation model from Hicks et al [24]. It means that for each number of alleles (1 to 31), we have a collection of partial DNA profiles (from 6 to 1000) depending on the number of alleles.

- 
- (3) For each of these profiles, their match probability (MP) is computed with a  $\theta$  of 0.02 using the allele frequencies from [23].
- (4) For a given number of alleles, at each simulation, we draw the match probability from the collection of match probabilities associated with the drawn number of alleles (corresponding to a given quantity of DNA).

During the simulation process, the number of alleles and the associated MPs are resampled only if this node is taken into consideration. Otherwise, its CPTs is set once for all simulations.

*Method used to perform simulations beyond the “initial case”*

The “initial case” represents the circumstances of the case, typically set by the chosen states in the nodes that have been instantiated. The BN also allows us to explore other sets of circumstances. Hence, we can repeat the simulation process to explore how each node impacts on the LRs under any set of circumstances. We have chosen to investigate two additional scenarios deviating from the conditions of the “initial case”. The first will adopt circumstances that will favour the transfer of the POI DNA (a rough surface to increase the deposition, a short time delay between the stabbing and the collection of the swabs, and favourable environmental conditions). The second adopts circumstances that are less favourable to the transfer of DNA. The states of the nodes for each scenario are presented in Table 10. The other nodes related to the case circumstances were kept constant with the same states as for the “initial case”.

<b>Node</b>	<b>“Initial case” Instantiated state or mean</b>	<b>“High transfer case” Instantiated state or mean</b>	<b>“Low transfer case” Instantiated state or mean</b>
Nature of target surface	TP1-TP3-TP4: Smooth TP2: Hand	TP1-TP3-TP4: Rough TP2: Hand	TP1-TP3-TP4: Smooth TP2: Hand
Days between both transfers or transfer and recovery	TP1-TP3-TP4: 2 TP2: 0.5	TP1-TP3-TP4: 0.5 TP2: 0.5	TP1-TP3-TP4: 20 TP2: 0.5
Environmental conditions	TP1-TP3-TP4: Favourable TP2: Unfavourable	TP1-TP3-TP4: Favourable TP2: Unfavourable	TP1-TP3-TP4: Unfavourable TP2: Unfavourable

*Table 10: Choice for node instantiations and mean for the “initial case”, for the “high transfer case” and for the “low transfer case”.*

*Method used for the analysis of the simulation results*

Following a set of simulations (one for each of the initial, the high transfer and the low transfer case), we obtain 100 LRs for each combination of results (quantity of POI and not POI DNA – 36 x 36 possibilities), each proposition retained for the defence (Hd<sub>1</sub> and Hd<sub>2</sub>) and for each node considered (10 nodes and the case with all simulated nodes considered jointly). It represents a total of 28'512 combinations and 2'851'200 LRs.

A dedicated Shiny application ([https://lydie-samie.shinyapps.io/DNA\\_Activity/](https://lydie-samie.shinyapps.io/DNA_Activity/)) has been designed to allow the visualisation of these results for each possible combinations of variables.

To explore which node (or variable) has impact on the LRs, we ordered them by range. That will be done by conditioning on the defence proposition. It is important to stress that with this approach, we aim at identifying the variables that have the most impact across all possible outcomes. The variable by variable analysis allows us to identify which variables contribute significantly to the whole. By significant, we mean that the range of LRs (meaning the difference between the maximum LR and the minimum LR) produced by sampling a variable exceeds 1 order of magnitude of LR. The range was chosen to cover up to extreme situations even if their probability of occurrence is low. At this stage, the issue to screen among variables with a rather wide net, instead of being too strict in their selection by adopting a more stringent criterion such as the interquartile range.

*Performing simulations adapting the number of data points informing the conditional probability tables*

As proposed in [7] we will mimic an increase of knowledge, to do so we have increased the counts informing the CPTs by a constant factor (retaining the observed proportions).

The variables (“Background”, “DNA quantity on hands”, “Sampling efficiency”, “Extraction efficiency” and “Transfer Proportion”) that proved to be significant based on the method described in section 2.4, have been studied using simulations. To simulate an increase in the size of datasets, we have adapted the simulation process (*Simulation method 2*) by tripling the number of data points resampled and then used to inform the corresponding CPTs. The original numbers of datapoints used for each variable are given in Table 7. It means that the relative proportions associated with each state of each variable remain the same, but the counts are multiplied by 3 as if the study had been conducted on a larger sample. The factor of 3 is what we considered reasonable for an operational laboratory.



## **Results and Discussion**

### *Ranges of simulated LRs obtained under $Hd_1$*

$Hd_1$  stipulates that the defence alleges that the DNA corresponding to the POI's profile is the consequence of a secondary transfer. The DNA corresponding to POI's profile is the contribution of the POI following the secondary transfer and the DNA different from POI's profile is to the potential joint contribution of the background and the DNA of the alternative offender (AO). Figure 3 illustrates the ranges of LRs for all outcomes (i.e. all considered amounts of POI and not POI's DNA, each possibility gives a range of LRs after 100 simulations) considering respectively all variables jointly and then each variable simulated in turn. The results are shown for the three cases considered: "initial", "high transfer" and "low transfer"). Regardless of the case, the significant variables (with an impact of more than an order of magnitude) are: "DNA quantity on hands", "Extraction efficiency", "Background", and "Sampling efficiency". Two of these variables "Extraction efficiency" and "Sampling efficiency" relates to the laboratory choices regarding their sampling devices and extraction techniques.

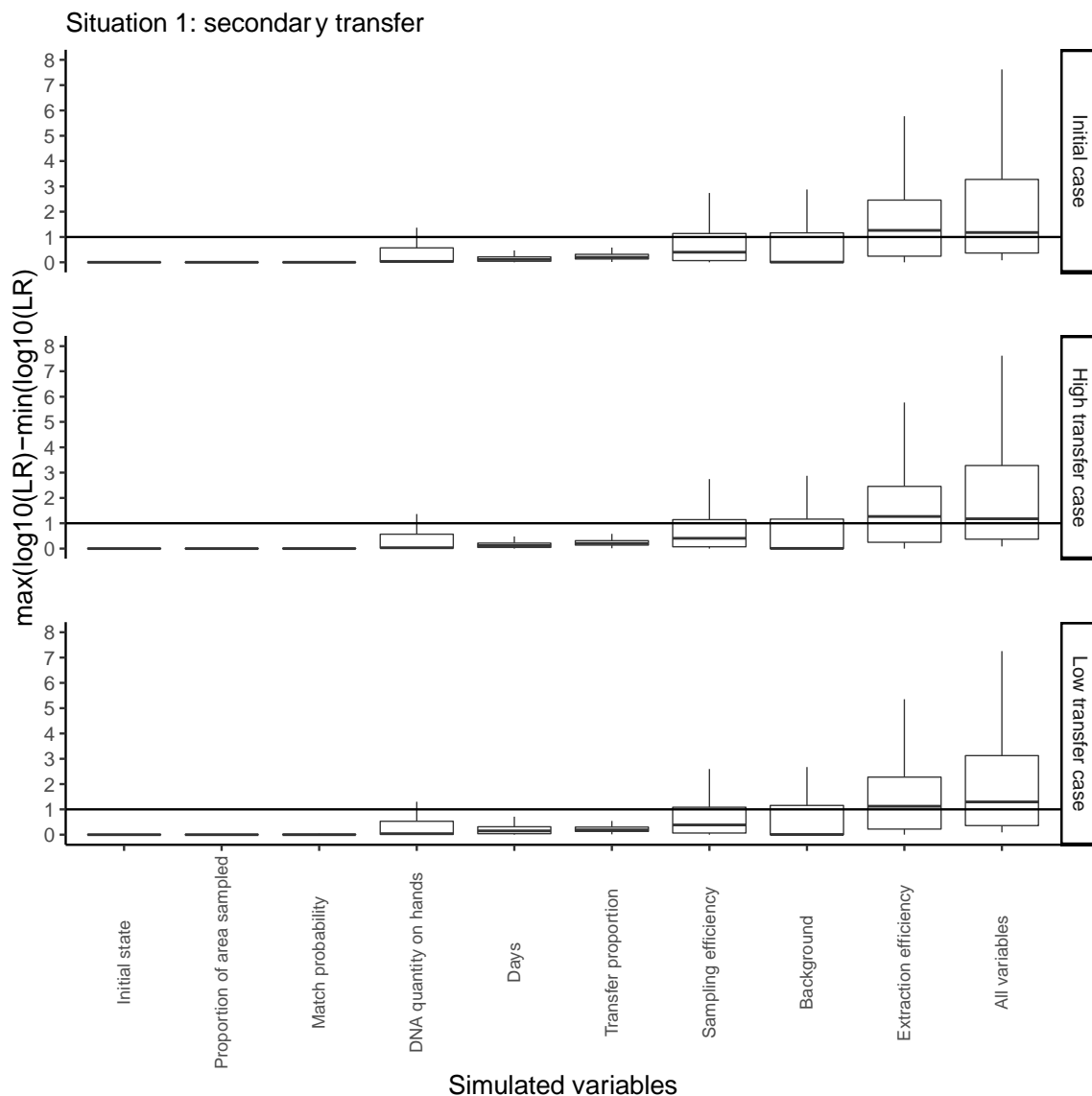


Figure 3: Boxplots presenting the ranges (min-max) expressed in  $\log_{10}$  of the LRs obtained following 100 simulations under  $H_{d1}$ . The panels present the results for the three cases considered. The horizontal line drawn at the difference of 1 (in  $\log_{10}$ ) set the limit above the variable considered will be declared as having a significant effect on the global variability shown when “all variables” are resampled jointly.

Hence, when secondary transfer is alleged, the current state of knowledge regarding the transfer and persistence of DNA is sparse and induces a large variability on the simulated LRs. One way to overcome this problem would be to increase the underpinning data (see section 3.4 below). It is rather easy for a laboratory to increase its knowledge base associated with some of the variables involved (see [25] for example regarding the acquisition of data in relation to sampling and extraction efficiency).

#### *Ranges of simulated LRs obtained under situation 2 ( $H_{d2}$ )*

In situation 2 the defence alleges that the DNA corresponding to POI's profile is due to the contribution of an unknown alternative offender (AO) ( $H_{d2}$ ). In this situation, the possibility

for a secondary transfer for the POI is not retained. Figure 4 illustrates the ranges of LR<sub>s</sub> plotted considering the same variables as before. The variable by variable analysis allows to identify the variables contributing to more than one order of magnitude. They are the following: “DNA quantity on hands”, “Match probability”, “Extraction efficiency”, “Proportion of transfer”, “Sampling efficiency” and “Background”, regardless of the scenario considered.

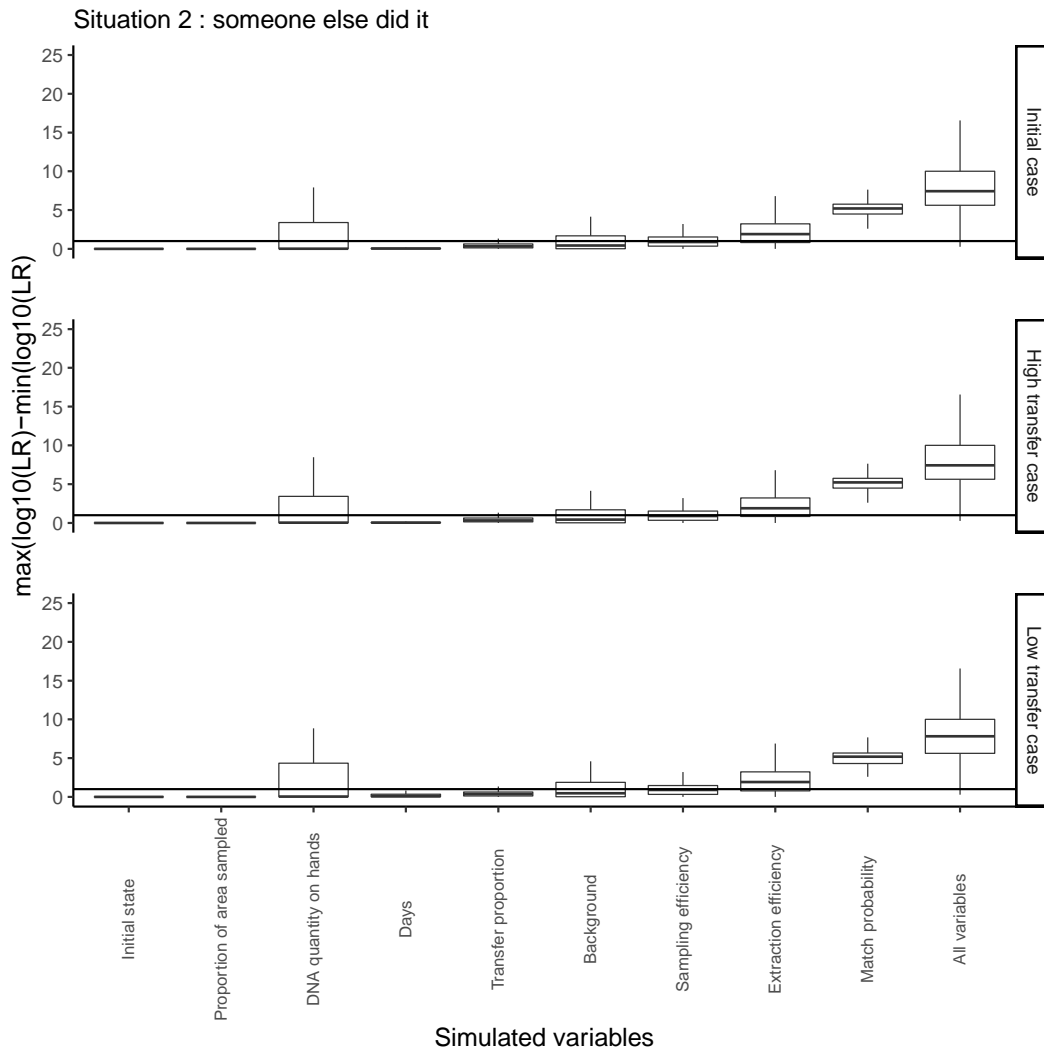


Figure 4: Boxplots presenting the ranges (between min-max) expressed in log<sub>10</sub> of the LR<sub>s</sub> obtained following 100 simulations under Hd<sub>2</sub>. The panels present the results for the three cases considered. The horizontal line drawn at the difference of 1 (in log<sub>10</sub>) set the limit above the variable considered will be declared as having a significant effect on the global variability shown when “all variables” are resampled jointly.

In addition to the variables identified for Hd<sub>1</sub>, the variables “Match probability” and “Transfer proportion” comes into play under Hd<sub>2</sub>. This is due to the fact that under that defence proposition, the DNA corresponding potentially to the POI may arise from the background level of DNA on the surface. That was much less critical under Hd<sub>1</sub> as the secondary transfer was dominating the considerations compared to the background.

The above results are showing the most impacting variables in a global sense, regardless of the outcome observed in a given case (i.e. the respective amounts of DNA corresponding to POI and to the not POI). It helps us to identify the variables that ought to receive attention if we want to treat the problem globally considering all possible outcomes. Also the case (initial, high transfer and low transfer) considered as no impact on the impacting variables.

However, in a given case with observed outcomes, the variables that will have the most impact will vary and could be different from the above globally selected variables. To illustrate this point, we present one example (with 0.5-0.6 ng of POI's DNA with 0.03-0.04 ng of not POI's DNA) leading to a different selection of the most contributing variables (Figure 5). The reader can refer to the Shiny application to select any set of outcomes and explore the most impacting variables. In this chosen particular case, under Hd<sub>1</sub>, the results show that the most significant variables are "DNA quantity on hands", "Background" and "Extraction Efficiency", whereas under Hd<sub>2</sub>, the most impacting variable is the "match probability". As before, a variable is declared to be significant (boxplots coloured in green in Figure 5) if the range (the whole height of the boxplots shown in Figure 5) is above 1 order of magnitude of LR.

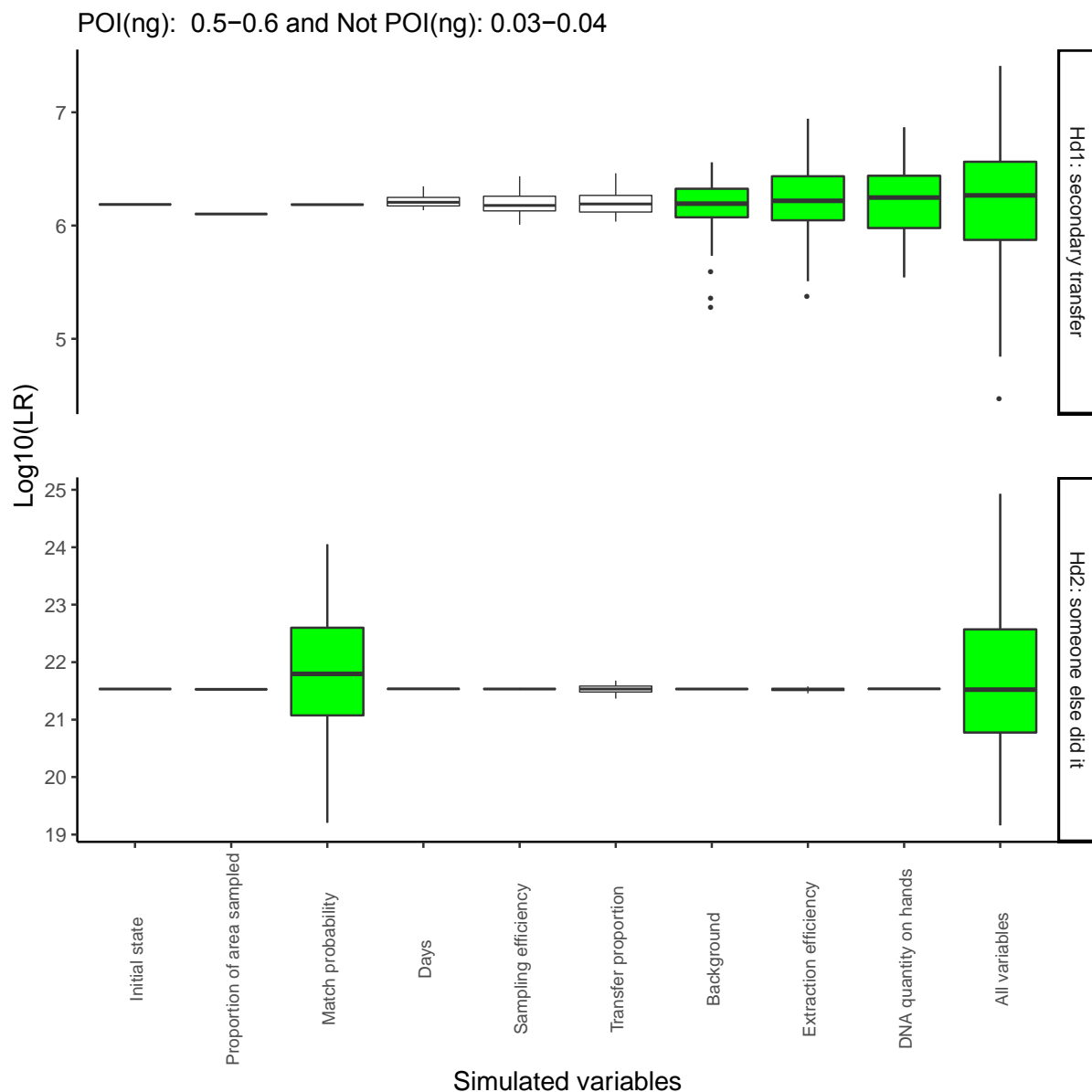


Figure 5: Boxplots of the LR (expressed in Log10) obtained after 100 simulations of each variable in both situations (Hd<sub>1</sub> and Hd<sub>2</sub>) with a quantity of POI's DNA between 0.5 and 0.6 ng with between 0.03 and 0.04 of not POI's DNA. The boxplots corresponding to the significant variables are coloured in green.

The purpose of setting up a rather complex simulation regime was twofold: (1) facing a large networks of connected variables, the simulations allow us to identify where there is a knowledge gap by identifying the most impacting variables and (2) to appraise the range of variations the computed LR may take given the limited data points that have informed each conditional probability table in the Bayesian network.

For the example in Figure 5, we note that when considering Hd<sub>1</sub> the variables significantly impacting are only 2 and relates to transfer mechanisms. Under Hd<sub>2</sub>, the match probability variable is dominating on all the other variables, making variables in relation to transfer less impacting in favour of the variable in relation to the source of the DNA. This is not surprising

as Hd<sub>2</sub> stipulated that the POI has no previous encounter with the knife and should his/her DNA profile be present, it must be from the AO whose profile, by chance, matches the POI or some background DNA which, by chance, matches the POI. Given that we are considering a relatively large amounts of DNA for the POI (0.5-0.6 ng), the chance of an adventitious correspondence is very small. The range of LR<sub>s</sub> observed stems from the fact that the simulation process accounts for the distribution of the number of alleles that could be derived from that quantity of DNA and for each of them the match probabilities that could potentially be associated. In other words, it accounts for a large range of potential matching DNA profiles each of them having different match probabilities depending on the number of alleles and their designations. In this paper the match probabilities are used as a plug-in to explore the relative importance of the variables, they do not guide precisely the sub-source level likelihood ratio. They are reasonable in terms of order of magnitude for single profiles, but they fail to account for the impact of mixtures on the sub-source level likelihood ratio. In a given case, with a specified DNA allelic profile, that variability will be extremely reduced as the match probability node will be directly informed with the sub-source level likelihood ratio computed for the case at hand. It will lead to different overall likelihood ratio compared to the values reported in Figure 5. When mixtures are involved, the node "Match probability" will be informed by 1 over the LR obtained for that mixture considering sub-source propositions.

In this example, when considering Hd<sub>1</sub> and Hd<sub>2</sub>, the overall range of the LR<sub>s</sub> is of 2 orders of magnitude but for a median log<sub>10</sub>(LR) of about 6 under Hd<sub>1</sub> and about 22 under Hd<sub>2</sub>. The 2 orders of magnitude have then a greater impact under Hd<sub>1</sub> rather than Hd<sub>2</sub> with regards to how it might influence the decision making of the recipient. We shall see below (in section 3.4) that one way to reduce the range is to increase the number of data points informing the impacting variables, that would be worth the investment for a case involving Hd<sub>1</sub> and not for a case involving Hd<sub>2</sub>.

Exploring the results in the Shiny application allows one to observe that, for some specific cases (a given amount of POI's DNA and not-POI's DNA), other variables may have a higher impact, hence be more critical to the case. That observation calls for a case specific approach if needed once a specific outcome has been observed in a given case. The general trend though is given by the variables identified globally.

*Impact on the LR of the quantities of DNA, POI and not POI respectively*

To show the respective impact of the quantity of POI's DNA and not POI's DNA on the LR obtained (summarized by their median), we can show two situations linked to the above example. In the first, we will maintain the quantity of not POI's DNA (0.03-0.04 ng) and vary the amount of POI's DNA. In the second, we will keep fixed the amount of POI's DNA (0.5-0.6 ng) and vary the not POI's DNA quantity. Both are shown in Figure 6.

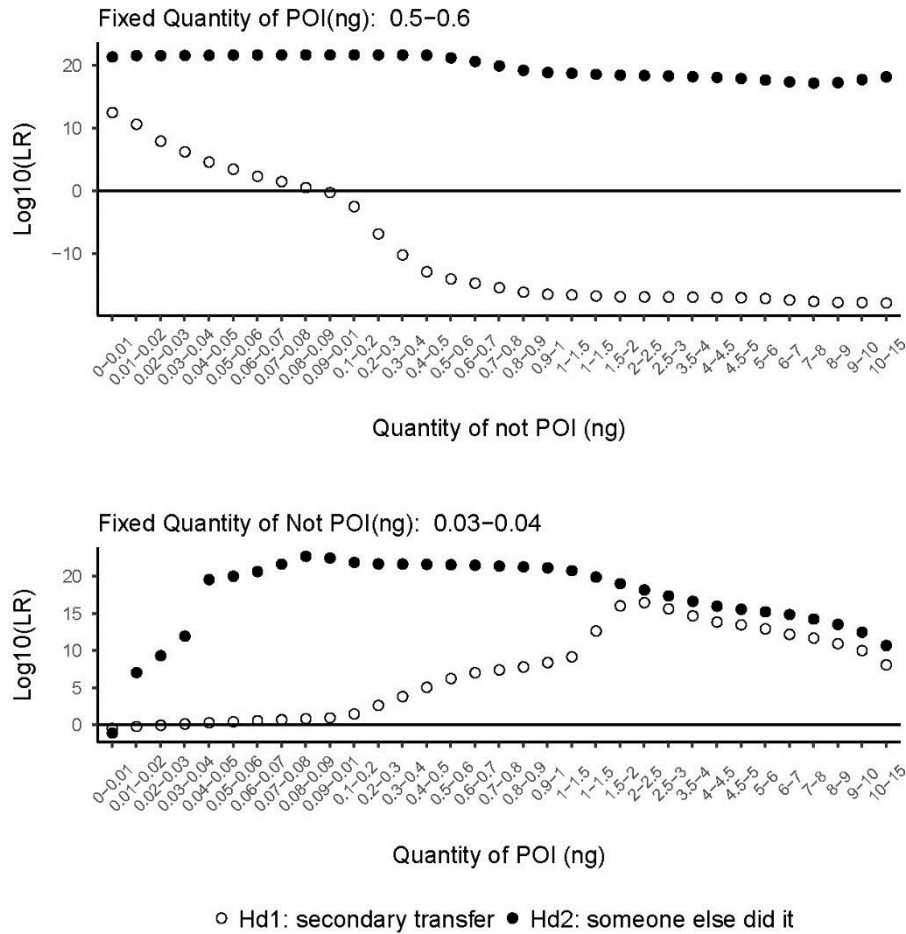


Figure 6: LRs obtained when the amount of DNA (POI and non-POI) are respectively varied for a given amount of DNA that remain fixed.

When the quantity of DNA of the POI is kept fixed (top graph of in Figure 6), under Hd<sub>2</sub>, the presence of non-matching DNA has little impact on the LR. It is the fact that Hd<sub>2</sub> stipulates that the POI has no link whatsoever of with the knife, hence the probability that some background DNA or AO DNA, would correspond to the POI, is the key consideration. The amounts have almost no effect. Under Hd<sub>1</sub>, however, the LR will be above 1 (the findings will provide support for Hp) with a maximum above one billion, but lower than 1 over the match

probability (value under  $Hd_1$ )<sup>5</sup>. When the quantity of not POI's DNA is increased, the LR drops gradually. The clinching point (when the LR is equal to 1 or 0 on a log10 scale) is with 0.07–0.08 ng of non-matching DNA. Above that amount, the non-matching DNA becomes more compatible with the stabbing activity, hence the findings overall would lend support for the defence.

When the quantity of DNA of the not POI is kept fixed and we increase the amount the quantity of DNA corresponding to the POI (bottom graph of in Figure 6), under  $Hd_2$ , the likelihood ratio gradually increases with the increased amount DNA corresponding to the POI. The maximum LR is obtained when the quantity is the most expected when handling a knife (0.07–0.08 ng). Above that quantity, the LR will gradually reduce. Under  $Hd_1$ , the LR will increase only when the quantity of DNA corresponding to POI reaches the point that it is more compatible with the POI stabbing scenario than under the POI handshaking scenario. That clinching point is at the same quantity as the quantity of not POI's DNA (0.03–0.04 ng). That is logical because we have modelled the transfer probabilities in the same way for both POI and AO. Then, adding quantity of DNA corresponding to the POI will gradually increase the LR up until the quantity that is best expected under the primary transfer and not under the scenario of a secondary transfer. Then it will reduce again. The LRs lending support for the defence spans over a large range of POI corresponding DNA quantities despite the presence of non-matching DNA. It stems from the fact that the non-matching DNA is a quantity that is more compatible with background level than with a quantity you would expect following primary transfer.

The above considerations will change as a function of the choice of quantities, hence in the Shiny application, the one can find the same representations as in Figure 6 but for any given choice of POI or not POI quantities of DNA.

#### *Increasing the datasets informing impacting variables*

Should we judge that the overall variation on the ranges of LRs obtained is too high (such in the example shown in Figure 5 under  $Hd_1$ ), one way to reduce it is to increase the number of data points informing the most impacting variables. Indeed, large variations in LRs translate limitations on the size of the datasets constituting the knowledge used to inform the CPTs. But before rushing into conducting specially designed experiments to acquire additional data

---

<sup>5</sup> Interestingly, while this is true of the values we found, there is nothing stopping the LR from being greater than 1 over the match probability (MP). If the probabilities of the transfers are extremely low, they can quite legitimately (in theory) lead to an LR that exceeds  $1/MP$ .



to inform the relevant conditional probability tables, we can again take advantage of the simulation strategy developed in this research. It will allow to assess if the analytical investment is worth the effort, hence lead to a reduction of the ranges of LR's that will be beneficial for the case at hand.

We could proceed considering any combination of results (POI and not POI's DNA) or jointly for all of them. In Figure 7, we show the results for the case presented before (POI's DNA of 0.5-0.6 ng and not POI's DNA of 0.03-0.04 ng).

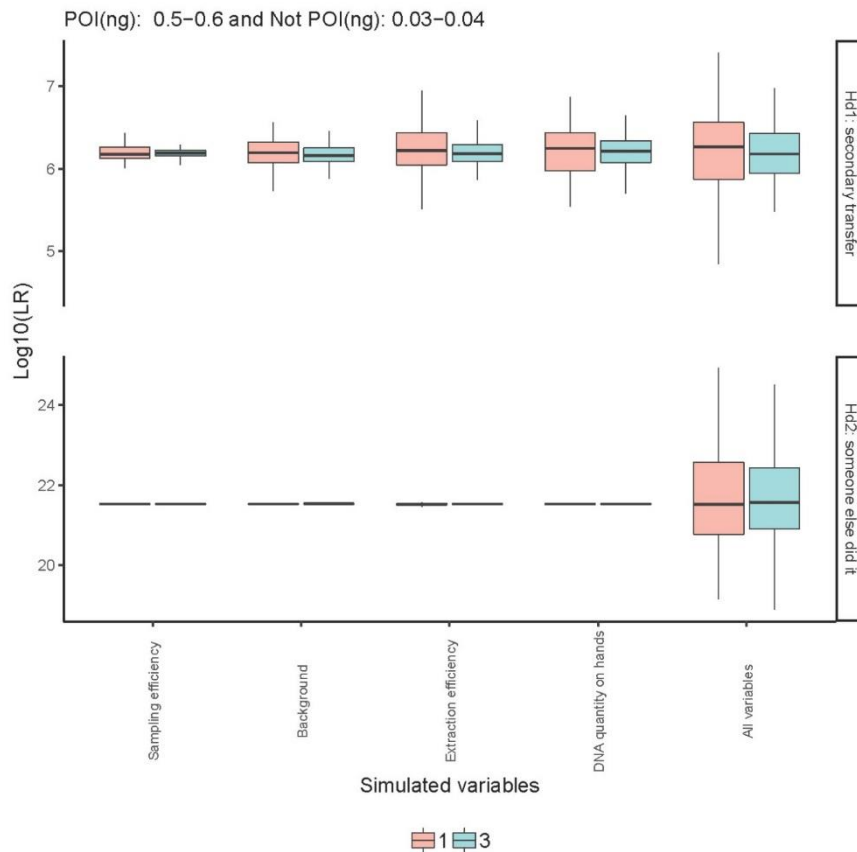


Figure 7: Boxplots of the LR's (expressed in Log10) obtained after 100 simulations of each variable in both situations (Hd<sub>1</sub> and Hd<sub>2</sub>) with a quantity of POI's DNA between 0.5 and 0.6 ng with between 0.03 and 0.04 of not POI's DNA. The boxplots in red are with the actual data, those in blue are obtained after increasing the underpinning data counts by a factor of 3 on the significant variables only.

As expected, when probabilities are informed by an increased number of experiments, the ranges of likelihood ratios decrease. Under Hd<sub>1</sub>, the reduction is greater than under Hd<sub>2</sub> due to the fact that the increase of data affected the significant variables (identified in Figure 4). For Hd<sub>2</sub> however, the increase of data does not have a strong impact on the range of LR observed for "All variables". This is due to the dominant impact on the LR of the match probability that is not affected by this increase in data. If Hd<sub>2</sub> is the defence proposition, there is no benefit in acquiring more data on the other variables that were significant under Hd<sub>1</sub>.

These simulations allow an assessment, in advance of investing in a large number of experiments to inform the CPTs, if the benefits, in terms of reduction of LR range, would be meaningful for the case at hand.

### **Conclusion**

The ENFSI guideline [3] is advising forensic scientist to evaluate biological traces with a low level of DNA considering activity level propositions. For that task, specific variables such as transfer, persistence, recovery and background need to be considered. Many experts face practical difficulties when considering these variables for various perceived reasons. Typically, expert will indicate that:

- the number of variables at play is overwhelming and unmanageable;
- every case represents a unique set of circumstances and any numerical assignment of probabilities is critically dependant on the specificities of the case;
- the paucity of current published studies to inform the parameters associated with these variables cannot reasonably be compensated by reasonable data acquisition campaigns.

In this study, we show that Bayesian networks can handle this complexity efficiently, when coupled with simulation techniques, they can be used to identify the most impacting variables, hence reducing the data acquisition burden by directing the laboratory to the key issue.

The method has been applied to a scenario involving trace DNA recovered from knife handles where the prosecution alleges that the person of interest (POI) stabbed a victim. The findings considered take the form of given quantities of DNA (in ng) corresponding or not the POI. As a general tendency, regardless of the findings, we showed that when the defence claims that the POI has nothing to do with the incident, the match probability associated with the POI will dictate most of the weight to be assigned to the findings. If the POI defence is invoking the possibility of secondary transfer, the key variables are associated with the sampling, the extraction efficiency, the background and the quantity of DNA on the hands.

Simulation techniques can also be used to assess the merit of increasing the knowledge base (in terms of size of studies carried out) when the significant variables had been identified. Finally, we noted that the identification of significant variables depends on the obtained DNA results and this selection may be refined on a case by case basis. To allow one exploring all

possibilities, a dedicated Shiny application has been designed ([https://lydie-samie.shinyapps.io/DNA\\_Activity/](https://lydie-samie.shinyapps.io/DNA_Activity/)).

## Bibliography

- [1] F. Taroni, A. Biedermann, J. Vuille, N. Morling, Whose DNA is this? How relevant a question? (a note for forensic scientists), *Forensic Sci. Int. Genet.* 7 (2013) 467–470.
- [2] C. Champod, DNA transfer: informed judgment or mere guesswork?, *Frontiers in Genetics.* 4 (2013) 300 in A. Biedermann, J. Vuille, F. Taroni, DNA, Statistics and the Law: A Cross-Disciplinary Approach to Forensic Inference, *Frontiers Media.* 4 (2014) 22-24.
- [3] S. Willis et al., ENFSI guideline for evaluative reporting in forensic science, European Network of Forensic Science Institutes, Dublin (2015). [http://enfsi.eu/wp-content/uploads/2016/09/m1\\_guideline.pdf](http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf) (accessed May 6, 2019).
- [4] A. Biedermann, C. Champod, G. Jackson, P. Gill, D. Taylor, J. Butler, N. Morling, T. Hicks, J. Vuille, F. Taroni, Evaluation of Forensic DNA Traces When Propositions of Interest Relate to Activities: Analysis and Discussion of Recurrent Concerns, *Frontiers in Genetics.* 7 (2016) 1–12.
- [5] D. Taylor, A. Biedermann, L. Samie, K.-M. Pun, T. Hicks, C. Champod, Helping to distinguish primary from secondary transfer events for trace DNA, *Forensic Sci. Int. Genet.* 28 (2017) 155-177.
- [6] F. Taroni, A. Biedermann, S. Bozza, P. Garbolino, C. Aitken, Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science, 2nd edition, *Statistics in Practice*, Chichester: John Wiley & Sons, Ltd (2014).
- [7] D. Taylor, T. Hicks, C. Champod. Using Sensitivity Analyses in Bayesian Networks to Highlight the Impact of Data Paucity and Direct Future Analyses: A Contribution to the Debate on Measuring and Reporting the Precision of Likelihood Ratios. *Science & Justice*, 56 (2017) 402-410.
- [8] P. Ka-Man, Interprétation des profils génétiques obtenus à partir des traces de contact, PhD thesis, School of Criminal Justice, University of Lausanne, 2016.
- [9] R. Palmer, The Evaluation of Fibre Evidence in the Investigation of Serious Crime, PhD thesis, School of Criminal Justice, University of Lausanne, 2016.
- [10] M. Khodabina, A. Ahmadabadi, Some properties of generalized gamma distribution, *Mathematical Sciences* 4 (2010) 9-28
- [11] D. Dufresne, The Beta Product Distribution with Complex Parameters, *Commun. Stat. —Theory Methods.* 39:5 (2010) 837–854

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

---

- [12] D.J. Daly, C. Murphy, S.D. McDermott, The transfer of touch DNA from hands to glass, fabric and wood, *Forensic Sci. Int. Genet.* 6 (2013) 41–46.
- [13] L. Bontadelli, Study of DNA Shedder Quality, MSc thesis, School of Criminal Justice, University of Lausanne, Lausanne, 2009.
- [14] M. Goray, R.J. Mitchell, R.A.H.V. Oorschot, Investigation of secondary transfer of skin cells under controlled conditions, *Leg. Med.* 12 (2010) 117–120.
- [15] T.J. Verdon, R.J. Mitchell, R.A.H.V. Oorschot, Evaluation of tapelifting as a collection method for contact DNA, *Forensic Sci. Int. Genet.* 8 (2014) 179–186.
- [16] J.M. Bernardo and A.F.M Smith, *Bayesian Theory*, John Wiley & Sons, Inc (2000)
- [17] E. Butts, Exploring DNA Extraction Efficiency Forensics@NIST 2012 Meeting, Gaithersburg [https://www.nist.gov/sites/default/files/documents/oles/3\\_Butts-DNA-extraction-2.pdf](https://www.nist.gov/sites/default/files/documents/oles/3_Butts-DNA-extraction-2.pdf) (2017).
- [18] G.E. Meakin, E. V. Butcher, R.A.H. van Oorschot, R.M. Morgan, The deposition and persistence of indirectly-transferred DNA on regularly-used knives, *Forensic Sci. Int. Genet. Suppl. Ser.* 5 (2015) e498-e500.
- [19] RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.
- [20] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [21] Kjell Konis. (2017). RHugin: RHugin. R package version 8.4. <http://rhugin.r-forge.r-project.org>] to link with the developed Hugin network.
- [22] L. Samie, T. Hicks, V. Castella, F. Taroni, Stabbing simulations and DNA transfer, *Forensic Science International: Genetics* 22 (2016) 73–80.
- [23] C. Gehrig, B. Balitzki, A. Kratzer, C. Cossu, N. Malik, V. Castella, Allelic proportions of 16 STR loci – including the new European Standard Set (ESS) loci – in a Swiss population sample, *Int. J. Legal Med.* 128 (2013) 461–465.
- [24] T. Hicks, F. Taroni, J. Curran, J. Buckleton, O. Ribaux, V. Castella, *Forensic Science International : Genetics Use of DNA profiles for investigation using a simulated national DNA database : Part I . Partial SGM Plus 1 profiles*, *Forensic Sci. Int. Genet.* 4 (2010) 232–238.
- [25] Samie L., Champod C., Glutz V., Garcia M., Castella, V. & F. Taroni, The efficiency of DNA extraction kit and the efficiency of recovery techniques to release DNA using flow cytometry, *Science & Justice*, 59 (2019) 405-410
-

Il.2 Samie et al., Estimating the quantity of transferred DNA in primary and secondary transfers, Science & Justice (2019) (Soumis pour publication, accepté, en cours de production)

Lydie Samie<sup>1,2</sup>, Franco Taroni<sup>1</sup> & Christophe Champod<sup>1</sup>

<sup>1</sup> Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice, University of Lausanne, Switzerland

<sup>2</sup> Forensic Genetics Unit, University Center of Legal Medicine Lausanne and Geneva, Lausanne, Switzerland

Corresponding Author: Lydie Samie: [lydie.samie-foucart@unil.ch](mailto:lydie.samie-foucart@unil.ch), [lydie.samie-foucart@chuv.ch](mailto:lydie.samie-foucart@chuv.ch)

**Keywords:** Quantity of DNA, DNA transfer, primary transfer, secondary transfer.

## Introduction

According to the ENFSI guideline on evaluative reports [1], the evaluation of biological stains, especially traces with a low quantity of DNA, should be carried out considering activity-level propositions. It involves a relative assessment of the expected quantities of recovered DNA under the alleged activities put forward by the parties. In order to do so, the respective shedder status (or shedding ability) of the person of interest and of the alternative offender should be investigated.

Previous studies have dealt with the shedder status of donors [2, 3, 4 and 5]. They all reported large variations between individuals in the amount of contact DNA that each donor may leave on a receiving surface; some individuals transfer more DNA than others. In addition, Pesaresi *et al.* [6], van Oorschot *et al.* [7] and Bright and Petricevic [8] show that variations can be observed in the amount of DNA a given individual may deposit. These studies show that variation within an individual should be taken into account to assess the probability of observing a given quantity of contact DNA.

In the present study, we will show that the DNA shedding ability of an individual should be characterized as a *distribution* of the quantity of DNA present on hands or transferred on surfaces. Individuals do not have fixed shedder status (such as “good” or “bad”) regardless of

the circumstances. Indeed, a given individual may deposit a mean quantity of DNA, but due to the inherent within-source variability, may also, at times, deposit, much more or less than this quantity. So, the probability of observing a given quantity of DNA should account for this distribution. We will inform this distribution by a measure of its mean and spread. In addition, the amount of DNA available to be shed from a hand to a surface depend on the conditions of the hands at the time of transfer (e.g. sweaty or dry). Lacerenza *et al.* [9] indicated that life habits have no impact on the recovered DNA quantity on hands except for the habit of touching the hairy surfaces. Touching his/her hairs increases the quantity of DNA recovered on hands. Our experimental design will consider a range of quantities of DNA on hands.

The above literature on the shedder status is mostly concerned with primary transfer and not with secondary or subsequent transfers. In this study, we will deal with two situations involving a knife handle; the first is a primary transfer from a hand to a knife handle and the second is a secondary transfer from a Person of Interest (POI) to the hand of an intermediate person who then took the knife handle. This is not the first time that transfer on surfaces is studied [3, 10, 11, 12, 13, 14], but these studies have some limitations. All researchers studied the probabilities of primary or secondary transfer of DNA but without considering the inherent variability due to the donor.

After the touch a surface by a POI, it is frequent to observe in addition to his/her DNA contribution, the DNA contribution of additional individuals [9, 14]. Modern probabilistic genotyping systems (such as STRmix, <https://www.strmix.com/>) allows to deconvolute these mixtures and, from the estimated mixing proportion, derive the effective quantity corresponding to the POI. That approach was already adopted by [11, 12]. In this study we will explore if such deconvolution is required to assess the quantity of DNA left by the POI or if the total quantity of DNA is sufficiently informative.

We will also investigate if the quantity of transferred DNA on an object can be predicted from the measure of the DNA quantity available on the hand and the application of a transfer proportion (TP) that will be fixed for each individual. Quantifying the amount of DNA on the hands has been made by McColl *et al.* [15] but only looking at the variability between donors and not reporting on the variability within donors.

To sum up, this study has three objectives: (1) to characterize the distribution of the quantity of DNA observed on the hands of individuals and transferred on surfaces either through

primary or secondary transfer; (2) to assess if deconvolution of the DNA profiles is required to estimate the quantity of DNA of the POI; and finally (3) to test if the transfer proportion (quantity transferred on the surface over the initial quantity on the hand) is similar across individuals and can be used to predict the quantity of transferred DNA.

## **Methodology**

### *Transfer Experiments*

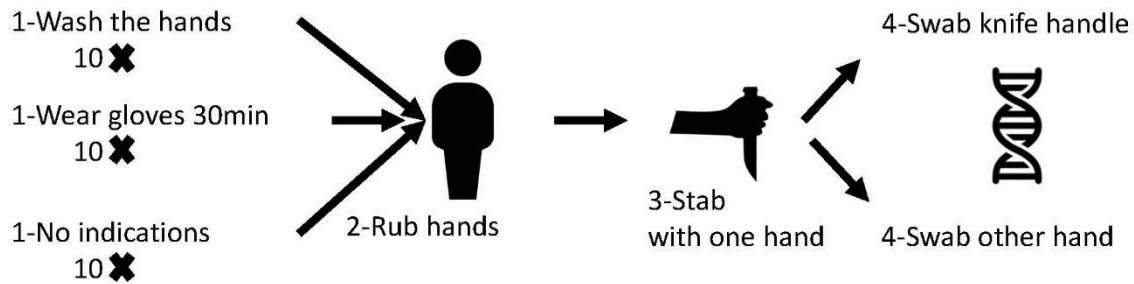
Six consenting participants, three men and three women, were randomly selected to deposit contact DNA following activities of primary or secondary transfer.

For primary transfer, each participant was asked to rub their hands during around five seconds [13] with a view to redistribute surface DNA evenly on both of them [16], then took a knife (Stainless Steel, X50 Cr Mo V15) handle with their usual hand and, immediately after, stab three times a ballistic soap (from Mettler SA). 30 stabbing experiments were performed for each participant, leading in total to 180 experiments. Before each experiment, the knife was thoroughly cleaned, using bleach and ethanol.

The duration of the contact, the type of contact and the force of the stabbing were not specified in order to simulated conditions as closed as possible than casework. The ballistic soap allowed mimicking the physical properties of a human body. This direct transfer on the knife handle is what will be considered as primary transfer. The entire surface of the knife handle and the inside part of the other hand, meaning the palm and the fingers inside the hand, not used for the activity, were swabbed just after the stabbing to collect DNA using the FLOQSwab™ from COPAN. One FLOQSwab™ was used per sample, following the procedure of the laboratory. The knife handle being a smooth surface, the FLOQSwabs™ were moist.

The stabbing conditions used for the experiments were adapted in order to increase or decrease the quantity of DNA initially on the surface of the hand and subsequently transferred. These variations aimed at reflecting an extreme range of life conditions for a given individual. For a first set of ten experiments out of thirty, each participant was asked to wash their hands just before performing the stabbing. For the next set of ten experiments, they were asked to wear gloves for 30 minutes to increase sweating. For the last set of experiments, no specific indication was given to the participants. Each set were performed on

different days. However, within each set, some experiments were conducted on the same day. For the washing and glove wearing conditions, it has no bearing. For the last condition (no specific indication), a sufficient time between experiments (about an hour) was allowed. The above-described experimental design, as performed by the six participants, is illustrated in Figure 1.



*Figure 1: Illustration of the experimental design to study the quantity of DNA on hands and the quantity of DNA transferred during the primary transfer.*

To study secondary transfer, only two participants were chosen in the light of the first set of experiments. Based on their mean quantities of transferred DNA, participant 2 and participant 6 have shown to be the “best” and the “worst” DNA donor respectively (See Figure 3, Figure 4, Figure 7, Table 1 and Table 2). Two identical knives, one for each participant, were used for all their experiments. Before each experiment however, the knife was thoroughly cleaned, using bleach and ethanol. Both participants were first asked to shake hands and then to stab the ballistic soap with the knife. No indication on the duration of the handshake was given to the participants in order to mimic real life conditions as closely as possible, the contact though didn't exceed by few seconds. The entire surface of the two knife handles were then swabbed for DNA just after the stabbing using one moist COPAN's FLOQSwab™, following our laboratory procedure. Thirty experiments were performed for each of the two participants (leading to 60 experiments in total). Experiments were subsequently performed with a minimum delay of five minutes between them. This experimental design is illustrated in Figure 2.



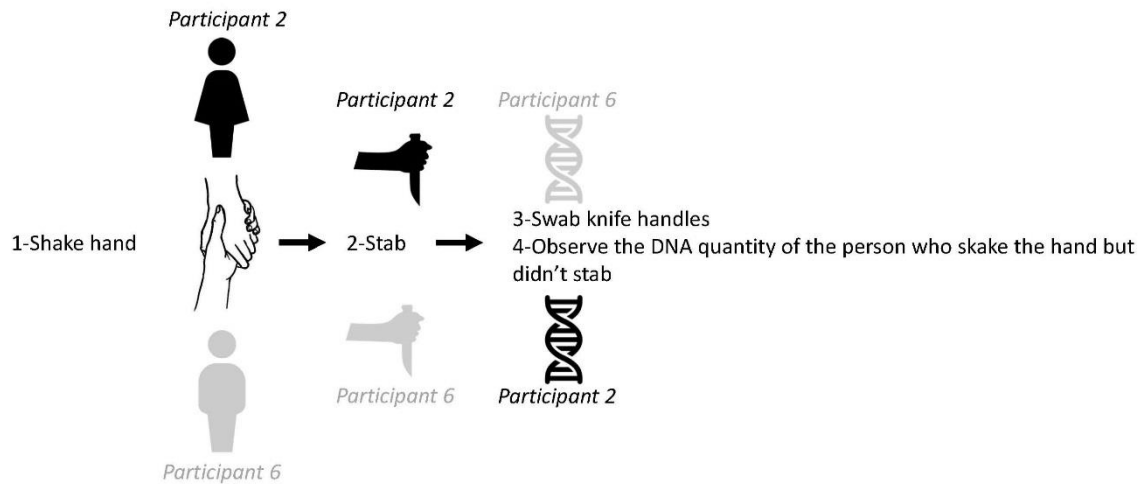


Figure 2 : Illustration of the experimental design to study the secondary transfer of the first participant's DNA (Participant 2, in black) and the second participant's DNA (Participant 6, in gray), respectively.

### *Quantification of DNA*

DNA was extracted from the swabs using a combination of two kits: QIAshredder and QIAamp DNA mini kit from Qiagen, concentrated to a final volume of 25 $\mu$ L with microcon® 30 spin column. Quantifications were performed directly following the DNA extraction using the Investigator® Quantiplex kit from Qiagen on Rotor-Gene® Q. DNA was then amplified at 30 cycles using 10  $\mu$ L of DNA extract per sample and the NGM SElect (Applied Biosystem™-Thermofisher) kit with a PCR system 9700 (Applied Biosystem™), analyzed on a 3500 Series Genetic Analyzers (Applied Biosystem™-Thermofisher Scientific) coupled with GeneMapper1IDX Software (Applied Biosystem™-Thermofisher Scientific). The kits were used as per manufacturer's instructions.

DNA quantification allows to obtain information about the total quantity of DNA recovered from the knife handle. That quantity may result from a mixture of DNA of the POI and of other contributors. To estimate the proportion of DNA corresponding to the POI, STRmix™ v2.5.11 software is used to assess the mixing ratio from each donor in the mixture. The number of contributors entered in the software for each case is based on the number of the peaks detected at each locus, peak height balance information and how the experiments were designed (i.e., we expected one, two or three person's DNA).

*Deriving the parameters of the transfer proportion*

For primary transfers, the parameters of the distribution for the log<sub>10</sub> of the transfer proportion (log<sub>10</sub>(TP)) for each individual is obtained by combining the results of the initial quantity of DNA on hands (Q<sub>i</sub>) and the results of the quantity of DNA observed on the knife handle (Q<sub>f</sub>), under the assumption that both Q<sub>i</sub> and Q<sub>f</sub> are Normally distributed [17]. When transformed in log<sub>10</sub>, the parameters of the distribution for log<sub>10</sub>(TP) are obtained as follows [18]:

$$\text{mean}(\log_{10}(\text{TP})) = \text{mean}(\log_{10}(\text{Q}_f) - \text{mean}(\log_{10}(\text{Q}_i))$$

and

$$\text{SD}(\log_{10}(\text{TP})) = \text{SD}(\log_{10}(\text{Q}_f)) + \text{SD}(\log_{10}(\text{Q}_i)) - 2 * \sqrt{\text{SD}(\log_{10}(\text{Q}_f)) * \text{SD}(\log_{10}(\text{Q}_i))}.$$

The same transfer proportion parameters can be computed for the secondary transfers taking the quantity of DNA matching the POI left on the knife handle following secondary transfer as Q<sub>f</sub>.

## **Results**

### *Quantity of DNA present on the hand and following primary transfer*

The initial quantity of DNA on hands and the quantity of DNA directly transferred on the knife handle are shown in Figure 3 and Figure 4, distinguishing the total quantity and the quantity corresponding to the POI (adjusted using the mixing proportions estimated using STRmix™).

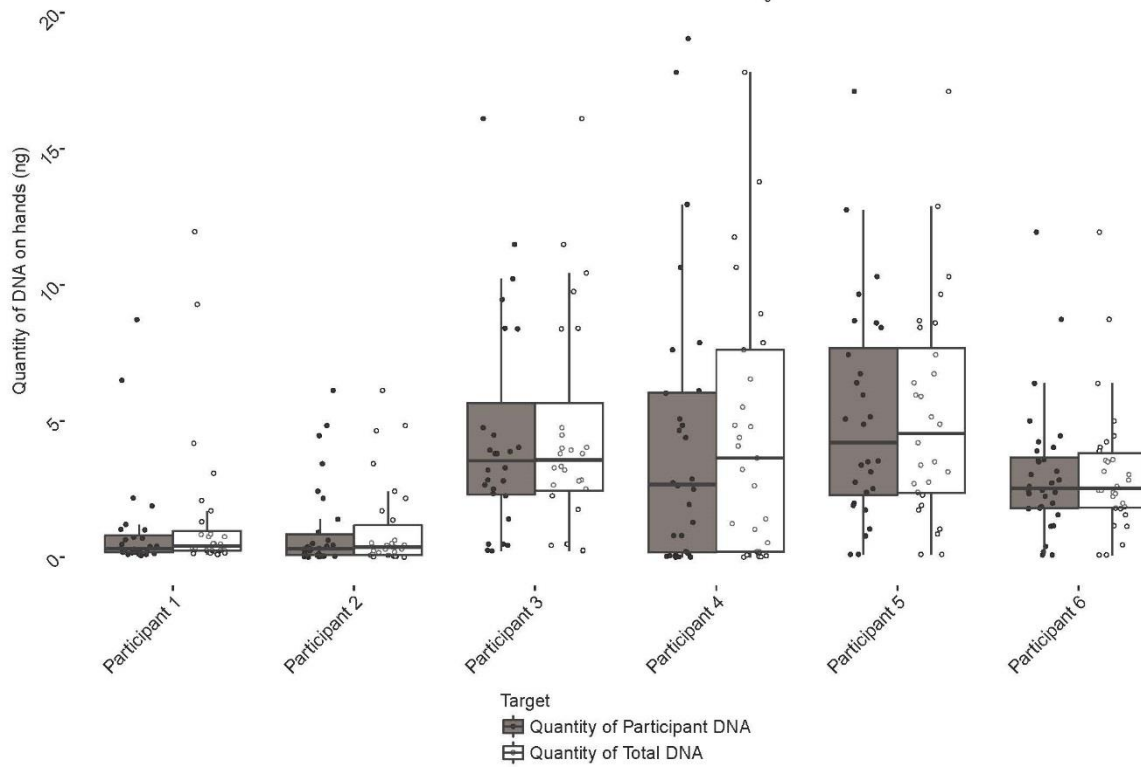


Figure 3 : Boxplots of the total quantity of DNA (and the quantity corresponding to each participant) recovered from the hand. Each dot corresponds to the corresponding quantity obtained after each experiment.

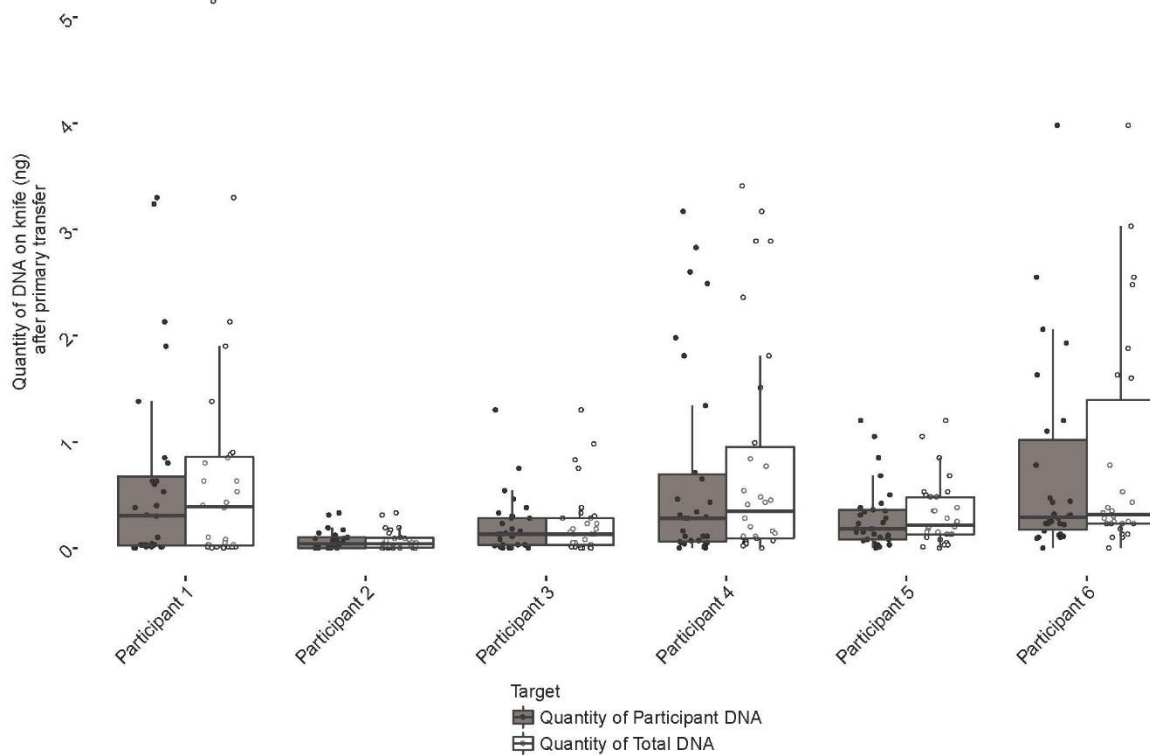


Figure 4: Boxplots of the total quantity of DNA (and the quantity corresponding to each participant) recovered from the knife handle. Each dot corresponds to the corresponding quantity obtained after each experiment.

A large variation of the quantity of DNA collected on participants' hand (Table 1) and the quantity recovered from the knife handle after a direct transfer (Table 2) is observed between participants. Indeed, the mean value of total DNA range from 1 ng to 5 ng. A large variation for each participant is also observed as can be seen from the ranges (max-min) of DNA quantities. For participant 1 for example, between 0 and more than 5ng of DNA can be recovered after directly handling the knife handle depending on the experiment (and between 0 and more than 11ng directly from his hand).

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

Table 1: Summary statistics of the quantities of total DNA and of the participant's DNA (obtained following mixture deconvolution) recovered on his hand.

The quantity of DNA on the other hand (ng)		Min	0.05 percentile	Median	Mean	SD	0.95 percentile	Max
Participant 1	Total DNA	0.10	0.11	0.42	1.47	2.78	7.50	11.95
	Participant 1 DNA	0.06	0.10	0.31	1.02	1.96	4.98	8.72
Participant 2	Total DNA	0.00	0.02	0.38	1.06	1.63	4.76	6.12
	Participant 2 DNA	0.00	0.01	0.31	1.02	1.62	4.66	6.12
Participant 3	Total DNA	0.23	0.31	3.57	5.03	4.94	14.48	21.03
	Participant 3 DNA	0.23	0.31	3.54	4.96	4.95	14.48	21.03
Participant 4	Total DNA	0.00	0.01	3.64	4.94	5.54	16.19	20.48
	Participant 4 DNA	0.00	0.00	2.68	4.39	5.23	16.01	19.04
Participant 5	Total DNA	0.10	0.36	4.54	5.29	4.04	11.97	17.10
	Participant 5 DNA	0.10	0.34	4.20	5.21	4.06	11.89	17.10
Participant 6	Total DNA	0.08	0.25	2.53	3.15	2.45	7.67	11.93
	Participant 6 DNA	0.08	0.22	2.53	3.15	2.43	7.67	11.93

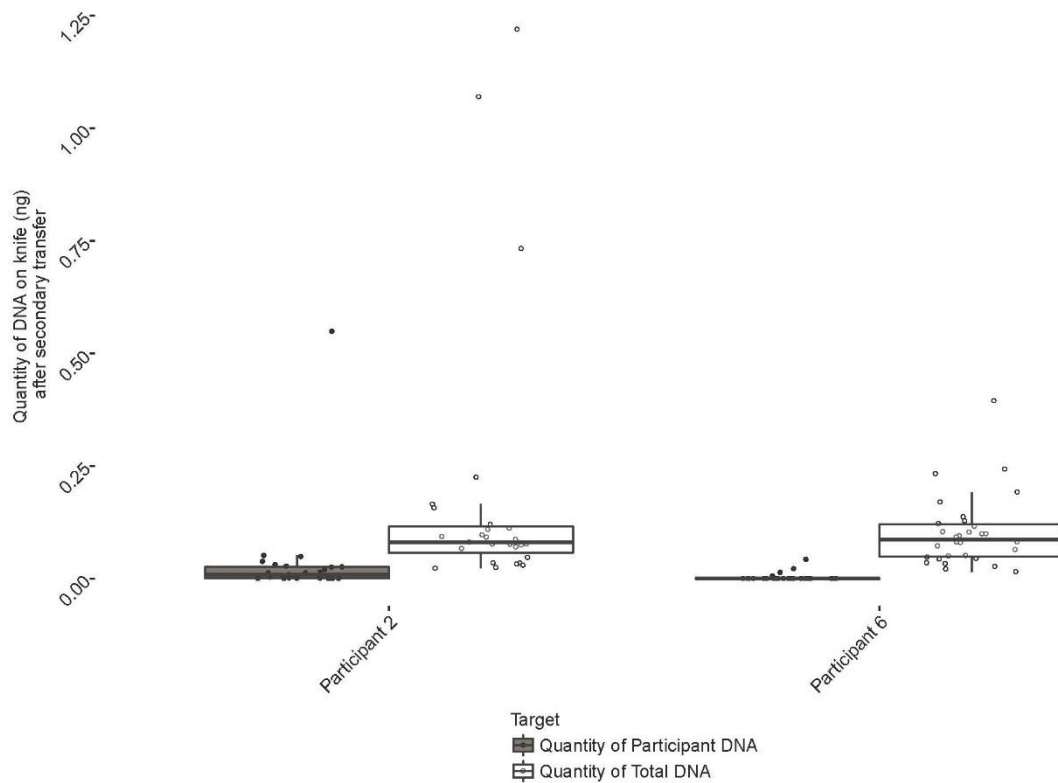
Table 2: Summary statistics of the quantities of the total DNA and of participant's DNA (obtained following mixture deconvolution) recovered on the knife handle after the participant directly stabbed a ballistic soap with the knife (primary transfer).

The quantity of DNA recovered on the knife handle after direct transfer (ng)		Min	0.05 percentile	Median	Mean	SD	0.95 percentile	Max
Participant 1	Total DNA	0.00	0.0035	0.39	0.74	1.17	2.89	5.15
	Participant 1 DNA	0.00	0.00	0.31	0.63	0.93	2.82	3.30
Participant 2	Total DNA	0.00	0.00	0.04	0.08	0.09	0.27	0.33
	Participant 2 DNA	0.00	0.00	0.04	0.07	0.09	0.29	0.33
Participant 3	Total DNA	0.00	0.00	0.13	0.24	0.32	0.92	1.30
	Participant 3 DNA	0.00	0.00	0.13	0.21	0.28	0.67	1.30
Participant 4	Total DNA	0.00	0.03	0.35	0.82	1.07	3.04	3.41
	Participant 4 DNA	0.00	0.00	0.28	0.70	0.97	2.73	3.17
Participant 5	Total DNA	0.00	0.02	0.22	0.32	0.30	0.96	1.20
	Participant 5 DNA	0.00	0.02	0.18	0.29	0.30	0.96	1.20
Participant 6	Total DNA	0.00	0.10	0.32	0.95	1.30	3.55	5.25
	Participant 6 DNA	0.00	0.10	0.29	0.84	1.23	3.34	5.25

---

*Quantity of DNA following secondary transfer*

The quantities of DNA obtained following the secondary transfer experiments are given in Figure 5 and Table 3. For participant 2 for example, about 0.2 ng of total DNA can be recovered on the knife handle after a secondary transfer with 0.03 ng of DNA corresponding to the participant's 2 DNA profile. POI. Whereas, for participant 6, 0.1 ng of total DNA can be recovered on the knife handle after a secondary transfer with only 0.003 ng of DNA corresponding to his DNA profile. A marked difference is observed for the two participants between the total quantity of DNA and the quantity of DNA corresponding to the POI.



*Figure 5: Boxplots of the DNA quantities for participant 2 and 6 obtained indirectly on the knife handle following secondary transfer. Each dot corresponds to the corresponding quantity obtained after each experiment.*

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

Table 3: Summary statistics of the quantities of the total of DNA and participant's DNA recovered on the knife handle after this participant shook hands with another participant who stabbed a ballistic soap with the knife. In this situation, Participant 2 shook hands with participant 6 then Participant 6 stabbed the ballistic soap and vice versa.

The quantity of DNA recovered on the knife handle after secondary transfer (ng)		Min	0.05 percentile	Median	Mean	SD	0.95 percentile	Max
Participant 2	Total DNA	0.02	0.03	0.08	0.19	0.31	0.97	1.22
	Participant 2 DNA	0.00	0.00	0.01	0.04	0.11	0.05	0.55
Participant 6	Total DNA	0.02	0.02	0.09	0.10	0.08	0.24	0.40
	Participant 6 DNA	0.00	0.00	0.00	0.00	0.01	0.02	0.04

POI's DNA: comparing hands, primary, and secondary transferred quantities

If we focus our attention on the quantity of DNA corresponding to the POI for the three cases studied (hand, primary transfer and secondary transfer), we recorded large variations of that quantity within participant and between participants. Figure 6 and Table 4 bring together these data (already shown in part before).

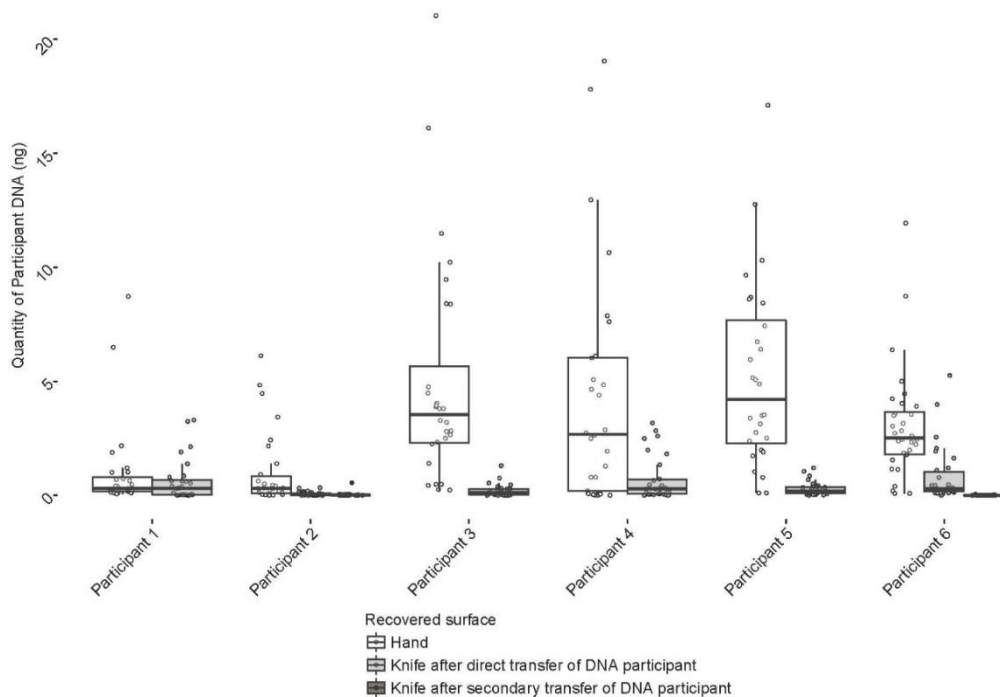


Figure 6 : Boxplots of the DNA quantities for each participant recovered on the participant's hand, on the knife handle after primary transfer and secondary transfer (only for participants 2 and 6). Each dot corresponds to the corresponding quantity obtained after each experiment

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

*Table 4: Summary statistics of the quantities of participants' DNA on participant's hands, participants' DNA on the knife handle after direct transfer and participants' DNA on the knife handle after secondary transfer respectively for each participant.*

	POI's quantity of DNA (ng)	Min	0.05 percentile	Median	Mean	SD	0.95 percentile	Max
<b>Participant 1</b>	On hands	0.06	0.10	0.31	1.02	1.96	4.98	8.72
	After direct transfer	0.00	0.00	0.31	0.63	0.93	2.82	3.30
<b>Participant 2</b>	On hands	0.00	0.02	0.38	1.06	1.62	4.74	6.12
	After direct transfer	0.00	0.00	0.04	0.07	0.09	0.29	0.33
	After secondary transfer	0.00	0.00	0.01	0.04	0.11	0.05	0.55
<b>Participant 3</b>	On hands	0.23	0.31	3.54	4.96	4.95	14.48	21.03
	After direct transfer	0.00	0.00	0.13	0.21	0.28	0.67	1.30
<b>Participant 4</b>	On hands	0.00	0.00	2.68	4.39	5.23	16.01	19.04
	After direct transfer	0.00	0.00	0.28	0.70	0.97	2.73	3.17
<b>Participant 5</b>	On hands	0.10	0.34	4.21	5.21	4.06	11.89	17.10
	After direct transfer	0.00	0.02	0.18	0.29	0.30	0.96	1.20
<b>Participant 6</b>	On hands	0.08	0.22	2.53	3.15	2.42	7.67	11.93
	After direct transfer	0.00	0.10	0.29	0.84	1.23	3.34	5.25
	After secondary transfer	0.00	0.00	0.00	0.00	0.01	0.02	0.04

The standard deviation (SD) observed on the quantity of DNA generally reduces for each donor when we move from hand, to primary transfer and subsequently to secondary transfer.

We have observed no obvious relationship between the quantity of POI's DNA recovered on the hand and the quantity of transferred DNA. For example, participant 3 has, in general, a large quantity of DNA on his hand compared to the other participants. However, this donor transferred a very small quantity of DNA on the knife handle through primary transfer. On the contrary, small quantities of DNA are recovered from the hand of the participant 1, compared to other participants, but he transferred a large part of that DNA on the handle. Hence, for primary transfer, there is no fixed transfer proportion (TP) for all participants as shown in Table 5. Participant 1 and participant 6 proportionally left more of their DNA than the other



participants. They both gave an average of 20% on the knife handle, whereas, participant 3 for example transferred an average of 7% only.

Primary TP	Mean	SD
<b>Participant 1</b>	0.20	0.25
<b>Participant 2</b>	0.13	0.19
<b>Participant 3</b>	0.07	0.14
<b>Participant 4</b>	0.14	0.23
<b>Participant 5</b>	0.11	0.19
<b>Participant 6</b>	0.20	0.25

Table 5: Means and standard deviations computed for the primary transfer proportion for each participant.

Figure 7 illustrates these differences in TPs between participants. Each boxplot represents 1000 data points that have been randomly selected from a Beta distribution with parameters set from the mean and the standard deviation specified in Table 6.

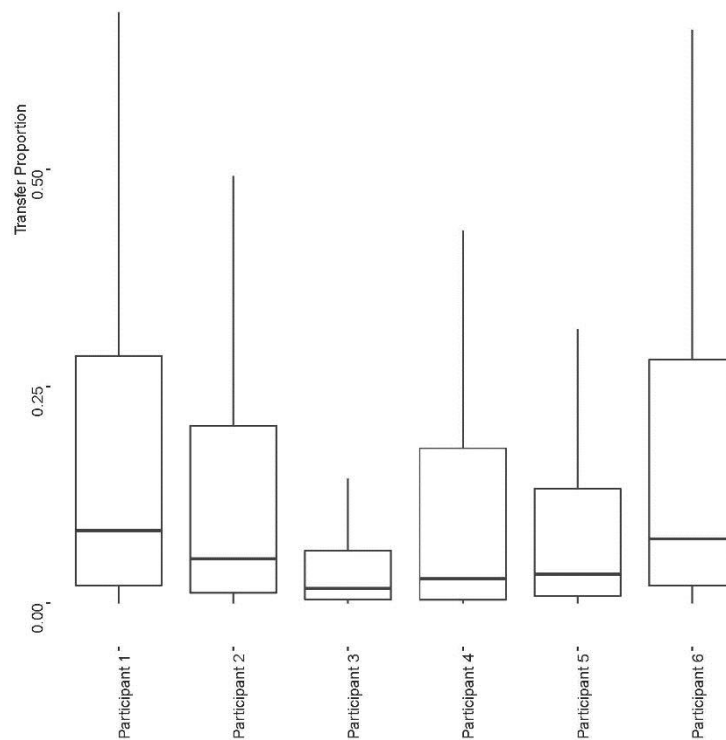


Figure 7: Boxplots of 1000 direct transfer proportions of DNA simulated from each participant corresponding distribution.

The proportions of transfer can also be computed for secondary transfers (against the quantity on the hand) as shown in Figure 7. The mean secondary transferred TP for participant 2 has an average of 1% whereas it is 3% for participant 6.

Secondary TP	Mean	SD
Participant 2	0.01	0.03
Participant 6	0.03	0.11

Table 6: Means and standard deviations computed for the secondary transfer proportion for the two participants

When these proportions are compared to the primary TP, they differ largely for the two participants. As before, these differences are illustrated graphically by re-sampling from the respective distributions (Figure 8).

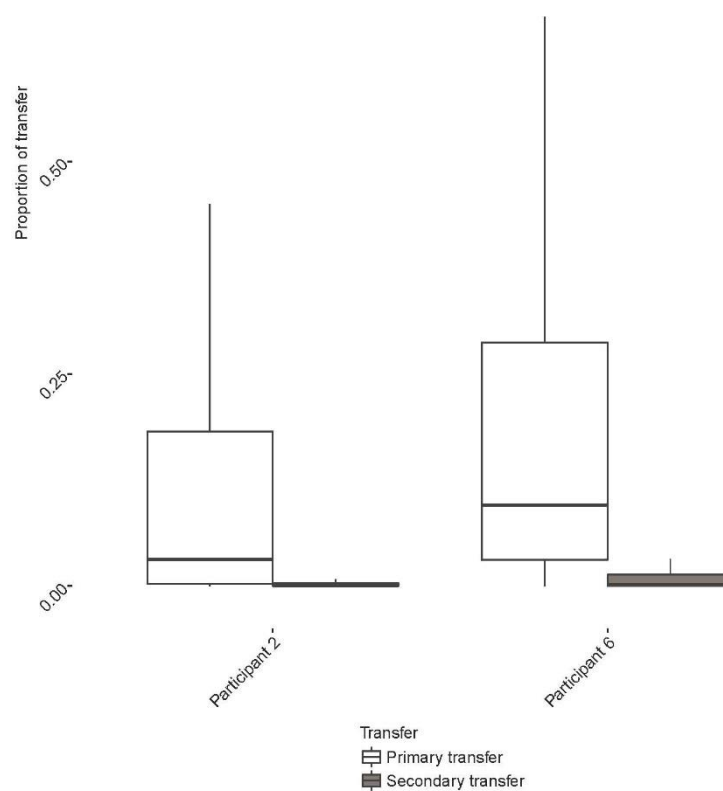


Figure 8: Boxplots of the direct transfer proportions and secondary transfer proportions for each participant (1000 data points randomly generated from the corresponding distributions).

## Discussion and conclusion

### Comparison of the results with other studies.

On the hands of the participants to this study, we observed a total quantity of DNA between 0 and 21 ng, made in majority of the donor's DNA with, on average, less than 8% of non-self DNA. That percentage of non-self-DNA can vary substantially between donors. Take

participant 1, for example, the total quantity of DNA obtained from his hand is contributed by only 70% of his own DNA. These quantities can be comparable to those obtained by Szkuta *et al.* [14]. They observed between 0.1 and 85.5 ng of DNA on 70 hands.

McColl *et al.* [15] observed higher quantities between 0 and 585ng. However, they studied a larger number of hands (120 hands), and that could explain the difference. However, the percentage of non-self DNA recovered on hands is similar to the percentage observed in the present study (an average of 8.5% of non self DNA and maximum less than 30%).

Following the primary transfers of DNA on the knife handles, we observed a total quantity of DNA ranging from 0 to 5ng. We observed (Table 2) that on average less than 8% of the total quantity originates from a different contributor than the donor. There are variations between donors with regards to the non-self DNA present on their hands and transferred on the handle. For example, 40% of the total quantity of DNA on the handles used by participant 1 comes from someone else. These results are in line with those obtained by Goray *et al.* [5], Samie *et al.* [10] and Szkuta *et al.* [14]. They reported recovered quantities of DNA between around 0 and 5ng [5], 0 and 5 ng [10] and 0 and 7ng [14].

We note however that other researchers have reported higher quantities of DNA transferred on knife handles, namely:

- Meakin *et al.* [11]: They reported between 3 and 10 ng of total DNA recovered on the knife handles with less than 3% of non-self DNA for 3 donors and 25% with one donor.
- Butcher *et al.* [12]: They reported between 1 and 10 ng of total DNA recovered on the knife handles with less than 16% of non-self DNA.

In our opinion, the differences observed may be due to the fact that in our experiments (as in others [5, 10, 14]), the surfaces were cleaned before each experiment, whereas in [11, 12] the handles were swabbed after the knife being used regularly for some time. In these conditions we could expect an accumulation of DNA, hence a higher yield.

#### *Novelty of the results.*

We set out three objectives to this study that we recall here:

- (1) to characterize the distribution of the quantity of DNA observed on the hands of individuals and transferred on surfaces either through primary or secondary transfer;
-

- (2) to assess if deconvolution of the DNA profiles is required to estimate the quantity of DNA of the POI and;
- (3) to test if the transfer proportion (quantity transferred on the surface over the initial quantity on the hand) is similar across individuals and can be used to predict the quantity of transferred DNA.

We were able to characterise for 6 individuals the distributions of the quantity of DNA observed on their hand and subsequently transferred on a knife handles either through primary contact or by a secondary mechanism. As already mentioned we have recorded very different quantities of DNA recovered on hands and on the knife handles after direct transfer for each participant and between participants. One person could then be judged as “good shedder” overall, but when considering a single experiment, that same person could be a very “poor shedder”. The shedder status, or for a better word the “shedding ability”, is better described by a distribution than by a single mean quantity. Our observations question the use, for a given individual, of a fixed label such as “good shedder” or “bad shedder”, irrespectively of time and circumstances. We propose alternatively to characterise a donor’s shedding ability by the parameters (mean and standard deviation) of the distribution of his/her quantities of DNA. Hence, when assessing the probability of observing a given quantity of DNA, for a given donor, that whole distribution should be accounted for and not only its mean (or a single shedder status label associated to it).

Regarding the second objective and the need to apply a deconvolution technique to mixed DNA profiles, we noted that, for each participant, both quantities (total DNA and POI’s DNA only) do not differ very much for primary transfers. It means that the total quantity of DNA can be used to study primary transfer without resorting to a mixture deconvolution process. However, in the experiments involving secondary transfers, we observed a marked difference between the total quantity of DNA and the quantity of DNA corresponding to the POI. It shows, as expected, that the total quantity of DNA left on the surface is dominated by the DNA coming from the handler. The POI’s DNA, who, in the secondary transfer scenario, did not touch the object but only the hand of the handler, is a minor contributor to the recovered mixed DNA profiles. Hence, the deconvolution is required when considering secondary transfers.

Regarding the third hypothesis postulating constant transfer proportions (TP) between donors, we have shown that TP may vary between participants and will depend on the type of the

---

transfer (primary versus secondary). It means that we cannot simply resort to a quantification of DNA on one hand to infer the shedder status and assess what will be transferred on a surface. Ideally, the measure of the distribution of the quantity of DNA should be carried out for a given person depositing on a given target surface following the alleged transfer mechanism.

We conclude in saying that in order to properly evaluate a given quantity of DNA considering different activities, the whole variation of DNA quantity should be accounted for. This can be done by using or measuring empirically the appropriate underpinning distribution that will be dependant on the donor, the substrate and the transfer mechanism.

### **Bibliography**

- [1] Willis S. et al., ENFSI guideline for evaluative reporting in forensic science, European Network of Forensic Science Institutes, Dublin (2015). <http://enfsi.eu/documents/forensic-guidelines/> (accessed April 28, 2018).
- [2] Lowe A., Murray C., Whitaker J., Tully G., Gill P., The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces, *Forensic Science International*. 129 (2002) 25–34.
- [3] Daly D.J., Murphy C., McDermott S.D., The transfer of touch DNA from hands to glass, fabric and wood, *Forensic Science International: Genetics*. 6 (2012) 41–46.
- [4] Phipps M., Petricevic S., The tendency of individuals to transfer DNA to handled items, *Forensic Science International*. 168 (2007) 162–168.
- [5] Goray M., Fowler S., Szkuta B., van Oorschot R.A.H., Shedder status—An analysis of self and non-self DNA in multiple handprints deposited by the same individuals over time, *Forensic Science International: Genetics* 23 (2016) 190–196.
- [6] Pesaresi M., Buscemi L., Alessandrini F., Cecati M., Tagliabracci A., Qualitative and quantitative analysis of DNA recovered from fingerprints, *International Congress Series*. 1239 (2003) 947–951.
- [7] Van Oorschot R.A.H., Phelan D., Furlong S., Scarfo G., Holding N., Cummins M., Are you collecting all the available DNA from touched objects?, *International Congress Series*. 1239 (2003) 803–807.
- [8] Bright J.-A., Petricevic S.F., Recovery of trace DNA and its application to DNA profiling of shoe insoles, *Forensic Science International*. 145 (2004) 7–12.
- [9] Lacerenza D., Aneli S., Omedei M., Gino S., Pasino S., Berchiolla P., Robino C., A molecular exploration of human DNA/RNA co-extracted from the palmar surface of the hands and fingers,

- Forensic Science International: Genetics. 22 (2016) 44–53.
- [10] Samie L., Hicks T., Castella C., Taroni F., Stabbing simulations and DNA transfer, *Forensic Science International: Genetics* 22 (2016) 73–80.
- [11] Meakin G.E., Butcher E.V., van Oorschot R.A.H., Morgan R.M., Trace DNA evidence dynamics: An investigation into the deposition and persistence of directly- and indirectly-transferred DNA on regularly-used knives, *Forensic Science International: Genetics* 29 (2017) 38–47.
- [12] Butcher E.V., van Oorschot R.A.H., Morgan R.M., Meakin G.E., Opportunistic crimes: Evaluation of DNA from regularly-used knives after a brief use by a different person, *Forensic Science International: Genetics* 42 (2019) 135–140.
- [13] Buckingham A.K, Harvey M.L., van Oorschot R.A.H., The origin of unknown source DNA from touched objects, *Forensic Science International: Genetics* 25 (2016) 26–33.
- [14] Szkuta B., Ballantyne K.N., van Oorschot R.A.H., Transfer and persistence of DNA on the hands and the influence of activities performed, *Forensic Science International: Genetics* 28 (2017) 10–20.
- [15] McColl D.L., Harvey M.L., van Oorschot R.A.H., DNA transfer by different parts of a hand, *Forensic Science International: Genetics Supplement Series* 6 (2017) e29–e31.
- [16] Kanokwongnuwut P., Kirkbride K.P., Kobus H., Linacre A., Enhancement of fingermarks and visualizing DNA, *Forensic Science International*. 300 (2019) 99–105.
- [17] Taylor D., Biedermann A., Samie L., Pun K.M. , Hicks T., Champod C., Helping to distinguish primary from secondary transfer events for trace DNA, *Forensic Science International: Genetics*. 28 (2017) 155–177.
- [18] Pinsky M. A. and Karlin S., *An Introduction to Stochastic Modeling*, Academic Press, 4th Edition (2011) chapter 1.

### Annexe III : Revue de la littérature centrée sur les traces ADN de contact

Une revue détaillée des variables intervenant dans les phénomènes de transfert, persistance, bruit de fond, collecte en laboratoire et contamination est présentée dans cette Annexe. Le scénario étudié dans ce projet de thèse faisant intervenir des traces de contact, la bibliographie est centrée sur ces dernières et non sur les traces issues de fluides biologiques (sang, salive, sperme, sécrétions vaginales par exemple).

Entre l'action d'intérêt et le prélèvement de la trace ADN, différents mécanismes interviennent qui peuvent mener au dépôt de la trace ADN puis à une perte plus ou moins importante de quantité d'ADN et au final donner des profils de plus ou moins bonne qualité. Ces mécanismes sont :

- le transfert d'ADN ou déposition de la trace,
- la persistance ou capacité de la trace à se maintenir (ou à subsister) sur le support,
- la présence de bruit de fond, qui est de l'ADN présent sur la surface prélevée autre que celui des personnes d'intérêt.
- la collecte et l'analyse de la trace en laboratoire.
- la pollution, qui est de l'ADN qui n'aurait pas dû se retrouver dans la trace. Elle est plus communément appelée « contamination »,

De nombreuses variables peuvent influencer sur ces mécanismes, donc sur la quantité d'ADN et la qualité des profils ADN observés et in fine sur le résultat de l'évaluation.

Ces variables ont fait l'objet de travaux précédents. Citons :

- Ka-Man Pun (2016) a réalisé une revue bibliographique de ces variables dans le cadre de sa thèse.
- Meakin et Jamieson ont écrit, en 2013, un article présentant l'ensemble des travaux relatifs au transfert et à la persistance d'ADN publiés jusqu'en 2013.
- Van Oorshot *et al.* (2018) ont par la suite publié à la fin de l'année 2018 une revue de la littérature sur ces variables.

La littérature présentée ici a servi de base pour déterminer les variables utilisées dans la conception du BN.

## 1- Le transfert

Les facteurs les plus étudiés à ce jour sont : le statut du donneur, le type de contact, le type de surface, les habitudes du donneur, son âge et son sexe.

**Statut du donneur :** le statut décrit la capacité d'un individu à déposer des traces de son ADN sur une surface, par contact, par transfert direct, sans l'intermédiaire d'un fluide biologique tel que le sang, la salive ou le sperme.

Selon certains auteurs, il est possible de distinguer un bon donneur d'un mauvais. Lowe *et al.* (2002) ont par exemple observé des différences importantes entre les profils ADN obtenus pour plusieurs personnes ayant touché à mains nues un tube en plastique 15 minutes après s'être lavé les mains. Un bon donneur est capable de transférer, juste après un lavage de main, une trace à partir de laquelle un profil ADN complet peut être établi.

A l'inverse, pour un mauvais donneur, un profil ADN complet ne pourra être obtenu que si celui-ci touche le tube en plastique au moins six heures après s'être lavé les mains. Dans cette même étude, une expérience a été réalisée avec deux couples chacun constitués d'un individu A, mauvais donneur, et d'un individu B, bon donneur. L'individu A sert la main de l'individu B avant de saisir un tube en plastique. Un prélèvement ADN est effectué sur le tube en plastique. Pour l'un des couples, le profil complet du teneur A est observé alors que pour l'autre couple, le profil du teneur A n'est pas observé ou partiellement. Une expérience similaire réalisée sur un échantillon de 9 individus, (Farmen *et al.*, 2008), donne des résultats similaires : un profil majoritaire correspondant à la personne B (bon donneur) est obtenu, alors que pour la personne A (mauvais donneur) ayant pris l'objet, seul un profil minoritaire complet ou partiel est observé, voire pas du tout.

Une étude réalisée par Daly *et al.* (2012) indique que le pourcentage de bons donneurs, « bon donneur » étant défini comme un individu pour qui un profil complet a pu être obtenu, est seulement de 13%. Ce pourcentage est de 22% si un « bon donneur » est défini comme un individu pour lequel un profil exploitable de 6 loci ou plus est obtenu.

Kanokwongnuwut *et al.* (2018) définissent, sur la base des observations faites dans leur étude, 3 catégories de donneurs. Le protocole de l'étude était le suivant : 11 personnes ont apposé leurs pouces durant 15 secondes sur une plaque de verre décontaminée. Pour chacune des personnes l'expérience a été réalisée immédiatement, 15 min, 60 min et 180 min après qu'elles se soient lavées les mains. En traçant la courbe de la quantité d'ADN en fonction du



temps, 3 catégories de donneurs ont été mises en évidence : les donneurs forts, les donneurs intermédiaires et les donneurs faibles.

Phipps et Petricevic (2007) montrent qu'il existe très peu de bons donneurs. De plus, les résultats qu'ils ont obtenus suggèrent que de nombreux facteurs interviennent dans le processus de transfert. Les auteurs pensent donc qu'il est plus difficile de qualifier des personnes comme « bons » ou « mauvais » donneurs que ne le laisse penser l'étude de Lowe *et al.*, 2002.

D'autres auteurs ont observé que la qualité du donneur varie de façon significative selon les individus (inter variabilité) mais aussi pour une même personne (intra variabilité) (Pesaresi *et al.*, 2001, Van Oorschot *et al.*, 2003 ; Bright et Petricevic, 2004). De manière générale, Il n'y aurait donc pas de classification entre donneurs, bons ou mauvais. Une personne peut être bon donneur à certains moments et mauvais donneur à d'autres.

Cependant, bien qu'il existe une certaine variabilité, c'est-à-dire qu'une même personne pouvait être bon donneur ou mauvais donneur à différents moments, Taylor *et al.* (2016) ont observé que deux individus pouvaient déposer en moyenne des quantités d'ADN très différentes : une personne peut donc être classée, en moyenne, comme bon ou mauvais donneur.

Goray *et al.* (2016) qui ont étudié la capacité de plusieurs individus à transférer leur ADN font également ce constat. L'expérience qu'ils ont menée est la suivante : dix participants, entre 20 et 40 ans dont 8 droitiers et 2 gauchers, ont apposé leurs deux mains sur des plaques de verre pendant 10 secondes. L'expérience est reconduite 3 fois par jour, sur 4 jours répartis sur une période de 4 à 6 semaines. Aucune instruction quant au lavage de main n'a été donnée aux participants. Leurs activités durant toute la période ont été documentées. Les auteurs ont étudié la quantité d'ADN et le nombre d'allèles correspondant au profil du donneur et ont observé une importante variabilité intra et inter individus. Cependant, sur les 4 jours, une tendance a pu être mise en évidence. Deux individus en particulier ont transféré de manière constante et significative plus d'ADN que les autres. Inversement, 2 autres individus ont transféré de manière constante et significative moins d'ADN que les autres. On retrouve alors moins de leur propre ADN et plus d'ADN étranger. Un des résultats intéressants est que pour 7 traces sur les 120 prélevées, le profil du donneur peut être exclu du profil de la trace. Seul de l'ADN étranger a été détecté. Quant au donneur, soit son ADN a été transféré mais en quantité si faible qu'il n'a pas pu être détecté, soit son ADN n'a pas été transféré. Les auteurs

précisent cependant que les profils de ces 7 traces étaient de mauvaise qualité avec des drops out et autres effets stochastiques.

La variabilité inter individu peut être expliqué par des facteurs comme le sexe, l'âge ou encore les habitudes du donneur, comme le montre les étapes suivantes :

- **Concernant le sexe du donneur**, Lacerenza *et al.* (2016) ont effectué des prélèvements ADN sur chacune des deux mains (paumes et doigts) de 60 individus (30 hommes et 30 femmes). Ils ont observé que la quantité d'ADN présente sur les mains était plus importante chez les hommes que chez les femmes. Ils ont également observé que les femmes avaient plus d'ADN étranger sur les mains que les hommes. Cependant, Bontadelli (2009) contredit ces résultats. Elle n'a observé aucune influence du sexe sur la quantité d'ADN présente sur les mains en étudiant les prélèvements effectués sur 30 femmes et 10 hommes âgés entre 21 et 24 ans.
- **Concernant l'âge du donneur**, une étude de Poetsch (2013) montre qu'il existe une corrélation entre la quantité d'ADN transféré, la qualité des profils provenant de traces digitales déposées et l'âge des individus. Cette étude constate aussi des différences dans les résultats selon qu'il s'agit d'un enfant ou d'un adulte. Ainsi, pour une population de 209 individus composée de 47 enfants de moins de 11 ans, 32 adolescents entre 11 et 20 ans, 81 adultes entre 20 et 60 ans et 49 personnes âgées de plus de 60 ans qui ont pressé leurs mains sur un support en plastique pendant 10 secondes, les résultats montrent qu'un profil complet est obtenu dans 75% des cas pour les enfants, 9% chez les adolescents, 25% chez les adultes et 8% chez les personnes âgées. Les résultats indiquent que la probabilité d'obtenir un profil ADN complet à partir de traces digitales décroît avec l'âge. Toutefois le résultat obtenu pour la tranche d'âge 11-20 ans pose question et aucune explication n'est avancée dans l'article.

A contrario, selon Raymond *et al.* (2008), l'âge n'influencerait pas la capacité à transférer son ADN. L'âge des volontaires n'est cependant pas précisé dans l'article.

D'autres facteurs pourraient expliquer la variabilité intra individus et contribuer également à augmenter la variabilité inter individus tels que les habitudes du donneur et ses propriétés biologiques.

- **Les habitudes du donneur** telles que le lavage des mains, la fréquence du toucher du visage ou des cheveux ou encore la pratique du sport sont des actions susceptibles d'avoir une influence sur le transfert d'ADN. Cependant l'influence de ces facteurs

n'a fait l'objet que de peu de recherches. Les expériences menées par Lacerenza *et al.* (2016) n'ont pas montré d'influence des habitudes personnelles et hygiéniques sur la quantité d'ADN présente sur les mains des donneurs. Seul le fait de se toucher les cheveux semble augmenter la quantité d'ADN récupérée.

Le lavage des mains est certainement le facteur le plus étudié à ce jour. Celui-ci augmente la probabilité de retrouver un profil simple car il enlève l'ADN étranger présent sur les mains (Pesaresi *et al.*, 2003). Cependant, une étude de Lowe *et al.* (2002), montre que plus le lavage est récent plus le profil obtenu peut être incomplet. Dans cette étude, sur les 8 individus testés, un seul profil complet est observé si le dépôt est effectué tout de suite après le lavage des mains. Si celui-ci a lieu 15 minutes après, on obtient alors quatre profils complets et 3 profils partiels.

Certaines études émettent l'hypothèse que la sueur faciliterait le transfert des cellules épithéliales (Wickenheiser, 2002 ; Bright et Petricevic, 2004). Van den Berge *et al.* (2016) ont d'ailleurs observé que le sébum et la sueur contribuent à l'augmentation de la quantité d'ADN présente sur les mains (ADN potentiellement transférable) par rapport à une absence de fluides. Afin d'étudier l'influence respective du sébum et de la sueur sur la quantité d'ADN présent sur les mains, deux prélèvements, l'un avant et l'autre après production de sueur ou sébum ont été effectués sur les 2 mains de respectivement de 5 et 16 individus. Les résultats montrent que la sueur a un effet moindre par rapport au sébum sur la quantité d'ADN présent sur les mains.

- Goray *et al.* (2016) quant à eux, ont observé que le moment de la journée et la dominance de la main ne semblent pas être des facteurs ayant un effet significatif sur les quantités d'ADN.

Outre le statut du donneur lui-même, ce qui influence le résultat ADN est le statut relatif entre les personnes impliquées. Par exemple, dans une première situation, un suspect dit « bon donneur » serre la main à un agresseur dit « mauvais donneur ». A l'inverse, dans une deuxième situation, un suspect dit « mauvais donneur » serre la main d'un agresseur dit « bon donneur ». Les deux situations ne donneront pas un type de profil similaire pour ce qui est des proportions dans le mélange d'ADN, comme en témoigne l'étude de Fonnelløp *et al.* (2017). Ces auteurs ont étudié cette problématique à partir d'un scénario fictif d'agression. Cette expérience a porté sur 20 volontaires.

Dans un premier temps, les bons et mauvais donneurs ont été identifiés. Pour cela, chacun des volontaires a serré un tube en plastique avec la main dominante pendant 10 secondes, 1 fois par jour sur 3 jours. Ont été considérés comme des bons donneurs les volontaires dont au moins deux profils sur les trois réalisés contenaient au moins 12 loci et une quantité supérieure à la quantité moyenne d'ADN déposée par l'ensemble des volontaires (quantité moyenne des 60 prélèvements). Les autres sont considérés comme de mauvais donneurs.

Dans un deuxième temps, les volontaires ont porté pendant 3h un tee-shirt préalablement décontaminé. A l'issue de cette période, le bruit de fond ADN était composé uniquement de celui du porteur du tee-shirt. Le test consistait à reproduire le scénario d'agression avec des couples composés différemment. Quatre types de couples ont été étudiés : bon donneur victime avec bon donneur attaquant - bon donneur victime avec mauvais donneur attaquant - mauvais donneur victime avec bon donneur attaquant et mauvais donneur victime avec mauvais donneur attaquant.

Pour simuler l'agression, l'attaquant attrape les épaules de la victime par derrière tandis que celle-ci doit essayer de se dégager vers l'avant. L'attaque dure 20s et est filmée. Chaque couple répète l'attaque 5 fois sauf le couple de bons donneurs, victime et attaquant, qui ont reproduit l'attaque seulement 2 fois. Quatre prélèvements sont effectués : 3 sur le tee-shirt de la victime (deux au niveau des épaules et un au milieu du dos) et 1 sur le tee-shirt de l'attaquant (au milieu du devant). Les prélèvements sont effectués seulement si le contact est bien direct (attaque filmée).

Sur les prélèvements effectués au niveau des épaules du tee-shirt de la victime, l'ADN de l'attaquant est détecté pour 16 attaques sur 17 sur au moins 1 échantillon sur 3. L'ADN de l'attaquant n'est pas détecté sur la 17ème attaque. Pour cette attaque, l'attaquant était mauvais donneur. En effet, lors du premier test utilisé pour définir son statut de donneur, l'ADN d'une collègue a été observé sur le prélèvement du tube qu'il avait serré dans sa main.

Si l'attaquant est bon donneur, son ADN est détecté sur tous les prélèvements effectués sur le tee-shirt, généralement en majoritaire mais aucun chiffre n'est fourni dans l'article. Si l'attaquant est un mauvais donneur, son ADN est détecté sur 15 prélèvements sur 26 et il n'y a généralement pas de « clairs » profils majoritaires, Cependant les auteurs n'expliquent pas ce qu'ils entendent par profils majoritaire « clair ».

Si la victime est un bon donneur, son profil est généralement détecté en majoritaire. Ici encore, les auteurs ne précisent pas de chiffre. Si, à l'inverse, la victime est un mauvais donneur, son profil est généralement détecté en minoritaire ou il n'y a pas de profil majoritaire clairement défini.

Les auteurs concluent qu'il y a un effet du statut de donneur de l'attaquant sur le transfert de son ADN sur le tee-shirt de la victime mais qu'il n'y a pas d'effet du statut de la victime sur le transfert de l'ADN de l'attaquant.

Concernant le prélèvement effectué sur le tee-shirt de l'attaquant, l'ADN de la victime est observé pour 6 prélèvements sur 17. Les auteurs concluent qu'il n'y a pas d'effet du statut de la victime sur le transfert de son ADN.

Un autre aspect à mettre en relation avec le statut de donneur, est la présence d'ADN « étranger » sur les mains du donneur. La capacité du donneur à transférer l'ADN « étranger » dépend alors de son statut (Szkuta *et al.*, 2017). L'ADN « étranger » est à différencier du bruit de fond qui est défini comme de l'ADN autre que celui du suspect et qui n'est pas en relation avec l'activité précise en investigation. Par conséquent, celui-ci ne peut pas être en relation avec les personnes impliquées.

**En conclusion**, le statut du donneur est une variable importante à prendre en compte. Elle ne peut cependant pas être définie suivant des catégories telles que « Bon/Mauvais donneur ». Ce statut dépend des caractéristiques propres à l'individu (âge, sexe, habitudes) ainsi que d'évènements ponctuels tels que le lavage des mains ou la transpiration qui influencent la capacité de transférer de l'ADN à un moment précis.

**Type de contact** : Trois types de contact sont définis dans la littérature :

- Contact par pression (toucher en appuyant),
- Contact passif (simple toucher),
- Contact par friction (toucher avec un mouvement de frottement).

Un contact par friction augmente la probabilité qu'un transfert secondaire se produise par rapport aux autres types de contact (Goray *et al.*, 2010).

La capacité de transfert secondaire double entre un contact passif et un contact par pression avec le vecteur secondaire. Dans le cas d'un contact passif, environ 6% de la quantité d'ADN

déposée sur la 1<sup>ère</sup> surface est retrouvée sur la 2<sup>ème</sup> alors que cette quantité est de 11% dans le cas d'un contact par pression.

Selon Van Oorschot (1997), le temps de contact n'aurait pas d'influence sur la quantité d'ADN déposée. L'expérience portait sur des temps de contact de 5 secondes, 30 secondes, 3 minutes et 10 minutes et les objets utilisés étaient des tubes en polypropylène. Une autre étude utilisant des câbles métalliques montre qu'un contact même bref suffit avec l'objet suffit pour transférer son ADN (Saravo *et al.*, 2004).

Kamphausen *et al.* (2015), ont étudié un cas de transfert d'ADN entre des vêtements lavés soit à la main, soit à la machine à laver. Quinze pulls, dit 'pulls donneurs', sur lesquels de l'ADN d'un individu avait été déposé au niveau du col, ont été utilisés. Chaque pull a été mis dans une machine à laver avec un deuxième pull, dit 'pull receveur', identique mais sans ADN. Sur les 15 tests réalisés, aucun profil ADN n'a été observé aussi bien sur le pull donneur que sur le pull receveur.

La même expérience a été réalisée avec un lavage à la main. Cette fois, sur le pull donneur, un profil partiel a été observé 11 fois et aucun profil ADN pour les 4 autres fois.

**Type de surface :** Les surfaces réceptrices sont de nature très diverse. En général, celles-ci sont classées en deux catégories, les surfaces poreuses et les surfaces non poreuses (lisses). Mais tous les objets ont des propriétés de surface spécifiques qui dépendent notamment des matériaux qui les composent et de leur forme.

Pour de nombreux auteurs, le type de surface influence la quantité d'ADN « de contact » transférée et la qualité des profils obtenus.

Ainsi les expériences réalisées par Daly *et al.* (2012) montrent que la quantité d'ADN transférée est plus importante sur des surfaces en bois que sur des surfaces en tissus (vêtements par exemple) ou en verre. De plus, dans les cas de transfert secondaire, la quantité d'ADN transférée est plus importante si la surface intermédiaire est non poreuse et la surface à prélever poreuse (Goray *et al.*, 2010).

Fonneløp A. *et al.* (2015) ont étudié la possibilité de transfert tertiaire, à partir d'un cas réel, le meurtre de Meredith Kercher. Dans cette affaire, l'ADN correspondant à un suspect a été prélevé sur l'attache d'un soutien-gorge retrouvé au sol. La défense a soutenu qu'un enquêteur avait transféré l'ADN du suspect, qui se trouvait sur la poignée d'une porte, sur le soutien-gorge *via* ses gants. Les auteurs ont cherché à vérifier si un tel transfert tertiaire était

possible et si oui, dans quelles conditions. Différents types de surfaces ont été testés : bois, tube en plastique, poignée de porte en métal pour la 1<sup>ère</sup> surface de contact, gants en nitrile pour la 2<sup>ème</sup> surface et tissus ou papier pour la 3<sup>ème</sup> surface. Tous les objets ont été préalablement décontaminés et un prélèvement de contrôle a été effectué afin de vérifier l'absence de profils ADN sur les surfaces avant que l'expérience ne commence. L'expérience est la suivante : un suspect, considéré comme bon donneur, touche un objet en bois, en plastique ou en métal (1<sup>ère</sup> surface). Celui-ci est saisi par un second individu, en l'occurrence l'enquêteur, bon donneur également, qui porte des gants en nitrile (2<sup>ème</sup> surface). Ce dernier touche ensuite une surface en tissu ou en papier (3<sup>ème</sup> surface). Les résultats indiquent que le type de surface a une influence sur le transfert. Ainsi, la quantité d'ADN retrouvée sur les gants est plus importante quand le transfert se fait à partir de la poignée de porte en métal qu'à partir du tube en plastique ou de l'objet en bois. Pour ces deux derniers matériaux, il n'y a pas de différence significative de quantité d'ADN transféré. Par contre, l'étude ne montre pas de différence de quantité d'ADN transféré des gants sur la troisième surface, que celle-ci soit en tissu ou en papier.

Cependant, d'autres auteurs, comme Alessandri *et al.* (2003), jugent ce facteur non significatif. Cale *et al.* (2015) qui ont étudié l'influence de ce facteur sur la possibilité de transfert secondaire arrivent à la même conclusion. L'expérience qu'ils ont menée est la suivante. Vingt-quatre couteaux identiques ont été utilisés dont la moitié avec un manche en plastique lisse et l'autre moitié avec un manche en plastique rugueux. Six couples de volontaires ont été formés de façon à ce qu'au sein de chaque couple, le profil ADN des deux partenaires soit le plus différenciable possible. Afin de favoriser le transfert d'ADN, tous ont porté des gants durant 1h30 avant le début de l'expérience. Tout de suite après avoir enlevé leurs gants, les deux partenaires de chaque couple se sont serrés la main pendant 2 minutes. Les auteurs reconnaissent que pour une poignée de main ce temps est long mais qu'il est raisonnable pour simuler d'autres formes de contacts intimes pouvant durer plus de 2 minutes. Puis, tous les volontaires ont chacun saisi un couteau avec un manche en plastique lisse pendant 2 minutes. L'expérience a été renouvelée à l'identique avec le couteau possédant un manche en plastique rugueux. Sur 4 couteaux, aucun profil ADN n'a été détecté. Pour les vingt autres, un profil ADN a été détecté. Pour ceux-ci, le profil du partenaire est observé dans 85% des cas, soit 17 cas. Pour 5 cas, celui-ci est le seul contributeur ou le contributeur majoritaire. Aucune différence significative dans les quantités prélevées n'a été observée entre les couteaux à manche lisse et ceux à manche rugueux. La différence des résultats

obtenus, en termes de qualité des profils, avec les différents types de manche n'a pas été détaillée dans les résultats de l'étude.

## 2- La persistance

La persistance est la capacité de la trace à se maintenir sur le support. Elle intervient entre le transfert et la collecte de la trace. Les principaux facteurs pouvant influencer la persistance de l'ADN « de contact » sur une surface sont : le temps écoulé entre le dépôt et le prélèvement, l'environnement et le type de la surface. Ces trois facteurs sont étudiés dans la littérature.

**Temps :** Plus le temps écoulé entre le dépôt de la trace et le prélèvement est long, plus la persistance des traces ADN diminue. L'ADN se dégrade et la qualité des profils obtenus est moins bonne (Raymond *et al.*, 2009).

**Environnement :** Les conditions environnementales (humidité, température, exposition aux UV par exemple) ont une incidence sur la persistance. Les résultats d'une expérience menée par Raymond *et al.* (2009) montrent qu'aucun profil ADN n'a pu être retrouvé deux semaines après le dépôt dans un milieu ouvert (en extérieur). A l'inverse, un profil ADN peut être observé jusqu'à six semaines si la trace est déposée en milieu fermé ou protégé (en intérieur).

Breathnach (2016) a étudié l'influence des conditions de stockage mises en place par son laboratoire sur la persistance d'ADN (*i.e.* quantité d'ADN et de qualité des profils). L'expérience est la suivante : 10 hommes ont porté pendant 12h un sous vêtement neuf, tout juste sorti de son emballage avant d'être stocké. Des zones spécifiques de prélèvement ont été délimitées sur la bande élastique du sous-vêtement en fonction du temps de stockage : une zone à prélever juste avant le stockage du sous vêtement, une zone à prélever après 6h de stockage puis une zone à prélever après 12 heures de stockage. Entre chaque prélèvement, les vêtements sont stockés en respectant les consignes du laboratoire. Les prélèvements effectués au bout de 6 heures et 12 heures de stockages n'ont pas révélé de différence significative par rapport au prélèvement de référence quant à la quantité d'ADN et la qualité des profils obtenus à partir des traces.

**Type de surface :** De nombreuses études se sont intéressées à la persistance de l'ADN sur différents types de support.

Une expérience réalisée par Goray *et al.* (2010) montre que la persistance est plus importante sur des surfaces en plastique plutôt qu'en coton. Pour une surface en coton, la quantité d'ADN



diminue de 50% lorsque le prélèvement est effectué 24 heures après le dépôt alors qu'elle reste identique pour une surface en plastique.

La persistance de l'ADN après lavage a également été étudiée sur les vêtements. Après un lavage à 40 degrés, la quantité d'ADN retrouvée diminue, tout comme la qualité des profils (Sczech, 2011).

Différentes études sur la persistance du profil d'une personne ayant utilisé régulièrement un objet (1<sup>er</sup> utilisateur, U1) avant que d'autres ne le fassent (nouvel utilisateur, U2) ont été réalisées. L'étude de Van Oorschot *et al.* (2014) met en évidence une diminution progressive au cours du temps, de la contribution d'U1 à la trace par rapport à celle des autres utilisateurs. Elle montre aussi que la diminution dépend du type de surface : un support poreux conservera plus longtemps le profil d'U1 qu'un support non poreux. Pour une surface poreuse, un prélèvement effectué 10 heures après l'échange d'objets, donne un profil d'U1 majoritaire. Si celui-ci est effectué 96 heures après l'échange, c'est le profil U2 qui devient majoritaire. Sur une surface non poreuse, la contribution des deux utilisateurs est de 50-50 si le prélèvement a lieu sitôt après qu'U2 se soit saisi de l'objet. Si celui-ci est effectué 96 heures après, le profil d'U2 est alors majoritaire à 88%. Pfeiffer et Wiegand (2017) observent aussi que la persistance de l'ADN dépend du type de surface.

Oldoni *et al.* (2016) ont étudié l'incidence des matériaux et du type d'objet sur la persistance des traces ADN d'un utilisateur régulier de différents types d'objets, supposé volés: plastique (souris d'ordinateur, stylo, montre et collier), tissu (une coiffe d'infirmière) et nitrile (gants). En utilisant un protocole décrit par Oldoni *et al.* (2015). L'expérience a été menée comme suit. Sept volontaires, « premiers utilisateurs », ont chacun manipulé régulièrement un ou deux bracelets ou montres durant 8 à 10 jours, puis sept autres volontaires, « deuxièmes utilisateurs, » ont pris le relais et utilisé ces objets pendant 5, 30 ou 120 minutes. Huit paires d'individus (« premiers utilisateurs », « deuxièmes utilisateurs ») ont été sélectionnés pour manipuler un ou deux colliers, neuf paires pour manipuler une ou deux souris d'ordinateurs, stylos ou coiffes et enfin onze paires d'individus pour manipuler une ou deux clés. Des prélèvements ont été effectués sur les objets, avec un total de 234 prélèvements. Les auteurs ont observé que le profil ADN des deuxièmes utilisateurs était majoritaire dans la plupart des cas après 120 minutes d'utilisation de l'objet.

Ils ont également observé une augmentation du pourcentage de la contribution du profil du second utilisateur (en terme de quantité relative, basée sur la hauteur des pics) avec la durée

---

d'utilisation. Après 5 minutes d'utilisation, le profil du second utilisateur représente 21% du mélange. Après 120 minutes, il représente 73% du mélange.

Les auteurs ont pu aussi observer des cas de transfert indirect. Des allèles de source inconnue (2 à 10% de l'ADN totale) ont été détectés pour toutes les traces. Pour 7 cas sur les 230 réalisés, un profil complet a été identifié. Pour cinq cas, celui-ci correspondait au profil d'un volontaire ayant participé à l'étude. Pour les deux autres cas, il s'agissait d'une personne du laboratoire n'ayant pas participé aux simulations.

Les auteurs avaient réalisés une étude similaire en 2015.

Fonnelløp et al. (2015) ont étudié la persistance de l'ADN d'un ancien utilisateur d'un objet sur les mains d'un nouvel utilisateur, la surface étudiée ici étant la main. Quatre personnes ayant régulièrement utilisé la souris et le clavier de leur ordinateur depuis au moins 8 mois ont été sollicitées. Au début de l'expérience, celles-ci ont échangé deux à deux leurs souris et leurs claviers, formant ainsi deux couples. Après l'échange, des prélèvements ont été effectués chaque jour pendant 42 jours sur les mains de chaque utilisateur. L'objectif était de voir s'il était possible de retrouver sur celles-ci des traces ADN du précédent utilisateur du matériel informatique et si oui pendant combien de temps.

Pour un des couples, l'ADN du premier utilisateur du matériel a pu être détecté sur les mains de son partenaire pendant huit jours. Tous les allèles ont été observés. Pour le second partenaire, aucun ADN du précédent utilisateur n'a été observé, et ceci, dès le 1<sup>er</sup> jour.

Pour l'autre couple, un profil présentant 70% de concordance avec celui du premier utilisateur a été observé sur les mains du partenaire mais seulement pour le prélèvement effectué le premier jour. Après le premier jour, aucune correspondance n'a été constatée. Pour le deuxième partenaire, l'ancien utilisateur ne peut pas être exclu comme contributeur du profil obtenu jusqu'au huitième jour. A partir du neuvième jour, son profil n'est plus observé.

Une étude de persistance a également été réalisée sur des couteaux par Meakin *et al.* en 2015. Quatre couples de volontaires formés d'un donneur A et d'un donneur B ont été constitués. Le scénario est le suivant : le donneur A utilise un couteau pendant 1 minute, 2 fois par jour pendant 2 jours. Le 3<sup>ème</sup> jour, il donne une poignée de main d'une durée de 10 secondes au donneur B. Immédiatement après cette poignée de main avec B, il utilise le couteau pour poignarder un appareil en mousse pendant 1 minute. Des prélèvements sont ensuite effectués sur le manche du couteau 1 heure, 1 jour et 1 semaine après l'action. Les prélèvements effectués une heure après l'action montrent, pour trois des couples, un mélange de 3 personnes constitué des profils du donneur A en majoritaire, du donneur B et d'un inconnu.

---

Les prélèvements effectués une heure et une semaine après montrent que le profil du donneur B est toujours observé dans la portion minoritaire en proportion plus réduite par rapport au prélèvement précédent. Pour le 4<sup>ème</sup> couple, le profil simple et complet du donneur A est observé lorsque le prélèvement est effectué 1 heure, 1 jour ou 1 semaine après l'action.

Michelle Breathnach (2016) a réalisé une étude dont une partie portait également sur la persistance de l'ADN d'une personne sur son sous-vêtement après qu'une autre personne ait touché ce sous-vêtement. L'expérience s'est déroulée de la façon suivante : des prélèvements ont été effectués sur la bande élastique des sous-vêtements masculins de 63 hommes. Ces sous-vêtements propres au début de l'expérience ont été portés pendant 12 heures avant d'être emballés dans un sachet plastique. Ensuite, des femmes volontaires, à qui on a demandé de se laver les mains au moins une heure avant l'expérience, ont chacune sorti un sous-vêtement du sachet et l'ont tenu par la bande élastique pendant 15 secondes. Une fraction majoritaire exploitable correspondant au profil ADN du porteur du sous-vêtement est observée pour 52% des cas environ. Le profil ADN de la personne ayant touché le sous-vêtement est observé dans la fraction majeure exploitable pour 9% des cas et dans la fraction mineure exploitable dans 1% des cas. La quantité moyenne d'ADN du porteur du sous-vêtement obtenue lorsque son profil est exploitable est de 14.37 ng. Elle est de 5.73 ng concernant l'ADN de la personne ayant touché le sous vêtement.

La persistance de l'ADN sous les ongles a aussi été étudiée. Une expérience réalisée à partir de prélèvements d'ADN retrouvé sous les ongles d'une personne ayant griffé une victime, montre que l'ADN est exploitable (plus de 5 loci détectés) dans 37% des cas si le prélèvement est effectué tout de suite après les faits. Si le prélèvement a lieu 6 heures après les faits, l'ADN n'est exploitable que dans 7% des cas (Matte et al., 2012).

### **3- Le bruit de fond**

Le bruit de fond (ou « contamination ») est défini comme de l'ADN autre que celui du suspect et qui n'est pas en relation avec l'activité précise en investigation. Quelle est la proportion du bruit de fond par rapport au profil de la personne ayant fait l'action reprochée ou les acteurs impliqués ? Généralement le bruit de fond est associé à la fraction minoritaire de la trace, constituée d'un ou plusieurs contributeurs. Mais si le profil du bruit de fond correspondant à une tierce personne (personne étrangère à la scène) est majoritaire et couvre celui de la

personne ayant fait l'action d'intérêt (qui se retrouve dans la fraction minoritaire), le risque est de considérer cette tierce personne comme impliquée dans les actions en enquête.

Des études sur cette thématique ont été publiées récemment. Lehmann *et al.* (2015) ont notamment étudié l'influence du bruit de fond sur la capacité à détecter le profil d'une personne cible. Cinq objets en verre et cinq en coton ont été utilisés pour l'expérience. Un ou plusieurs volontaires ont frotté leurs mains pendant 30 secondes sur tous les objets. Ensuite, une personne cible, autre que le/les volontaire(s) ayant déposé le bruit de fond, s'est saisie successivement des cinq objets. Si le bruit de fond initial est constitué de l'ADN d'une seule personne, aucune influence significative de la présence d'ADN de contact en bruit de fond sur la capacité à détecter le profil cible n'est observé, quel que soit l'objet. En revanche, plus le nombre de contributeurs au bruit de fond augmente, plus le profil cible est difficile à détecter. Sur le 1<sup>er</sup> objet saisi, le profil cible est majoritaire, que celui-ci soit en verre ou en coton. Concernant les objets en verre, un mélange 50/50 est obtenu pour les 2<sup>ème</sup> et 3<sup>ème</sup> objets saisis tandis que le profil cible n'est pas observé sur le 5<sup>ème</sup> et dernier objet saisi. Concernant les objets en coton, le profil cible est majoritaire sur le 2<sup>ème</sup> objet puis est observé en minoritaire ou pas du tout sur les autres objets.

Dans l'étude de Fonnøløp *et al.* (2015), 4 claviers d'ordinateurs ont été utilisés régulièrement par une même personne pendant au moins 8 mois. Des prélèvements sur les touches de claviers et entre les touches ont ensuite été effectués. Pour un des claviers, un profil de mélange comprenant le profil d'un contributeur inconnu en plus de celui de l'utilisateur régulier a été observé sur les prélèvements. Pour un autre clavier, aucun profil ADN n'est observé. Enfin pour les 2 autres claviers, de l'ADN de l'utilisateur régulier est observé entre les touches mais les prélèvements effectués sur les touches se sont révélés négatifs.

Dans une de leur étude, Van den Berge *et al.* (2016) ont effectué des prélèvements sur 1a surface de 105 objets publics et de 164 objets privés. Concernant les objets publics, un profil majoritaire a pu être déduit pour 17% des objets. Concernant les objets privés, un profil majoritaire correspondant au profil de l'utilisateur régulier de l'objet est observé dans la plupart des cas.

Plusieurs études sur le bruit de fond de l'ADN présent sous les ongles ont été réalisées afin d'avoir des données sur l'exploitabilité des traces ADN en bruit de fond. En 2007, Cook et Dixon, ont réalisé des prélèvements d'ADN sous les ongles sur 200 volontaires. Environ 13% des prélèvements ont donné du bruit de fond. Sur ces 13%, 6% sont des profils ADN exploitables (plus que 8 allèles observés). Pour ces 6%, les auteurs peuvent donc reportés les

profils dans une base de données alors qu'il s'agit de bruit de fond. Pour ces cas-là, les volontaires ont déclaré avoir eu des contacts physiques peu avant le prélèvement. Les auteurs de l'article avancent l'hypothèse que la correspondance entre le profil d'un suspect et le profil ADN retrouvé sous les ongles de la « victime » ne peut s'expliquer que par un contact intime ou physique ayant eu lieu entre les deux personnes, suspect et victime.

Dans une autre étude (Malsom *et al.*, 2009), un plus grand nombre de cas (17%) donnait un profil exploitable (plus que 8 allèles observés) c'est-à-dire un profil que les auteurs pouvaient reporter. L'hypothèse avancée dans cette étude est que les volontaires vivaient avec leurs partenaires, ce qui expliquerait ce plus grand pourcentage. Habiter avec quelqu'un augmenterait donc la probabilité d'avoir du bruit de fond exploitable et de retrouver l'ADN du partenaire.

Matte *et al.* (2012) ont étudié l'ADN présent sous les ongles des deux mains d'un échantillon de population constitué de 89 personnes. Un même écouvillon est utilisé pour collecter l'ADN sous tous les ongles d'une main, ce qui fait 178 prélèvements : 19% des prélèvements contenaient du bruit de fond dont 35% seulement étaient exploitables (5 loci ou plus).

Des études sur le bruit de fond dans les cas de strangulation ont également été effectuées (Graham et Rutty, 2008 ; Rutty, 2002). Celles-ci montrent que la présence d'ADN autre que celui de la victime peut être détectée sur le cou sans que celui de l'agresseur ne soit systématiquement retrouvé. Il est donc possible de retrouver un profil ADN autre que celui qui a étranglé la victime surtout si celle-ci vivait avec quelqu'un.

Concernant les données sur des habits, Breathnach *et al.* (2016) ont étudié la présence de bruit de fond sur des sous-vêtements de 63 individus de sexe masculin. Les sous-vêtements ont été portés pendant 12 heures avant que les prélèvements sur la bande élastique du sous-vêtement ne soient effectués. Un bruit de fond est observé dans 87% des cas, soit sur 55 sous-vêtements. Pour ces sous-vêtements, le bruit de fond est exploitable seulement dans 15 % des cas, soit sur 9 sous-vêtements. Pour ces derniers, le profil du bruit de fond est observé dans la fraction majoritaire dans 1 cas et dans la fraction minoritaire dans 8 cas.

Une étude s'est également intéressée au bruit de fond présent sur des vêtements portés habituellement par une personne. Un prélèvement ADN a été effectué sur 11 vêtements portés par différents individus. Dans trois cas, un profil simple correspondant au profil du porteur régulier est obtenu. Dans 8 cas, des profils de mélange sont obtenus avec, en profil majoritaire, celui du porteur régulier. Dans 6 de ces cas, le profil minoritaire correspond au

partenaire du porteur régulier. Concernant les 2 autres cas, le profil minoritaire ne correspond pas au profil des collègues ou de la famille du porteur (Stouder *et al.*, 2001).

Meakin *et al.* (2017) ont étudié la persistance du bruit de fond sur un manche de couteau. Pour cette expérience, quatre volontaires ont été choisis pour obtenir un bruit de fond sur les manches des couteaux. Deux fois par jour et pendant deux jours, chaque volontaire a tenu le manche du même couteau pendant 15 secondes puis l'a frotté pendant 15 secondes avec chaque main. Puis le couteau a ensuite été utilisé par une autre personne. L'ADN du volontaire ayant créé le bruit de fond était toujours détecté après une semaine mais la quantité diminuait au cours du temps.

Afin d'avoir une idée du bruit de fond, spécifique à chaque cas, une solution est d'effectuer les prélèvements sur les zones où il existe une forte probabilité qu'il n'y ait pas de dépôt d'ADN d'intérêt mais une contamination. Par exemple, si une corde est impliquée, suivant la position dans laquelle elle a été retrouvée et l'action d'intérêt, elle peut présenter des zones non pertinentes par rapport aux actions alléguées, comme, par exemple, les parties de cordes qui n'ont pas été en contact avec la ou les personnes concernées par l'action en question. Cependant, les résultats de l'analyse effectuée sur de telles zones, quels qu'ils soient, n'enlèvent rien à la valeur des résultats obtenus sur les traces pertinentes, prélevées, par exemple, sur des zones de la corde où il est possible de retrouver de l'ADN d'intérêt (Coquoz *et al.*, 2013).

L'étude de la quantité/qualité relative entre les différents contributeurs (contributeur majeur/contributeur minoritaire) peut permettre dans certains cas de faire la différence entre du bruit de fond et des profils d'intérêt. Récemment, le transfert d'ADN a été étudié lors des interactions sociales entre trois individus allant boire un verre dans un bar pendant 20 minutes. Les prélèvements ADN ont été effectués sur les verres et les chaises. L'expérience a été répétée quatre fois avec des individus différents. Pour ces quatre expériences, les résultats montrent que le profil du contributeur majoritaire est celui de la dernière personne ayant touché la chaise ou le verre. Le profil minoritaire est un profil d'une personne n'ayant pas directement été en contact avec l'objet. Il s'agit donc de profils d'ADN transférés indirectement (Goray et Van Oorschot, 2013 ; Goray et Van Oorschot, 2015).

Le bruit de fond a donc une influence sur la quantité d'ADN prélevé : plus il y a de bruit de fond, plus la quantité d'ADN prélevé sera grande. Il intervient aussi sur la qualité du profil de la trace : plus le bruit de fond est important, plus le nombre de contributeurs peut être élevé et

---

plus le profil de la trace ADN est susceptible d'être de moins bonne qualité. De plus, plus la quantité de bruit de fond est élevée, plus celui-ci aura tendance à masquer le profil d'intérêt.

Sur la surface, il y a alors du bruit de fond ainsi que de l'ADN des personnes impliquées qui s'est transféré et qui a persisté. La surface est ensuite prélevée et la trace analysée.

#### **4- La collecte et l'analyse de la trace en laboratoire**

##### **Outils de collecte et d'analyse.**

Pour une même trace ADN, la qualité des profils dépend des protocoles et des outils utilisés : kits de prélèvements, taille de la surface prélevée, kits d'extraction, d'amplification et systèmes d'électrophorèses capillaires. Deux études (Cook et Dixon, 2007 et Malsom *et al.*, 2009) sur le bruit de fond présent sous les ongles d'individus montrent des résultats différents qui s'expliquent par l'utilisation de techniques de prélèvements différentes (Meakin et Jamieson, 2013).

Dans l'étude de Lacerenza *et al.* (2016) décrite précédemment, concernant le prélèvement d'ADN sur les mains, les auteurs ont observé des différences significatives de la quantité d'ADN prélevé selon la méthode de collecte utilisée. Ainsi, pour des prélèvements effectués sur des mains, l'utilisation d'un coton, (« swab ») humide permet de recueillir une plus grande quantité d'ADN que l'utilisation d'un « swab » sec ou d'une bande collante (« tape lift »). Dans l'étude de Noël *et al.* (2016), les auteurs ont remarqué que lorsqu'ils prélevaient l'ADN sur des sous-vêtements en coton avec un « swab », aucun profil n'était observable, contrairement à des prélèvements effectués en coupant le sous-vêtement.

De même, certaines méthodes d'extraction sont plus performantes que d'autres et permettent d'obtenir de plus grandes quantités d'ADN et de limiter les pertes d'ADN. Cet effet ne doit pas être ignoré (Verdon, 2014).

Dans leurs études, Fonnèløp A. *et al.* (2015) obtiennent, pour des traces ayant de faibles quantités d'ADN, un pourcentage de profil ADN plus élevé que celui habituellement donné dans la littérature. D'après les auteurs, cette divergence pourrait s'expliquer par l'utilisation d'un kit d'amplification différent et d'un système d'électrophorèse capillaire plus sensible.

Concernant l'ensemble des études effectuées en laboratoire, les conditions étaient contrôlées et connues : l'instant du lavage des mains avant de toucher une surface, l'intervalle de temps entre le serrage de mains et le transfert sur la surface d'intérêt, le statut des donneurs ainsi que le type des surfaces intermédiaires. Ces expériences contrôlées sont utiles afin de comprendre

les processus de transfert, persistance et de bruit de fond et surtout l'incidence des variables qui interviennent dans ceux-ci. Cependant, dans une situation réelle, *in situ*, tous les paramètres ne sont pas connus ou peuvent être différents de ceux considérés dans les expériences. Les résultats en termes de quantité d'ADN prélevé et/ou de qualité peuvent ainsi différer de ceux donnés par la littérature et obtenus en conditions contrôlées.

## 5- La pollution

La pollution se manifeste par la présence d'ADN qui n'aurait pas dû se retrouver dans la trace. Elle est communément appelée « contamination », à ne pas confondre avec la contamination définie précédemment. Cet ADN est transféré lorsque les protocoles de prélèvements et de manipulations ne sont pas respectés ou lorsque les mesures de précaution préconisées sont insuffisantes. Les pollutions sont possibles à tout moment entre l'instant où l'action d'intérêt a eu lieu et les traces pertinentes déposées jusqu'à l'obtention du profil.

Fonneløp *et al.* (2016) ont rapportés, entre 2009 et 2015, 16 cas de pollution par la police sur 3545 cas étudiés. Au département de biologie forensique du Norwegian Institute of Public Health, Department of forensic biology (NIPH), 9 cas ont été reportés sur les 10590 traités. Basset et Castella (2018) ont quant à eux reportés 709 cas de pollution entre 2011 et 2015, ce qui représente une moyenne de 11,5 cas par an par 1000 profils envoyées dans la base de données ADN suisse dont 85% résulte de pollution par la police, 11% des employées du laboratoire et 3% d'autres sources. De nouvelles procédures au laboratoire ont ensuite été mises en place et entre 2015 et 2017, 110 cas de pollution sur 11 852 profils envoyées dans la base de données ADN suisse ont été détectés dont seulement 3% provenait des employées du laboratoire (Basset et Castella, 2019). Lapointe et al. (2015) ont observé 58 cas de pollution par la police du Québec lors de la comparaison entre 327 profils ADN de policiers et la base de données ADN canadienne en juillet 2014 qui contenait alors 31 071 profils ADN. De nombreuses causes peuvent être à l'origine de pollutions (Gill, 1997; Gill et Kirkham, 2004; Coquoz et al., 2013) telles que :

- Le non-respect des consignes de précaution par les intervenants sur la scène de crime et en laboratoire (Lee *et al.*, 1998; Lee et Ladd, 2001; Rutty *et al.*, 2000 ; Scherzinger *et al.*, 1999). Par exemple si le masque couvrant le nez et la bouche n'est pas mis, il a été démontré que des profils ADN complets de personnes ayant parlé ou éternué à proximité d'objets pouvaient être obtenus à partir des prélèvements effectués sur ces objets (Port *et al.*, 2006). Concernant les gants, si ceux-ci ne sont pas changés



régulièrement, le risque d'un transfert d'ADN au travers de ces derniers et de contamination des pièces à conviction, est augmenté (Bussard, 2008).

- La mauvaise décontamination des emballages ou des instruments servant à la manipulation des pièces à conviction (Sullivan et al., 2004; Tamariz et al., 2006; Shaw et al., 2008). Par exemple les pinces pour la détection de traces digitales peuvent être pollués et contaminer les pièces à conviction entre différents cas par transfert secondaire ou multiple (Proff et al., 2006). De même pour les gants et les différents instruments d'examen des pièces (Poy et Van Oorschot, 2006). Une étude a montré qu'un transfert d'ADN peut s'effectuer entre des pièces à conviction mises dans un même emballage. De même, un transfert d'ADN entre les différentes parties d'une même pièce est possible. L'ADN se trouvant sur la pièce à conviction peut se transférer sur l'emballage, qui, à son tour, peut transférer cet ADN sur une autre partie de la pièce (Goray et al., 2012). Ces expériences ont été réalisées avec des traces riches déposées sur les pièces mais il est possible que ce processus se produise également pour des traces de contact.

Deux autres études ont été menées sur la pollution via les instruments d'autopsie dans les morgues. Dans la première, 20 morgues ont été analysées. En moyenne, pour ces 20 établissements, 50% des instruments étaient recouverts d'ADN, les ciseaux étant les plus concernés alors qu'ils servent aux prélèvements sous les ongles (Rutty et al., 2000). Dans la seconde étude, deux morgues ont été étudiées et des résultats similaires ont été trouvés. Dans de nombreux cas, l'ADN prélevé sur d'anciens corps était toujours présent sur les instruments, la table d'autopsie ou la table des instruments lors de l'autopsie d'un nouveau corps (Schwark et al., 2012; Cemper-Kiesslich J et al., 2013).

- Lorsqu'un objet non décontaminé, comme par exemple le dossier du cas, est introduit dans un laboratoire. Taylor et al. 2016 ont ainsi observé le cas d'un transfert d'ADN d'un scientifique sur le dossier, puis de ce dossier sur les gants de la personne en charge de l'examen de la pièce à conviction et pour finir de ses gants à la pièce examinée.
- La pollution croisée : pollution ayant lieu par transfert d'ADN d'un échantillon à un autre lorsque ceux-ci sont analysés en même temps (Scherczinger et al., 1999).

Ces pollutions peuvent être minimisées lorsque les règles et les normes de protection et de sécurité sont respectées telles que le port de masques pour le nez et la bouche, lunettes, combinaisons et gants changés régulièrement, la décontamination des instruments entre

---

chaque analyse de chaque trace, la disposition des laboratoires évitant le croisement entre des pièces avec des traces riches en ADN et des pièces avec des traces pauvres en ADN.

Cependant, la pollution peut être inévitable, par exemple lorsqu'il y a l'intervention des secours mais, dans ce cas elle peut être documentée.

#### **6- Les erreurs de laboratoire**

Les erreurs de laboratoire sont à différencier de la pollution. Elles sont dues à des erreurs humaines. Elles résultent par exemple d'un mauvais étiquetage ou de tout événement qui induirait une inversion entre les échantillons. D'après Kloosterman et al. (2014), le NFI a relaté, entre 2008 et 2012, en moyenne 123 cas d'erreur par année dont la plupart n'ont pas eu de conséquences néfastes car elles ont pu être corrigées à temps.

Annexe IV: Données issues de l'étude sur l'efficacité d'extraction et de  
prélèvement

IV.1 Efficacité d'extraction

Présentation des résultats de quantification obtenus après l'extraction d'ADN en utilisant  
différents kits

Kit	Laboratoire	Nombre initial de cellule	Quantité théorique initiale d'ADN (pg)	Concentration d'ADN extrait (pg/uL)	Quantité d'ADN extrait (pg)	Efficacité d'extraction
<b>QIAshredder- QIAamp DNA mini</b>	ESC	100	600	5	121	0.20
				4	106	0.18
				3	84	0.14
				6	139	0.23
				5	113	0.19
				2	60	0.10
				5	113	0.19
				5	115	0.19
				3	73	0.12
				3	67	0.11
				4	106	0.18
				6	149	0.25
				6	154	0.26
				6	148	0.25
				8	212	0.35
				5	134	0.22
				7	185	0.31
				3	87	0.15
				8	196	0.33
				5	130	0.22
				9	228	0.38
				5	114	0.19
				5	123	0.21
4	90	0.15				
3	83	0.14				
8	208	0.35				
3	77	0.13				
6	155	0.26				
10	259	0.43				
9	232	0.39				
				7	405	0.68
				8	498	0.83
				5	301	0.50

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

<b>Manually Investigator Lyse&amp;Spin - QIAamp DNA Mini</b>	UGF	100	600	5	274	0.46
				7	429	0.72
				2	97	0.16
				4	243	0.40
				7	406	0.68
				2	124	0.21
				8	489	0.82
				8	478	0.80
				8	459	0.76
				10	572	0.95
				8	456	0.76
				5	294	0.49
				5	312	0.52
				6	335	0.56
				7	398	0.66
				6	363	0.60
				5	286	0.48
				4	259	0.43
	6	342	0.57			
	ESC	100	600	6	366	0.61
				4	247	0.41
				6	383	0.64
				6	378	0.63
				9	554	0.92
				8	501	0.83
				9	549	0.92
				2	114	0.19
				8	500	0.83
				10	597	0.99
				6	340	0.57
				4	252	0.42
				7	421	0.70
8				477	0.79	
6	376	0.63				
5	282	0.47				
6	354	0.59				
5	309	0.52				
5	302	0.50				
7	433	0.72				
5	316	0.53				
5	289	0.48				
			1	64	0.11	
			4	239	0.40	
			2	138	0.23	

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

<b>PrepFiler Express</b>	UGF	100	600	3	187	0.31
				3	189	0.32
				2	102	0.17
				3	206	0.34
				0	27	0.04
				0	24	0.04
				4	239	0.40
				2	129	0.22
				3	181	0.30
				4	217	0.36
				5	294	0.49
				4	252	0.42
				6	332	0.55
				1	69	0.11
				5	271	0.45
				3	190	0.32
2	111	0.19				
2	121	0.20				
8	454	0.76				

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

IV.2 Efficacité de prélèvement et d'extraction

Présentation des résultats de quantification obtenus après le prélèvement et l'extraction d'ADN en utilisant différents kits.

Quantité initiale théorique d'ADN (pg)	Concentration d'ADN extrait (pg/uL)	Quantité d'ADN extrait (pg)	Efficacité de prélèvement et d'extraction
3850	63	1583	0.41
3850	48	1192	0.31
3850	37	919	0.24
3850	27	665	0.17
3850	26	657	0.17
3850	31	780	0.20
3850	90	2245	0.58
3850	30	739	0.19
3850	25	623	0.16
3850	76	1903	0.49
3850	17	432	0.11
3850	18	462	0.12
3850	15	381	0.10
3850	28	704	0.18
3850	23	571	0.15
3850	49	1236	0.32
3850	33	825	0.21
3850	19	475	0.12
3850	36	897	0.23
3850	23	565	0.15
3850	34	838	0.22
3850	28	711	0.18
3850	27	681	0.18
3850	26	641	0.17
3850	21	517	0.13
3850	24	600	0.16
3850	44	1090	0.28
3850	35	876	0.23
3850	28	690	0.18
3850	30	742	0.19

## Annexe V: Données issues de l'étude sur la proportion de transfert : Transfert direct et prélèvement sur les mains

Présentation des résultats de quantifications, de la proportion des différents contributeurs obtenus après analyse STRmix et de l'analyse des profils ADN pour chaque participant issus de l'étude sur la proportion de transfert.

La qualité du profil partiel est décrite comme suit :

- Profil complet : la totalité des allèles est observé
- Profil presque complet : 1 à 4 allèles sont manquants
- Profil partiel : plus de 8 allèles sont présents mais au moins 4 allèles manquants
- Pas de profil : moins de 8 allèles sont présents

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

V.1 Participant 1

*Quantité d'ADN sur les mains*

Conditions	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	750	0.92	0.08	Profil complet
	275	0.80	0.20	Profil complet
	250	0.79	0.21	Profil complet
	250	1.00	0.00	Profil complet
	350	0.94	0.06	Profil complet
	250	0.67	0.33	Profil complet
	175	1.00	0.00	Profil partiel
	250	0.64	0.36	Profil partiel
	150	0.67	0.33	Profil presque complet
	250	0.77	0.23	Profil presque complet
Gants	50	0.00	0.00	Pas de profil
	125	1.00	0.00	Profil presque complet
	100	1.00	0.00	Profil presque complet
	2075	0.58	0.42	Profil presque complet
	825	0.78	0.22	Profil presque complet
	475	0.82	0.18	Profil presque complet
	1700	0.60	0.40	Profil presque complet
	850	0.86	0.14	Profil presque complet
	1300	0.77	0.23	Profil presque complet
	3075	0.61	0.39	Profil partiel
Lavage de mains	4175	0.52	0.48	Profil partiel
	9275	0.70	0.30	Profil presque complet
	11950	0.73	0.27	Profil presque complet
	100	0.61	0.39	Profil presque complet
	500	0.57	0.43	Profil presque complet
	350	0.66	0.34	Profil presque complet
	750	0.63	0.37	Profil partiel
	150	0.64	0.36	Profil partiel
	500	0.79	0.21	Profil presque complet



Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

*Quantité d'ADN transférée*

Condition	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	100	1.00	0.00	Profil partiel
	0.00	0.00	0.00	Pas de profil
	75	0.54	0.46	Profil presque complet
	50	0.62	0.38	Profil presque complet
	25	1.00	0.00	Profil partiel
	25	1.00	0.00	Profil partiel
	25	1.00	0.00	Profil partiel
	8	1.00	0.00	Profil partiel
	8	0.63	0.37	Profil partiel
Gants	8	1.00	0.00	Profil partiel
	13	0.71	0.29	Profil partiel
	8	1.00	0.00	Profil partiel
	850	1.00	0.00	Profil partiel
	2125	1.00	0.00	Profil partiel
	875	0.68	0.32	Profil partiel
	1900	1.00	0.00	Profil partiel
	5150	0.63	0.37	Profil partiel
	38900	1.00	0.00	Profil presque complet
	33175	0.73	0.27	Profil partiel
Lavage de mains	1375	1.00	0.00	Profil partiel
	800	1.00	0.00	Profil partiel
	375	1.00	0.00	Profil partiel
	400	1.00	0.00	Profil partiel
	625	1.00	0.00	Profil partiel
	425	0.72	0.28	Profil partiel
	525	1.00	0.00	Profil presque complet
	900	0.33	0.67	Profil partiel
	625	1.00	0.00	Profil partiel
	3300	1.00	0.00	Profil partiel

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

V.2 Participant 2

*Quantité d'ADN sur les mains*

Condition	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	441	1.00	0.00	profil complet
	377	1.00	0.00	Profil partiel
	186	1.00	0.00	Profil partiel
	32	1.00	0.00	Profil partiel
	69	1.00	0.00	Profil partiel
	66	0.59	0.41	Profil partiel
	4	1.00	0.00	Profil partiel
	24	1.00	0.00	Profil partiel
	37	1.00	0.00	Profil partiel
	27	1.00	0.00	Profil partiel
6	1.00	0.00	Pas de Profil	
Gants	287	1.00	0.00	Profil partiel
	433	0.74	0.26	Profil partiel
	3431	1.00	0.00	Profil complet
	4641	0.96	0.04	Profil complet
	2157	1.00	0.00	Profil complet
	6121	1.00	0.00	Profil complet
	4834	1.00	0.00	Profil complet
	1700	0.82	0.18	Profil complet
	1360	0.67	0.33	Profil complet
Lavage de mains	299	1.00	0.00	Profil complet
	622	1.00	0.00	Profil presque complet
	400	1.00	0.00	Profil presque complet
	116	1.00	0.00	Profil presque complet
	497	1.00	0.00	Profil presque complet
	171	0.69	0.31	Profil presque complet
	235	1.00	0.00	Profil partiel
	2422	1.00	0.00	Pas de profil
	524	0.61	0.39	Profil presque complet
	373	0.78	0.22	Profil presque complet

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

*Quantité d'ADN transférée*

Condition	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	28	1.00	0.00	Pas de profil
	36	1.00	0.00	Pas de profil
	0	0.00	0.00	Pas de Profil
	2	1.00	0.00	Pas de Profil
	10	1.00	0.00	Pas de Profil
	3	1.00	0.00	Pas de Profil
	3	1.00	0.00	Pas de Profil
	6	1.00	0.00	Pas de Profil
Gants	4	1.00	0.00	Pas de Profil
	4	1.00	0.00	Pas de Profil
	167	1.00	0.00	Profil partiel
	307	1.00	0.00	Profil partiel
	331	1.00	0.00	Profil partiel
	92	1.00	0.00	Pas de Profil
	139	1.00	0.00	Profil partiel
	91	1.00	0.00	Profil partiel
	73	1.00	0.00	Profil partiel
Lavage de mains	191	0.64	0.00	Profil partiel
	0	1.00	0.00	Pas de profil
	82	1.00	0.00	Pas de Profil
	51	1.00	0.00	Pas de Profil
	98	1.00	0.00	Profil partiel
	22	1.00	0.00	Pas de profil
	23	1.00	0.00	Pas de profil
	186	1.00	0.00	Profil partiel

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

V.3 Participant 3

*Quantité d'ADN sur les mains*

Conditions	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	21025	1.00	0.00	Profil complet
	8375	1.00	0.00	Profil complet
	16100	1.00	0.00	Profil complet
	10425	0.98	0.02	Profil complet
	8400	1.00	0.00	Profil complet
	11475	1.00	0.00	Profil complet
	4750	1.00	0.00	Profil complet
	3800	1.00	0.00	Profil complet
Gants	4000	0.97	0.03	Profil complet
	2650	1.00	0.00	Profil complet
	4025	1.00	0.00	Profil complet
	3200	1.00	0.00	Profil complet
	4475	1.00	0.00	Profil complet
	9750	0.97	0.03	Profil complet
	3275	1.00	0.00	Profil complet
	3925	1.00	0.00	Profil complet
	3800	1.00	0.00	Profil complet
	2825	1.00	0.00	Profil complet
	2800	1.00	0.00	Profil complet
Lavage de mains	2500	1.00	0.00	Profil complet
	3325	0.7	0.30	Profil complet
	2250	1.00	0.00	Profil complet
	1750	0.8	0.20	Profil complet
	225	1.00	0.00	Profil complet
	475	1.00	0.00	Profil presque complet
	425	1.00	0.00	Profil presque complet
	250	1.00	0.00	Profil complet
	475	1.00	0.00	Profil complet

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

*Quantité d'ADN transférée*

Conditions	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	975	0.55	0.45	Profil partiel
	750	1.00	0.00	Profil partiel
	375	1.00	0.00	Profil partiel
	225	1.00	0.00	Profil partiel
	1300	1.00	0.00	Profil partiel
	825	0.56	0.44	Profil partiel
	225	0.69	0.31	Profil partiel
	25	1.00	0.00	Pas de profil
Gants	5	0.86	0.14	Profil partiel
	175	0.81	0.19	Profil partiel
	150	0.73	0.27	Profil presque complet
	75	1.00	0.00	Profil presque complet
	25	0.69	0.31	Profil partiel
	50	0.69	0.31	Profil partiel
	10	1.00	0.00	Pas de profil
	325	1.00	0.00	Profil complet
	175	1.00	0.00	Profil complet
	300	1.00	0.00	Profil complet
	125	1.00	0.00	Profil partiel
Lavage de mains	275	1.00	0.00	Pas de profil
	275	1.00	0.00	Profil partiel
	125	1.00	0.00	Pas de profil
	125	1.00	0.00	Pas de profil
	0	0.80	0.20	Profil partiel
	50	0.68	0.32	Profil complet
	0	0.61	0.39	Pas de profil
	3	1.00	0.00	Pas de profil
	25	1.00	0.00	Pas de profil
	0	1.00	0.00	Pas de profil

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

V.4 Participant 4

*Quantité d'ADN sur les mains*

Conditions	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	0	0.86	0.14	Profil complet
	10636	1.00	0.00	Profil complet
	7607	1.00	0.00	Profil complet
	20476	0.93	0.07	Profil complet
	11750	0.52	0.48	Profil partiel
	4388	1.00	0.00	Profil presque complet
	13775	0.94	0.06	Profil presque complet
	4837	1.00	0.00	Profil complet
	2622	1.00	0.00	Profil presque complet
	5507	0.92	0.08	Profil complet
	17802	1.00	0.00	Profil complet
Gants	3220	0.89	0.11	Profil complet
	6540	0.92	0.08	Profil complet
	114	0.62	0.38	Profil partiel
	19	0.37	0.63	Profil partiel
	14	0.27	0.73	Profil partiel
	190	0.75	0.25	Profil complet
	49	0.61	0.39	Profil partiel
	77	0.67	0.33	Profil partiel
Lavage de mains	196	1.00	0.00	Pas de profil
	7871	1.00	0.00	Profil complet
	3637	0.53	0.47	Profil complet
	8943	0.52	0.48	Profil complet
	4080	0.61	0.39	Profil complet
	4792	0.57	0.43	Profil partiel
	1404	0.91	0.09	Profil complet
	1233	0.64	0.36	Profil partiel
	1019	0.78	0.22	Profil partiel

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

*Quantité d'ADN transférée*

Conditions	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	432	1.00	0.00	Profil presque complet
	98	0.66	0.34	Profil partiel
	543	0.85	0.15	Profil partiel
	1810	1.00	0.00	Profil partiel
	2886	0.90	0.10	Profil partiel
	108	1.00	0.00	Pas de profil
	3172	1.00	0.00	Profil complet
	284	1.00	0.00	Profil partiel
	93	0.73	0.27	Pas de profil
	42	0.60	0.40	Profil partiel
	413	0.69	0.31	Profil partiel
	844	0.77	0.23	Profil presque complet
Gants	454	0.63	0.37	Profil partiel
	71	0.68	0.32	Profil partiel
	3415	0.83	0.17	Profil complet
	2892	0.86	0.14	Profil presque complet
	2362	0.84	0.16	Profil complet
	19	0.76	0.24	Profil complet
	0	0.38	0.62	Profil partiel
	774	0.92	0.08	Profil complet
	1509	0.89	0.11	Profil complet
Lavage de mains	156	1.00	0.00	Pas de profil
	479	0.70	0.30	Pas de profil
	143	0.74	0.26	Pas de profil
	992	0.31	0.69	Profil partiel
	204	0.34	0.66	Pas de profil
	47	0.79	0.21	Pas de profil
	63	0.00	1.00	Pas de profil
	107	1.00	0.00	Pas de profil
	62	0.75	0.25	Pas de profil

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

V.5 Participant 5

*Quantité d'ADN sur les mains*

Conditions	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	100	1.00	0.00	Profil presque complet
	100	1.00	0.00	Profil presque complet
	2375	1.00	0.00	Profil presque complet
	3500	1.00	0.00	Profil presque complet
	1725	1.00	0.00	Profil presque complet
	5150	1.00	0.00	Profil presque complet
	2750	1.00	0.00	Profil presque complet
	8425	1.00	0.00	Profil presque complet
Gants	8675	1.00	0.00	Profil presque complet
	1025	1.00	0.00	Profil presque complet
	850	0.92	0.08	Profil complet
	12875	0.99	0.01	Profil complet
	17100	1.00	0.00	Profil complet
	10300	1.00	0.00	Profil complet
	6725	1.00	0.00	Profil complet
	2700	0.93	0.07	Profil complet
	2275	0.87	0.13	Profil complet
1900	1.00	0.00	Profil complet	
Lavage de mains	6400	1.00	0.00	Profil complet
	7425	1.00	0.00	Profil complet
	9650	1.00	0.00	Profil complet
	5950	1.00	0.00	Profil complet
	8600	1.00	0.00	Profil complet
	3125	1.00	0.00	Profil complet
	4875	1.00	0.00	Profil complet
	3375	1.00	0.00	Profil complet
	4200	0.84	0.16	Profil complet
	5900	0.86	0.14	Profil complet



Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

*Quantité d'ADN transférée*

Conditions	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	275	1.00	0.00	Pas de profil
	75	1.00	0.00	Pas de profil
	50	0.79	0.21	Pas de profil
	475	0.89	0.11	profil presque complet
	150	0.68	0.32	Profil partiel
	175	1.00	0.00	profil partiel
	350	0.68	0.32	profil partiel
	475	0.65	0.35	Profil presque complet
	225	1.00	0.00	profil partiel
	500	1.00	0.00	profil partiel
Gants	150	1.00	0.00	profil partiel
	0	0.00	0.00	Pas de profil
	125	1.00	0.00	profil presque complet
	1050	1.00	0.00	Profil complet
	125	0.71	0.29	Profil très partiel
	1200	1.00	0.00	Profil complet
	850	1.00	0.00	Profil presque complet
	200	0.59	0.41	Profil partiel
	125	0.60	0.40	Profil partiel
	375	1.00	0.00	Profil partiel
Lavage de mains	25	1.00	0.00	Profil partiel
	675	1.00	0.00	profil presque complet
	350	1.00	0.00	profil presque complet
	525	0.68	0.32	Profil presque complet
	525	0.65	0.35	Profil presque complet
	8	1.00	0.00	Pas de profil
	100	0.72	0.28	Pas de profil
	200	0.76	0.24	Profil partiel
	25	1.00	0.00	Pas de profil
	250	0.70	0.30	Profil partiel

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

V.6 Participant 6

*Quantité d'ADN sur les mains*

Conditions	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	2375	1.00	0.00	Profil presque complet
	5000	1.00	0.00	Profil complet
	4225	1.00	0.00	Profil complet
	3025	1.00	0.00	Profil complet
	4025	1.00	0.00	Profil complet
	11925	1.00	0.00	Profil complet
	75	1.00	0.00	Profil complet
	3550	1.00	0.00	Profil complet
	3150	1.00	0.00	Profil complet
	3575	1.00	0.00	Profil complet
Gants	2825	1.00	0.00	Profil complet
	3500	1.00	0.00	Profil complet
	1800	1.00	0.00	Profil complet
	1850	1.00	0.00	Profil complet
	1125	1.00	0.00	Profil complet
	1150	1.00	0.00	Profil complet
	1975	1.00	0.00	Profil complet
	1775	1.00	0.00	Profil complet
	2325	1.00	0.00	Profil complet
	1550	1.00	0.00	Profil complet
Lavage de mains	75	1.00	0.00	Profil complet
	2600	1.00	0.00	Profil complet
	2450	1.00	0.00	Profil complet
	2225	1.00	0.00	Profil complet
	3900	1.00	0.00	Profil complet
	6375	1.00	0.00	Profil complet
	8725	1.00	0.00	Profil complet
	4450	1.00	0.00	Profil complet
	450	0.86	0.14	Profil complet
	2450	1.00	0.00	Profil complet

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

*Quantité d'ADN transférée*

Conditions	Quantité d'ADN (en pg)	Proportion POI	Proportion Non POI	Profil du POI
Aucune	1625	1.00	0.00	Profil partiel
	1875	0.64	0.36	Profil partiel
	3025	0.68	0.32	Profil partiel
	5250	1.00	0.00	Profil complet
	1600	0.69	0.31	Profil complet
	2550	1.00	0.00	Profil complet
	2475	0.78	0.22	Profil complet
	0	1.00	0.00	Pas de profil
	100	1.00	0.00	Profil partiel
	225	1.00	0.00	Profil presque complet
Gants	225	0.59	0.41	Profil presque complet
	125	0.73	0.27	Profil presque complet
	225	0.70	0.3	Profil presque complet
	350	0.88	0.12	Profil presque complet
	250	1.00	0.00	Profil complet
	325	0.67	0.33	Profil presque complet
	525	0.83	0.17	Profil partiel
	300	1.00	0.00	Profil complet
	125	0.88	0.12	Profil presque complet
	275	1.00	0.00	Profil complet
Lavage de mains	250	1.00	0.00	Profil presque complet
	225	1.00	0.00	Profil presque complet
	225	1.00	0.00	Profil presque complet
	175	0.69	0.31	Profil presque complet
	100	1.00	0.00	Profil partiel
	425	1.00	0.00	Profil complet
	775	1.00	0.00	Profil complet
	525	0.89	0.11	Profil complet
	3975	1.00	0.00	Profil complet
	375	0.84	0.16	Profil complet

Annexe VI: Données utilisées pour informer les variables du BN

VI.1 Variable « Efficacité d'extraction »

Présentation des données utilisées pour informer la variable « Efficacité d'extraction » (en jaune).

	Quantité théorique initiale d'ADN (pg)	Concentration d'ADN extrait (pg/uL)	Quantité d'ADN extrait (pg)	Efficacité
<b>Efficacité d'extraction</b>	600	5	121	0.20
		4	106	0.18
		3	84	0.14
		6	139	0.23
		5	113	0.19
		2	60	0.10
		5	113	0.19
		5	115	0.19
		3	73	0.12
		3	67	0.11
		4	106	0.18
		6	149	0.25
		6	154	0.26
		6	148	0.25
		8	212	0.35
		5	134	0.22
		7	185	0.31
		3	87	0.15
		8	196	0.33
		5	130	0.22
		9	228	0.38
		5	114	0.19
		5	123	0.21
4	90	0.15		
3	83	0.14		
8	208	0.35		
3	77	0.13		
6	155	0.26		
10	259	0.43		
9	232	0.39		

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

VI.2 Variable « Efficacité de prélèvement »

Présentation des données utilisées pour informer la variable « Efficacité de prélèvement » (en  
jaune).

	Quantité théorique initiale d'ADN (pg)	Concentration d'ADN extrait (pg/uL)	Quantité d'ADN extrait (pg)	Efficacité
<b>Efficacité d'extraction</b>	600	5	121	0.2
		4	106	0.18
		3	84	0.14
		6	139	0.23
		5	113	0.19
		2	60	0.1
		5	113	0.19
		5	115	0.19
		3	73	0.12
		3	67	0.11
		4	106	0.18
		6	149	0.25
		6	154	0.26
		6	148	0.25
		8	212	0.35
		5	134	0.22
		7	185	0.31
		3	87	0.15
		8	196	0.33
		5	130	0.22
		9	228	0.38
		5	114	0.19
		5	123	0.21
4	90	0.15		
3	83	0.14		
8	208	0.35		
3	77	0.13		
6	155	0.26		
10	259	0.43		
9	232	0.39		

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

<b>Efficacité d'extraction et de prélèvement</b>	3850	63	1583	0.41
		48	1192	0.31
		37	919	0.24
		27	665	0.17
		26	657	0.17
		31	780	0.2
		90	2245	0.58
		30	739	0.19
		25	623	0.16
		76	1903	0.49
		17	432	0.11
		18	462	0.12
		15	381	0.1
		28	704	0.18
		23	571	0.15
		49	1236	0.32
		33	825	0.21
		19	475	0.12
		36	897	0.23
		23	565	0.15
		34	838	0.22
		28	711	0.18
		27	681	0.18
		26	641	0.17
		21	517	0.13
		24	600	0.16
44	1090	0.28		
35	876	0.23		
28	690	0.18		
30	742	0.19		

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

VI.3 « Quantité d'ADN sur les mains

Présentation des données utilisées pour informer les variables « Quantités d'ADN sur les mains » (en jaune) en fonction du participant.

n° échantillon	Expérience	Participant	Quantité (en ng)	% POI	% AO	% Non_POI	Quantité POI
MM1	Prélèvement sur les mains	2	0.441	1.00	0.00	0.00	0.441
MM2	Prélèvement sur les mains	2	0.377	1.00	0.00	0.00	0.377
MM3	Prélèvement sur les mains	2	0.186	1.00	0.00	0.00	0.186
MM4	Prélèvement sur les mains	2	0.032	1.00	0.00	0.00	0.032
MM5	Prélèvement sur les mains	2	0.069	1.00	0.00	0.00	0.069
MM6	Prélèvement sur les mains	2	0.066	0.56	0.00	0.44	0.037
MM7	Prélèvement sur les mains	2	0.004	1.00	0.00	0.00	0.004
MM8	Prélèvement sur les mains	2	0.024	1.00	0.00	0.00	0.024
MM9	Prélèvement sur les mains	2	0.037	1.00	0.00	0.00	0.037
MM10	Prélèvement sur les mains	2	0.027	1.00	0.00	0.00	0.027
MM11	Prélèvement sur les mains	2	0.006	1.00	0.00	0.00	0.006
MM12	Prélèvement sur les mains	2	0.287	1.00	0.00	0.00	0.287
MM13	Prélèvement sur les mains	2	0.433	1.00	0.00	0.00	0.433
MM14	Prélèvement sur les mains	2	3.431	1.00	0.00	0.00	3.431
MM15	Prélèvement sur les mains	2	4.641	0.91	0.00	0.09	4.223
MM16	Prélèvement sur les mains	2	2.157	1.00	0.00	0.00	2.157
MM17	Prélèvement sur les mains	2	6.121	1.00	0.00	0.00	6.121
MM18	Prélèvement sur les mains	2	4.834	1.00	0.00	0.00	4.834
MM19	Prélèvement sur les mains	2	1.700	0.83	0.00	0.17	1.411
MM20	Prélèvement sur les mains	2	1.360	0.67	0.00	0.33	0.911
MM21	Prélèvement sur les mains	2	0.299	1.00	0.00	0.00	0.299
MM22	Prélèvement sur les mains	2	0.622	1.00	0.00	0.00	0.622
MM23	Prélèvement sur les mains	2	0.400	1.00	0.00	0.00	0.400
MM24	Prélèvement sur les mains	2	0.116	1.00	0.00	0.00	0.116
MM25	Prélèvement sur les mains	2	0.497	1.00	0.00	0.00	0.497
MM26	Prélèvement sur les mains	2	0.171	0.73	0.00	0.27	0.125

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

MM27	Prélèvement sur les mains	2	0.235	1.00	0.00	0.00	0.235
MM28	Prélèvement sur les mains	2	2.422	1.00	0.00	0.00	2.422
MM29	Prélèvement sur les mains	2	0.524	0.61	0.00	0.39	0.320
MM30	Prélèvement sur les mains	2	0.373	0.78	0.00	0.22	0.291
FM1	Prélèvement sur les mains	6	2.375	1.00	0.00	0.00	2.375
FM2	Prélèvement sur les mains	6	5.000	1.00	0.00	0.00	5.000
FM3	Prélèvement sur les mains	6	4.225	1.00	0.00	0.00	4.225
FM4	Prélèvement sur les mains	6	3.025	1.00	0.00	0.00	3.025
FM5	Prélèvement sur les mains	6	4.025	1.00	0.00	0.00	4.025
FM6	Prélèvement sur les mains	6	11.925	1.00	0.00	0.00	11.925
FM7	Prélèvement sur les mains	6	0.075	1.00	0.00	0.00	0.075
FM8	Prélèvement sur les mains	6	3.550	1.00	0.00	0.00	3.550
FM9	Prélèvement sur les mains	6	3.150	1.00	0.00	0.00	3.150
FM10	Prélèvement sur les mains	6	3.575	1.00	0.00	0.00	3.575
FM11	Prélèvement sur les mains	6	0.075	1.00	0.00	0.00	0.075
FM12	Prélèvement sur les mains	6	2.600	1.00	0.00	0.00	2.600
FM13	Prélèvement sur les mains	6	2.450	0.93	0.00	0.07	2.279
FM14	Prélèvement sur les mains	6	2.225	1.00	0.00	0.00	2.225
FM15	Prélèvement sur les mains	6	3.900	1.00	0.00	0.00	3.900
FM16	Prélèvement sur les mains	6	6.375	1.00	0.00	0.00	6.375
FM17	Prélèvement sur les mains	6	8.725	1.00	0.00	0.00	8.725
FM18	Prélèvement sur les mains	6	4.450	1.00	0.00	0.00	4.450
FM19	Prélèvement sur les mains	6	0.450	0.90	0.00	0.10	0.405
FM20	Prélèvement sur les mains	6	2.450	0.93	0.00	0.07	2.279
FM21	Prélèvement sur les mains	6	2.825	0.95	0.00	0.05	2.684
FM22	Prélèvement sur les mains	6	3.500	1.00	0.00	0.00	3.500
FM23	Prélèvement sur les mains	6	1.800	0.94	0.00	0.06	1.692
FM24	Prélèvement sur les mains	6	1.850	1.00	0.00	0.00	1.850
FM25	Prélèvement sur les mains	6	1.125	0.94	0.00	0.06	1.058
FM26	Prélèvement sur les mains	6	1.150	0.91	0.00	0.09	1.047
FM27	Prélèvement sur les mains	6	1.975	0.92	0.00	0.08	1.817
FM28	Prélèvement sur les mains	6	1.775	1.00	0.00	0.00	1.775
FM29	Prélèvement	6	2.325	1.00	0.00	0.00	2.325



Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

	sur les mains						
FM30	Prélèvement sur les mains	6	1.550	1.00	0.00	0.00	1.550

VI.4 « Proportion de transfert » : Propositions du procureur est vraie

Présentation des données utilisées pour informer les variables « Proportion de transfert » (en jaune) en fonction du participant qui est le POI lorsque la proposition du Procureur  $H_p$  est vraie

n° échantillon	Proposition	POI	Quantité (en ng)	% POI	% AO	% Non_POI	Quantité POI
MC1	Procureur	2	0.025	1.00	0.00	0.00	0.025
MC2	Procureur	2	0.025	1.00	0.00	0.00	0.025
MC4	Procureur	2	0.000	1.00	0.00	0.00	0.000
MC5	Procureur	2	2.278	1.00	0.00	0.00	2.278
MC6	Procureur	2	9.986	1.00	0.00	0.00	9.986
MC7	Procureur	2	0.110	0.00	0.00	0.00	0.000
MC8	Procureur	2	4.807	1.00	0.00	0.00	4.807
MC10	Procureur	2	8.522	1.00	0.00	0.00	8.522
MC12	Procureur	2	5.443	1.00	0.00	0.00	5.443
MC13	Procureur	2	5.257	1.00	0.00	0.00	5.257
MC14	Procureur	2	0.175	1.00	0.00	0.00	0.175
MC15	Procureur	2	0.300	1.00	0.00	0.00	0.300
MC16	Procureur	2	0.325	1.00	0.00	0.00	0.325
MC17	Procureur	2	0.100	1.00	0.00	0.00	0.100
MC18	Procureur	2	0.150	1.00	0.00	0.00	0.150
MC19	Procureur	2	0.100	1.00	0.00	0.00	0.100
MC20	Procureur	2	0.075	0.12	0.00	0.88	0.009
MC21	Procureur	2	0.200	1.00	0.00	0.00	0.200
MC22	Procureur	2	0.000	1.00	0.00	0.00	0.000
MC27	Procureur	2	0.100	1.00	0.00	0.00	0.100
MC28	Procureur	2	0.025	1.00	0.00	0.00	0.025
MC29	Procureur	2	0.025	1.00	0.00	0.00	0.025
MC30	Procureur	2	0.175	1.00	0.00	0.00	0.175
FC1	Procureur	6	1.625	1.00	0.00	0.00	1.625
FC2	Procureur	6	1.875	0.61	0.00	0.39	1.144
FC3	Procureur	6	3.025	0.49	0.00	0.51	1.482
FC4	Procureur	6	5.250	1.00	0.00	0.00	5.250
FC5	Procureur	6	1.600	0.84	0.00	0.16	1.344
FC6	Procureur	6	2.550	1.00	0.00	0.00	2.550
FC7	Procureur	6	2.475	0.78	0.00	0.22	1.931
FC8	Procureur	6	0.001	1.00	0.00	0.00	0.001
FC9	Procureur	6	0.100	1.00	0.00	0.00	0.100
FC10	Procureur	6	0.225	1.00	0.00	0.00	0.225

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

FC11	Procureur	6	0.250	1.00	0.00	0.00	0.250
FC12	Procureur	6	0.225	1.00	0.00	0.00	0.225
FC13	Procureur	6	0.225	1.00	0.00	0.00	0.225
FC14	Procureur	6	0.175	1.00	0.00	0.00	0.175
FC15	Procureur	6	0.100	1.00	0.00	0.00	0.100
FC16	Procureur	6	0.425	1.00	0.00	0.00	0.425
FC17	Procureur	6	0.775	1.00	0.00	0.00	0.775
FC18	Procureur	6	0.525	0.91	0.00	0.09	0.478
FC19	Procureur	6	3.975	1.00	0.00	0.00	3.975
FC20	Procureur	6	0.375	0.83	0.00	0.17	0.311
FC21	Procureur	6	0.225	0.85	0.00	0.15	0.191
FC22	Procureur	6	0.125	0.73	0.00	0.27	0.091
FC23	Procureur	6	0.225	0.57	0.00	0.43	0.128
FC24	Procureur	6	0.350	0.89	0.00	0.11	0.312
FC25	Procureur	6	0.250	1.00	0.00	0.00	0.250
FC26	Procureur	6	0.325	0.70	0.00	0.30	0.228
FC27	Procureur	6	0.525	0.83	0.00	0.17	0.436
FC28	Procureur	6	0.300	1.00	0.00	0.00	0.300
FC29	Procureur	6	0.125	0.88	0.00	0.12	0.110
FC30	Procureur	6	0.275	1.00	0.00	0.00	0.275

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

VI.5 « Proportion de transfert » : Propositions de la défense Hd<sub>1</sub> est vraie

Présentation des données utilisées pour informer les variables « Proportion de transfert » concernant le POI et l'AO (en jaune), en fonction du participant qui est le POI, lorsque la proposition de la défense Hd<sub>1</sub> est vraie.

n° échantillon	Proposition	POI	Quantité (en ng)	% POI	% AO	% Non_POI	Quantité POI	Quantité AO
MTS1	Défense	6	10.900	0.00	0.25	0.75	0.000	68.125
MTS2	Défense	6	1.070	0.21	0.18	0.61	0.225	4.815
MTS3	Défense	6	0.733	0.58	0.00	0.42	0.425	0.000
MTS4	Défense	6	1.220	0.23	0.45	0.32	0.281	13.725
MTS5	Défense	6	0.093	0.12	0.01	0.87	0.011	0.023
MTS6	Défense	6	0.077	0.30	0.00	0.70	0.023	0.000
MTS7	Défense	6	0.076	0.15	0.36	0.49	0.011	0.687
MTS8	Défense	6	0.087	0.18	0.59	0.23	0.016	1.280
MTS9	Défense	6	0.092	0.34	0.22	0.44	0.031	0.506
MTS10	Défense	6	0.120	0.26	0.26	0.48	0.031	0.780
MTS11	Défense	6	0.077	0.25	0.33	0.42	0.019	0.631
MTS12	Défense	6	0.165	0.18	0.23	0.59	0.030	0.949
MTS14	Défense	6	0.074	0.15	0.18	0.67	0.011	0.334
MTS15	Défense	6	0.157	0.17	0.00	0.83	0.027	0.000
MTS17	Défense	6	0.225	0.13	0.45	0.42	0.029	2.531
MTS18	Défense	6	0.081	0.41	0.11	0.48	0.033	0.222
MTS19	Défense	6	0.109	0.00	0.00	1.00	0.000	0.000
MTS20	Défense	6	0.097	0.57	0.03	0.40	0.055	0.073
MTS21	Défense	6	0.034	0.14	0.03	0.83	0.005	0.026
MTS22	Défense	6	0.067	0.15	0.73	0.12	0.010	1.219
MTS23	Défense	6	0.047	0.37	0.11	0.52	0.017	0.128
MTS24	Défense	6	0.035	0.03	0.36	0.61	0.001	0.313
MTS25	Défense	6	0.024	0.00	0.51	0.49	0.000	0.311
MTS26	Défense	6	0.023	0.13	0.05	0.82	0.003	0.029
MTS27	Défense	6	0.112	0.00	0.23	0.77	0.000	0.644
MTS28	Défense	6	0.070	0.64	0.01	0.35	0.045	0.018
MTS29	Défense	6	0.030	0.00	0.19	0.81	0.000	0.142
MTS30	Défense	6	0.032	0.92	0.01	0.07	0.030	0.008
FTS1	Défense	2	0.128	0.00	1.00	0.00	0.000	3.200
FTS2	Défense	2	0.395	0.00	1.00	0.00	0.000	9.875
FTS3	Défense	2	0.072	0.00	1.00	0.00	0.000	1.808
FTS4	Défense	2	0.103	0.00	1.00	0.00	0.000	2.575
FTS5	Défense	2	0.233	0.00	1.00	0.00	0.000	5.825
FTS6	Défense	2	0.064	0.00	1.00	0.00	0.000	1.610
FTS7	Défense	2	0.099	0.00	1.00	0.00	0.000	2.480

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

FTS8	Défense	2	0.081	0.00	1.00	0.00	0.000	2.025
FTS9	Défense	2	0.015	0.00	1.00	0.00	0.000	0.378
FTS10	Défense	2	0.095	0.00	1.00	0.00	0.000	2.370
FTS11	Défense	2	0.021	0.00	1.00	0.00	0.000	0.530
FTS12	Défense	2	0.027	0.00	1.00	0.00	0.000	0.670
FTS13	Défense	2	0.035	0.00	1.00	0.00	0.000	0.875
FTS14	Défense	2	0.082	0.00	1.00	0.00	0.000	2.038
FTS15	Défense	2	0.243	0.00	1.00	0.00	0.000	6.075
FTS16	Défense	2	0.051	0.00	1.00	0.00	0.000	1.283
FTS17	Défense	2	0.099	0.00	1.00	0.00	0.000	2.480
FTS18	Défense	2	0.122	0.11	0.61	0.28	0.013	1.861
FTS19	Défense	2	0.080	0.00	1.00	0.00	0.000	2.000
FTS20	Défense	2	0.044	0.15	0.50	0.35	0.007	0.544
FTS21	Défense	2	0.104	0.00	1.00	0.00	0.000	2.600
FTS22	Défense	2	0.092	0.00	0.46	0.54	0.000	1.061
FTS23	Défense	2	0.137	0.00	0.51	0.49	0.000	1.747
FTS24	Défense	2	0.048	0.00	1.00	0.00	0.000	1.195
FTS25	Défense	2	0.033	0.17	0.00	0.83	0.006	0.000
FTS26	Défense	2	0.116	0.00	0.00	1.00	0.000	0.000
FTS27	Défense	2	0.045	0.00	1.00	0.00	0.000	1.120
FTS28	Défense	2	0.170	0.00	1.00	0.00	0.000	4.250
FTS29	Défense	2	0.050	0.00	1.00	0.00	0.000	1.258
FTS30	Défense	2	0.192	0.00	1.00	0.00	0.000	4.800

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

VI.6 « Proportion de transfert » : Propositions de la défense Hd<sub>3</sub> est vraie

Présentation des données utilisées pour informer les variables « Proportion de transfert » concernant le POI et l'AO (en jaune), en fonction du participant qui est le POI, lorsque la proposition de la défense Hd<sub>3</sub> est vraie.

n° échantillon	Proposition	POI	Quantité (en ng)	% POI	% AO	% Non_POI	Quantité POI	Quantité AO
MPC1	Défense	6	0.077	0.20	0.13	0.67	0.015	0.010
MPC2	Défense	6	0.156	0.80	0.16	0.04	0.125	0.025
MPC3	Défense	6	0.693	0.69	0.31	0.00	0.478	0.215
MPC4	Défense	6	1.040	0.64	0.35	0.01	0.666	0.365
MPC5	Défense	6	0.088	0.19	0.81	0.00	0.017	0.071
MPC6	Défense	6	0.795	0.59	0.15	0.26	0.469	0.119
MPC7	Défense	6	0.324	0.56	0.10	0.34	0.181	0.032
MPC8	Défense	6	0.086	0.47	0.21	0.32	0.040	0.018
MPC9	Défense	6	0.104	0.27	0.73	0.00	0.028	0.076
MPC10	Défense	6	0.121	0.72	0.08	0.20	0.087	0.010
MPC11	Défense	6	0.080	0.34	0.37	0.29	0.027	0.030
MPC12	Défense	6	0.246	0.56	0.27	0.17	0.138	0.066
MPC13	Défense	6	0.203	0.70	0.17	0.13	0.142	0.034
MPC14	Défense	6	0.265	0.23	0.14	0.63	0.061	0.037
MPC15	Défense	6	0.273	0.12	0.77	0.11	0.033	0.210
MPC16	Défense	6	0.327	0.29	0.36	0.35	0.095	0.118
MPC17	Défense	6	0.080	0.24	0.34	0.42	0.019	0.027
MPC18	Défense	6	0.151	0.09	0.77	0.14	0.014	0.116
MPC19	Défense	6	0.116	0.18	0.23	0.59	0.021	0.027
MPC20	Défense	6	0.178	0.21	0.24	0.55	0.037	0.043
MPC21	Défense	6	0.039	0.33	0.40	0.27	0.013	0.015
MPC22	Défense	6	0.120	0.52	0.13	0.35	0.062	0.016
MPC23	Défense	6	0.435	0.75	0.17	0.08	0.326	0.074
MPC24	Défense	6	0.623	0.37	0.56	0.07	0.231	0.349
MPC25	Défense	6	0.123	0.69	0.25	0.06	0.085	0.031
MPC26	Défense	6	0.072	0.24	0.13	0.63	0.017	0.009
MPC27	Défense	6	0.197	0.23	0.49	0.28	0.045	0.097
MPC28	Défense	6	0.210	0.20	0.26	0.54	0.042	0.055
MPC29	Défense	6	0.167	0.56	0.36	0.08	0.094	0.060
MPC30	Défense	6	0.291	0.57	0.43	0.00	0.166	0.125
FPC1	Défense	2	0.032	0.68	0.32	0.00	0.021	0.010
FPC2	Défense	2	0.225	0.61	0.27	0.12	0.137	0.061
FPC3	Défense	2	0.112	0.52	0.34	0.14	0.058	0.038
FPC4	Défense	2	0.078	0.35	0.29	0.36	0.027	0.023
FPC5	Défense	2	0.563	0.71	0.14	0.15	0.400	0.079

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

FPC6	Défense	2	0.593	0.66	0.31	0.03	0.391	0.184
FPC7	Défense	2	0.363	0.22	0.47	0.31	0.080	0.170
FPC8	Défense	2	0.115	0.60	0.18	0.22	0.069	0.021
FPC9	Défense	2	0.177	0.45	0.10	0.45	0.080	0.018
FPC10	Défense	2	0.085	0.24	0.27	0.49	0.020	0.023
FPC11	Défense	2	0.088	0.35	0.14	0.51	0.031	0.012
FPC12	Défense	2	0.082	0.63	0.23	0.14	0.052	0.019
FPC13	Défense	2	0.050	0.79	0.06	0.15	0.039	0.003
FPC14	Défense	2	0.350	0.84	0.16	0.00	0.294	0.056
FPC15	Défense	2	0.187	0.73	0.17	0.10	0.137	0.032
FPC16	Défense	2	0.084	0.58	0.13	0.29	0.049	0.011
FPC17	Défense	2	0.189	0.91	0.02	0.07	0.172	0.004
FPC18	Défense	2	0.428	0.23	0.67	0.10	0.098	0.286
FPC19	Défense	2	0.166	0.48	0.45	0.07	0.080	0.075
FPC20	Défense	2	0.181	0.52	0.48	0.00	0.094	0.087
FPC21	Défense	2	0.490	0.60	0.12	0.28	0.294	0.059
FPC22	Défense	2	0.253	0.70	0.22	0.08	0.177	0.056
FPC23	Défense	2	0.161	0.82	0.18	0.00	0.132	0.029
FPC24	Défense	2	0.109	0.33	0.66	0.01	0.036	0.072
FPC25	Défense	2	0.457	0.68	0.29	0.03	0.311	0.133
FPC26	Défense	2	0.535	0.65	0.19	0.16	0.348	0.102
FPC27	Défense	2	0.111	0.64	0.22	0.14	0.071	0.024
FPC28	Défense	2	0.207	0.65	0.13	0.22	0.135	0.027
FPC29	Défense	2	0.134	0.73	0.13	0.14	0.098	0.017
FPC30	Défense	2	0.122	0.52	0.30	0.18	0.063	0.036

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

VI.7 « Bruit de fond » : Scénario 1

Présentation des données utilisées concernant le scénario 1 pour informer la variable « Bruit de fond » (en jaune), en fonction du participant qui est le POI.

n° échantillon	Proposition	POI	Quantité (en ng)	% POI	% AO	% Non_POI	Quantité Non_POI
MC1	Procureur	2	0.025	1.00	0.00	0.00	0.000
MC2	Procureur	2	0.025	1.00	0.00	0.00	0.000
MC4	Procureur	2	0.000	1.00	0.00	0.00	0.000
MC5	Procureur	2	2.278	1.00	0.00	0.00	0.000
MC6	Procureur	2	9.986	1.00	0.00	0.00	0.000
MC7	Procureur	2	0.110	0.00	0.00	0.00	0.000
MC8	Procureur	2	4.807	1.00	0.00	0.00	0.000
MC10	Procureur	2	8.522	1.00	0.00	0.00	0.000
MC12	Procureur	2	5.443	1.00	0.00	0.00	0.000
MC13	Procureur	2	5.257	1.00	0.00	0.00	0.000
MC14	Procureur	2	0.175	1.00	0.00	0.00	0.000
MC15	Procureur	2	0.300	1.00	0.00	0.00	0.000
MC16	Procureur	2	0.325	1.00	0.00	0.00	0.000
MC17	Procureur	2	0.100	1.00	0.00	0.00	0.000
MC18	Procureur	2	0.150	1.00	0.00	0.00	0.000
MC19	Procureur	2	0.100	1.00	0.00	0.00	0.000
MC20	Procureur	2	0.075	0.12	0.00	0.88	1.650
MC21	Procureur	2	0.200	1.00	0.00	0.00	0.000
MC22	Procureur	2	0.000	1.00	0.00	0.00	0.000
MC27	Procureur	2	0.100	1.00	0.00	0.00	0.000
MC28	Procureur	2	0.025	1.00	0.00	0.00	0.000
MC29	Procureur	2	0.025	1.00	0.00	0.00	0.000
MC30	Procureur	2	0.175	1.00	0.00	0.00	0.000
FC1	Procureur	6	1.625	1.00	0.00	0.00	0.000
FC2	Procureur	6	1.875	0.61	0.00	0.39	18.281
FC3	Procureur	6	3.025	0.49	0.00	0.51	38.569
FC4	Procureur	6	5.250	1.00	0.00	0.00	0.000
FC5	Procureur	6	1.600	0.84	0.00	0.16	6.400
FC6	Procureur	6	2.550	1.00	0.00	0.00	0.000
FC7	Procureur	6	2.475	0.78	0.00	0.22	13.613
FC8	Procureur	6	0.001	1.00	0.00	0.00	0.000
FC9	Procureur	6	0.100	1.00	0.00	0.00	0.000
FC10	Procureur	6	0.225	1.00	0.00	0.00	0.000
FC11	Procureur	6	0.250	1.00	0.00	0.00	0.000
FC12	Procureur	6	0.225	1.00	0.00	0.00	0.000
FC13	Procureur	6	0.225	1.00	0.00	0.00	0.000
FC14	Procureur	6	0.175	1.00	0.00	0.00	0.000
FC15	Procureur	6	0.100	1.00	0.00	0.00	0.000

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

FC16	Procureur	6	0.425	1.00	0.00	0.00	0.000
FC17	Procureur	6	0.775	1.00	0.00	0.00	0.000
FC18	Procureur	6	0.525	0.91	0.00	0.09	1.181
FC19	Procureur	6	3.975	1.00	0.00	0.00	0.000
FC20	Procureur	6	0.375	0.83	0.00	0.17	1.594
FC21	Procureur	6	0.225	0.85	0.00	0.15	0.844
FC22	Procureur	6	0.125	0.73	0.00	0.27	0.844
FC23	Procureur	6	0.225	0.57	0.00	0.43	2.419
FC24	Procureur	6	0.350	0.89	0.00	0.11	0.963
FC25	Procureur	6	0.250	1.00	0.00	0.00	0.000
FC26	Procureur	6	0.325	0.70	0.00	0.30	2.438
FC27	Procureur	6	0.525	0.83	0.00	0.17	2.231
FC28	Procureur	6	0.300	1.00	0.00	0.00	0.000
FC29	Procureur	6	0.125	0.88	0.00	0.12	0.375
FC30	Procureur	6	0.275	1.00	0.00	0.00	0.000
MTS1	Défense	6	10.900	0.00	0.25	0.75	204.375
MTS2	Défense	6	1.070	0.21	0.18	0.61	16.318
MTS3	Défense	6	0.733	0.58	0.00	0.42	7.697
MTS4	Défense	6	1.220	0.23	0.45	0.32	9.760
MTS5	Défense	6	0.093	0.12	0.01	0.87	2.029
MTS6	Défense	6	0.077	0.30	0.00	0.70	1.344
MTS7	Défense	6	0.076	0.15	0.36	0.49	0.935
MTS8	Défense	6	0.087	0.18	0.59	0.23	0.499
MTS9	Défense	6	0.092	0.34	0.22	0.44	1.012
MTS10	Défense	6	0.120	0.26	0.26	0.48	1.440
MTS11	Défense	6	0.077	0.25	0.33	0.42	0.803
MTS12	Défense	6	0.165	0.18	0.23	0.59	2.434
MTS14	Défense	6	0.074	0.15	0.18	0.67	1.245
MTS15	Défense	6	0.157	0.17	0.00	0.83	3.258
MTS17	Défense	6	0.225	0.13	0.45	0.42	2.363
MTS18	Défense	6	0.081	0.41	0.11	0.48	0.970
MTS19	Défense	6	0.109	0.00	0.00	1.00	2.725
MTS20	Défense	6	0.097	0.57	0.03	0.40	0.970
MTS21	Défense	6	0.034	0.14	0.03	0.83	0.712
MTS22	Défense	6	0.067	0.15	0.73	0.12	0.200
MTS23	Défense	6	0.047	0.37	0.11	0.52	0.605
MTS24	Défense	6	0.035	0.03	0.36	0.61	0.531
MTS25	Défense	6	0.024	0.00	0.51	0.49	0.299
MTS26	Défense	6	0.023	0.13	0.05	0.82	0.467
MTS27	Défense	6	0.112	0.00	0.23	0.77	2.156
MTS28	Défense	6	0.070	0.64	0.01	0.35	0.615
MTS29	Défense	6	0.030	0.00	0.19	0.81	0.603
MTS30	Défense	6	0.032	0.92	0.01	0.07	0.057
FTS1	Défense	2	0.128	0.00	1.00	0.00	0.000
FTS2	Défense	2	0.395	0.00	1.00	0.00	0.000



Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

FTS3	Défense	2	0.072	0.00	1.00	0.00	0.000
FTS4	Défense	2	0.103	0.00	1.00	0.00	0.000
FTS5	Défense	2	0.233	0.00	1.00	0.00	0.000
FTS6	Défense	2	0.064	0.00	1.00	0.00	0.000
FTS7	Défense	2	0.099	0.00	1.00	0.00	0.000
FTS8	Défense	2	0.081	0.00	1.00	0.00	0.000
FTS9	Défense	2	0.015	0.00	1.00	0.00	0.000
FTS10	Défense	2	0.095	0.00	1.00	0.00	0.000
FTS11	Défense	2	0.021	0.00	1.00	0.00	0.000
FTS12	Défense	2	0.027	0.00	1.00	0.00	0.000
FTS13	Défense	2	0.035	0.00	1.00	0.00	0.000
FTS14	Défense	2	0.082	0.00	1.00	0.00	0.000
FTS15	Défense	2	0.243	0.00	1.00	0.00	0.000
FTS16	Défense	2	0.051	0.00	1.00	0.00	0.000
FTS17	Défense	2	0.099	0.00	1.00	0.00	0.000
FTS18	Défense	2	0.122	0.11	0.61	0.28	0.854
FTS19	Défense	2	0.080	0.00	1.00	0.00	0.000
FTS20	Défense	2	0.044	0.15	0.50	0.35	0.381
FTS21	Défense	2	0.104	0.00	1.00	0.00	0.000
FTS22	Défense	2	0.092	0.00	0.46	0.54	1.246
FTS23	Défense	2	0.137	0.00	0.51	0.49	1.678
FTS24	Défense	2	0.048	0.00	1.00	0.00	0.000
FTS25	Défense	2	0.033	0.17	0.00	0.83	0.685
FTS26	Défense	2	0.116	0.00	0.00	1.00	2.900
FTS27	Défense	2	0.045	0.00	1.00	0.00	0.000
FTS28	Défense	2	0.170	0.00	1.00	0.00	0.000
FTS29	Défense	2	0.050	0.00	1.00	0.00	0.000
FTS30	Défense	2	0.192	0.00	1.00	0.00	0.000

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

VI.8 « Bruit de fond » : Scénario 3

Présentation des données utilisées concernant le scénario 3 pour informer la variable « Bruit de fond » (en jaune), en fonction du participant qui est le POI.

n° échantillon	Proposition	POI	Quantité (en ng)	% POI	% AO	% Non_POI	Quantité Non_POI
MC1	Procureur	2	0.025	1.00	0.00	0.00	0.000
MC2	Procureur	2	0.025	1.00	0.00	0.00	0.000
MC4	Procureur	2	0.000	1.00	0.00	0.00	0.000
MC5	Procureur	2	2.278	1.00	0.00	0.00	0.000
MC6	Procureur	2	9.986	1.00	0.00	0.00	0.000
MC7	Procureur	2	0.110	0.00	0.00	0.00	0.000
MC8	Procureur	2	4.807	1.00	0.00	0.00	0.000
MC10	Procureur	2	8.522	1.00	0.00	0.00	0.000
MC12	Procureur	2	5.443	1.00	0.00	0.00	0.000
MC13	Procureur	2	5.257	1.00	0.00	0.00	0.000
MC14	Procureur	2	0.175	1.00	0.00	0.00	0.000
MC15	Procureur	2	0.300	1.00	0.00	0.00	0.000
MC16	Procureur	2	0.325	1.00	0.00	0.00	0.000
MC17	Procureur	2	0.100	1.00	0.00	0.00	0.000
MC18	Procureur	2	0.150	1.00	0.00	0.00	0.000
MC19	Procureur	2	0.100	1.00	0.00	0.00	0.000
MC20	Procureur	2	0.075	0.12	0.00	0.88	1.650
MC21	Procureur	2	0.200	1.00	0.00	0.00	0.000
MC22	Procureur	2	0.000	1.00	0.00	0.00	0.000
MC27	Procureur	2	0.100	1.00	0.00	0.00	0.000
MC28	Procureur	2	0.025	1.00	0.00	0.00	0.000
MC29	Procureur	2	0.025	1.00	0.00	0.00	0.000
MC30	Procureur	2	0.175	1.00	0.00	0.00	0.000
FC1	Procureur	6	1.625	1.00	0.00	0.00	0.000
FC2	Procureur	6	1.875	0.61	0.00	0.39	18.281
FC3	Procureur	6	3.025	0.49	0.00	0.51	38.569
FC4	Procureur	6	5.250	1.00	0.00	0.00	0.000
FC5	Procureur	6	1.600	0.84	0.00	0.16	6.400
FC6	Procureur	6	2.550	1.00	0.00	0.00	0.000
FC7	Procureur	6	2.475	0.78	0.00	0.22	13.613
FC8	Procureur	6	0.001	1.00	0.00	0.00	0.000
FC9	Procureur	6	0.100	1.00	0.00	0.00	0.000
FC10	Procureur	6	0.225	1.00	0.00	0.00	0.000
FC11	Procureur	6	0.250	1.00	0.00	0.00	0.000
FC12	Procureur	6	0.225	1.00	0.00	0.00	0.000
FC13	Procureur	6	0.225	1.00	0.00	0.00	0.000
FC14	Procureur	6	0.175	1.00	0.00	0.00	0.000
FC15	Procureur	6	0.100	1.00	0.00	0.00	0.000

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

FC16	Procureur	6	0.425	1.00	0.00	0.00	0.000
FC17	Procureur	6	0.775	1.00	0.00	0.00	0.000
FC18	Procureur	6	0.525	0.91	0.00	0.09	1.181
FC19	Procureur	6	3.975	1.00	0.00	0.00	0.000
FC20	Procureur	6	0.375	0.83	0.00	0.17	1.594
FC21	Procureur	6	0.225	0.85	0.00	0.15	0.844
FC22	Procureur	6	0.125	0.73	0.00	0.27	0.844
FC23	Procureur	6	0.225	0.57	0.00	0.43	2.419
FC24	Procureur	6	0.350	0.89	0.00	0.11	0.963
FC25	Procureur	6	0.250	1.00	0.00	0.00	0.000
FC26	Procureur	6	0.325	0.70	0.00	0.30	2.438
FC27	Procureur	6	0.525	0.83	0.00	0.17	2.231
FC28	Procureur	6	0.300	1.00	0.00	0.00	0.000
FC29	Procureur	6	0.125	0.88	0.00	0.12	0.375
FC30	Procureur	6	0.275	1.00	0.00	0.00	0.000
MPC1	Défense	6	0.077	0.20	0.13	0.67	0.052
MPC2	Défense	6	0.156	0.80	0.16	0.04	0.006
MPC3	Défense	6	0.693	0.69	0.31	0.00	0.000
MPC4	Défense	6	1.040	0.64	0.35	0.01	0.010
MPC5	Défense	6	0.088	0.19	0.81	0.00	0.000
MPC6	Défense	6	0.795	0.59	0.15	0.26	0.207
MPC7	Défense	6	0.324	0.56	0.10	0.34	0.110
MPC8	Défense	6	0.086	0.47	0.21	0.32	0.028
MPC9	Défense	6	0.104	0.27	0.73	0.00	0.000
MPC10	Défense	6	0.121	0.72	0.08	0.20	0.024
MPC11	Défense	6	0.080	0.34	0.37	0.29	0.023
MPC12	Défense	6	0.246	0.56	0.27	0.17	0.042
MPC13	Défense	6	0.203	0.70	0.17	0.13	0.026
MPC14	Défense	6	0.265	0.23	0.14	0.63	0.167
MPC15	Défense	6	0.273	0.12	0.77	0.11	0.030
MPC16	Défense	6	0.327	0.29	0.36	0.35	0.114
MPC17	Défense	6	0.080	0.24	0.34	0.42	0.034
MPC18	Défense	6	0.151	0.09	0.77	0.14	0.021
MPC19	Défense	6	0.116	0.18	0.23	0.59	0.068
MPC20	Défense	6	0.178	0.21	0.24	0.55	0.098
MPC21	Défense	6	0.039	0.33	0.40	0.27	0.010
MPC22	Défense	6	0.120	0.52	0.13	0.35	0.042
MPC23	Défense	6	0.435	0.75	0.17	0.08	0.035
MPC24	Défense	6	0.623	0.37	0.56	0.07	0.044
MPC25	Défense	6	0.123	0.69	0.25	0.06	0.007
MPC26	Défense	6	0.072	0.24	0.13	0.63	0.045
MPC27	Défense	6	0.197	0.23	0.49	0.28	0.055
MPC28	Défense	6	0.210	0.20	0.26	0.54	0.113
MPC29	Défense	6	0.167	0.56	0.36	0.08	0.013
MPC30	Défense	6	0.291	0.57	0.43	0.00	0.000

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

FPC1	Défense	2	0.032	0.68	0.32	0.00	0.000
FPC2	Défense	2	0.225	0.61	0.27	0.12	0.027
FPC3	Défense	2	0.112	0.52	0.34	0.14	0.016
FPC4	Défense	2	0.078	0.35	0.29	0.36	0.028
FPC5	Défense	2	0.563	0.71	0.14	0.15	0.084
FPC6	Défense	2	0.593	0.66	0.31	0.03	0.018
FPC7	Défense	2	0.363	0.22	0.47	0.31	0.112
FPC8	Défense	2	0.115	0.60	0.18	0.22	0.025
FPC9	Défense	2	0.177	0.45	0.10	0.45	0.080
FPC10	Défense	2	0.085	0.24	0.27	0.49	0.042
FPC11	Défense	2	0.088	0.35	0.14	0.51	0.045
FPC12	Défense	2	0.082	0.63	0.23	0.14	0.011
FPC13	Défense	2	0.050	0.79	0.06	0.15	0.007
FPC14	Défense	2	0.350	0.84	0.16	0.00	0.000
FPC15	Défense	2	0.187	0.73	0.17	0.10	0.019
FPC16	Défense	2	0.084	0.58	0.13	0.29	0.024
FPC17	Défense	2	0.189	0.91	0.02	0.07	0.013
FPC18	Défense	2	0.428	0.23	0.67	0.10	0.043
FPC19	Défense	2	0.166	0.48	0.45	0.07	0.012
FPC20	Défense	2	0.181	0.52	0.48	0.00	0.000
FPC21	Défense	2	0.490	0.60	0.12	0.28	0.137
FPC22	Défense	2	0.253	0.70	0.22	0.08	0.020
FPC23	Défense	2	0.161	0.82	0.18	0.00	0.000
FPC24	Défense	2	0.109	0.33	0.66	0.01	0.001
FPC25	Défense	2	0.457	0.68	0.29	0.03	0.014
FPC26	Défense	2	0.535	0.65	0.19	0.16	0.086
FPC27	Défense	2	0.111	0.64	0.22	0.14	0.016
FPC28	Défense	2	0.207	0.65	0.13	0.22	0.045
FPC29	Défense	2	0.134	0.73	0.13	0.14	0.019
FPC30	Défense	2	0.122	0.52	0.30	0.18	0.022

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

Annexe VII: Résultats obtenus pour les cas reproduits

VII.1. Scénario 1 : Transfert secondaire allégué par la défense

Présentation des résultats de quantification, d'analyse du profil par STRmix et d'évaluation grâce à l'utilisation du BN pour chaque cas reproduit si la proposition  $H_p$  est vraie et si la proposition de la défense  $H_{d1}$  (impliquant un transfert secondaire) est vraie.

Bonne Proposition	Vrai agresseur	Quantité totale (en ng)	% POI	% AO	% Bruit de fond	Quantité POI (en ng)	Quantité Non POI (en ng)	LR	LR calibré	Ratio non POI/ POI
Procureur	6 « Bon donneur »	1.625	1.00	0.00	0.00	1.625	0.000	2E+09	3E+01	0,0
		0.225	1.00	0.00	0.00	0.225	0.000	1E+07	1E+01	0,0
		0.250	1.00	0.00	0.00	0.250	0.000	1E+07	1E+01	0,0
		0.225	1.00	0.00	0.00	0.225	0.000	1E+07	1E+01	0,0
		0.225	1.00	0.00	0.00	0.225	0.000	1E+07	1E+01	0,0
		0.175	1.00	0.00	0.00	0.175	0.000	9E+05	6E+00	0,0
		0.100	1.00	0.00	0.00	0.100	0.000	9E+05	6E+00	0,0
		0.425	1.00	0.00	0.00	0.425	0.000	5E+08	2E+01	0,0
		0.775	1.00	0.00	0.00	0.775	0.000	4E+09	3E+01	0,0
		0.525	0.91	0.00	0.09	0.478	0.047	8E+03	2E+00	0,1
		3.975	1.00	0.00	0.00	3.975	0.000	5E+07	1E+01	0,0
		1.875	0.61	0.00	0.39	1.144	0.731	1E+00	<b>4E-01</b>	0,6
		0.375	0.83	0.00	0.17	0.311	0.064	3E+02	1E+00	0,2
		0.225	0.85	0.00	0.15	0.191	0.034	4E+03	2E+00	0,2
		0.125	0.73	0.00	0.27	0.091	0.034	2E+03	2E+00	0,4
		0.225	0.57	0.00	0.43	0.128	0.097	3E+00	<b>5E-01</b>	0,8
		0.350	0.89	0.00	0.11	0.312	0.039	7E+04	4E+00	0,1
		0.250	1.00	0.00	0.00	0.250	0.000	1E+07	1E+01	0,0
		0.325	0.70	0.00	0.30	0.228	0.098	8E+00	<b>6E-01</b>	0,4
		0.525	0.83	0.00	0.17	0.436	0.089	1E+02	1E+00	0,2
		0.300	1.00	0.00	0.00	0.300	0.000	8E+07	1E+01	0,0
		0.125	0.88	0.00	0.12	0.110	0.015	9E+04	4E+00	0,1
		3.025	0.49	0.00	0.51	1.482	1.543	2E+00	<b>5E-01</b>	1,0
0.275	1.00	0.00	0.00	0.275	0.000	1E+07	1E+01	0,0		
5.250	1.00	0.00	0.00	5.250	0.000	3E+06	8E+00	0,0		
1.600	0.84	0.00	0.16	1.344	0.256	2E+00	<b>4E-01</b>	0,2		
2.550	1.00	0.00	0.00	2.550	0.000	2E+08	2E+01	0,0		
2.475	0.78	0.00	0.22	1.931	0.545	1E+00	<b>4E-01</b>	0,3		
0.001	1.00	0.00	0.00	0.001	0.000	2E+03	2E+00	1,0		
0.100	1.00	0.00	0.00	0.100	0.000	9E+05	6E+00	0,0		
Procureur	2 « Mauvais donneur »	0.025	1.00	0.00	0.00	0.025	0.000	2E+08	2E+01	0,0
		8.522	1.00	0.00	0.00	8.522	0.000	2E+07	1E+01	0,0
		5.443	1.00	0.00	0.00	5.443	0.000	7E+10	6E+01	0,0
		5.257	1.00	0.00	0.00	5.257	0.000	7E+10	6E+01	0,0
		0.175	1.00	0.00	0.00	0.175	0.000	6E+13	2E+02	0,0

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

		0.300	1.00	0.00	0.00	0.300	0.000	3E+18	2E+03	0,0
		0.325	1.00	0.00	0.00	0.325	0.000	3E+18	2E+03	0,0
		0.100	1.00	0.00	0.00	0.100	0.000	6E+13	2E+02	0,0
		0.150	1.00	0.00	0.00	0.150	0.000	6E+13	2E+02	0,0
		0.100	1.00	0.00	0.00	0.100	0.000	6E+13	2E+02	0,0
		0.025	1.00	0.00	0.00	0.025	0.000	2E+08	2E+01	0,0
		0.075	0.12	0.00	0.88	<b>0.009</b>	<b>0.066</b>	<b>3E-16</b>	<b>3E-04</b>	7,3
		0.200	1.00	0.00	0.00	0.200	0.000	7E+16	9E+02	0,0
		0.000	1.00	0.00	0.00	0.000	0.000	8E+04	4E+00	10,0
		0.100	1.00	0.00	0.00	0.100	0.000	6E+13	2E+02	0,0
		0.025	1.00	0.00	0.00	0.025	0.000	2E+08	2E+01	0,0
		0.025	1.00	0.00	0.00	0.025	0.000	2E+08	2E+01	0,0
		0.175	1.00	0.00	0.00	0.175	0.000	6E+13	2E+02	0,0
		0.000	1.00	0.00	0.00	0.000	0.000	8E+04	4E+00	10,0
		2.278	1.00	0.00	0.00	2.278	0.000	3E+14	3E+02	0,0
		9.986	1.00	0.00	0.00	9.986	0.000	2E+07	1E+01	0,0
		0.110	0.00	0.00	0.00	0.000	0.000	8E+04	4E+00	1,0
		4.807	1.00	0.00	0.00	4.807	0.000	7E+11	9E+01	0,0
Défense	2 « Mauvais donneur »	10.900	0.00	0.25	0.75	0.000	10.899	2E-02	2E-01	10899,4
		0.120	0.26	0.26	0.48	0.031	0.089	7E-01	4E-01	2,8
		0.077	0.25	0.33	0.42	<b>0.019</b>	<b>0.057</b>	<b>3E+00</b>	5E-01	3,0
		0.165	0.18	0.23	0.59	0.030	0.135	4E-02	2E-01	4,5
		0.074	0.15	0.18	0.67	<b>0.011</b>	<b>0.063</b>	<b>2E+00</b>	4E-01	5,7
		0.157	0.17	0.00	0.83	0.027	0.131	4E-02	2E-01	4,9
		0.225	0.13	0.45	0.42	0.029	0.196	4E-02	2E-01	6,7
		0.081	0.41	0.11	0.48	<b>0.033</b>	<b>0.048</b>	<b>2E+01</b>	7E-01	1,4
		0.109	0.00	0.00	1.00	0.000	0.109	2E-02	2E-01	108,5
		1.070	0.21	0.18	0.61	0.225	0.847	2E-01	3E-01	3,8
		0.097	0.57	0.03	0.40	<b>0.055</b>	<b>0.042</b>	<b>3E+01</b>	8E-01	0,8
		0.034	0.14	0.03	0.83	<b>0.005</b>	<b>0.029</b>	<b>1E+02</b>	1E+00	6,1
		0.067	0.15	0.73	0.12	<b>0.010</b>	<b>0.057</b>	<b>3E+00</b>	5E-01	5,7
		0.047	0.37	0.11	0.52	<b>0.017</b>	<b>0.029</b>	<b>4E+02</b>	1E+00	1,7
		0.035	0.03	0.36	0.61	<b>0.001</b>	<b>0.034</b>	<b>5E+01</b>	8E-01	32,3
		0.024	0.00	0.51	0.49	<b>0.000</b>	<b>0.024</b>	<b>1E+02</b>	1E+00	24,4
		0.023	0.13	0.05	0.82	<b>0.003</b>	<b>0.020</b>	<b>2E+02</b>	1E+00	6,7
		0.112	0.00	0.23	0.77	0.000	0.112	2E-02	2E-01	111,5
		0.070	0.64	0.01	0.35	<b>0.045</b>	<b>0.025</b>	<b>2E+03</b>	<b>2E+00</b>	0,6
		0.030	0.00	0.19	0.81	<b>0.000</b>	<b>0.030</b>	<b>1E+02</b>	1E+00	29,8
		0.733	0.58	0.00	0.42	0.425	0.308	3E-01	3E-01	0,7
		0.032	0.92	0.01	0.07	<b>0.030</b>	<b>0.003</b>	<b>1E+04</b>	<b>3E+00</b>	0,1
		1.220	0.23	0.45	0.32	0.281	0.936	2E-01	3E-01	3,3
		0.093	0.12	0.01	0.87	0.011	0.082	5E-01	3E-01	7,3
		0.077	0.30	0.00	0.70	<b>0.023</b>	<b>0.054</b>	<b>4E+00</b>	5E-01	2,3
		0.076	0.15	0.36	0.49	<b>0.011</b>	<b>0.065</b>	<b>2E+00</b>	4E-01	5,7
0.087	0.18	0.59	0.23	0.016	0.071	9E-01	4E-01	4,6		
0.092	0.34	0.22	0.44	<b>0.031</b>	<b>0.061</b>	<b>2E+00</b>	5E-01	1,9		
		0.128	0.00	1.00	0.00	0.000	0.128	1E-18	1E-04	128,0

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

Défense	6 « Bon donneur »	0.095	0.00	1.00	0.00	0.000	0.095	9E-18	2E-04	94,8
		0.021	0.00	1.00	0.00	0.000	0.021	3E-04	8E-02	21,2
		0.027	0.00	1.00	0.00	0.000	0.027	3E-04	8E-02	26,8
		0.035	0.00	1.00	0.00	0.000	0.035	6E-09	8E-03	35,0
		0.082	0.00	1.00	0.00	0.000	0.082	2E-17	2E-04	81,5
		0.243	0.00	1.00	0.00	0.000	0.243	3E-19	8E-05	243,3
		0.051	0.00	1.00	0.00	0.000	0.051	2E-15	4E-04	51,3
		0.099	0.00	1.00	0.00	0.000	0.099	9E-18	2E-04	99,2
		0.122	0.11	0.61	0.28	0.013	0.108	8E-18	1E-04	8,1
		0.080	0.00	1.00	0.00	0.000	0.080	2E-17	2E-04	80,0
		0.395	0.00	1.00	0.00	0.000	0.395	2E-19	7E-05	395,0
		0.044	0.15	0.50	0.35	0.022	0.037	1E-06	2E-02	1,7
		0.104	0.00	1.00	0.00	0.000	0.104	1E-18	1E-04	103,5
		0.092	0.00	0.46	0.54	0.042	0.092	2E-15	4E-04	2,2
		0.137	0.00	0.51	0.49	0.000	0.137	1E-18	1E-04	136,5
		0.048	0.00	1.00	0.00	0.000	0.048	1E-13	1E-03	47,8
		0.033	0.17	0.00	0.83	0.006	0.027	3E-04	8E-02	4,9
		0.116	0.00	0.00	1.00	0.000	0.116	1E-18	1E-04	116,3
		0.045	0.00	1.00	0.00	0.000	0.045	1E-13	1E-03	44,8
		0.170	0.00	1.00	0.00	0.000	0.170	1E-18	1E-04	170,3
		0.050	0.00	1.00	0.00	0.000	0.050	2E-15	4E-04	50,3
		0.072	0.00	1.00	0.00	0.000	0.072	6E-17	2E-04	72,3
		0.192	0.00	1.00	0.00	0.000	0.192	1E-18	1E-04	191,5
		0.103	0.00	1.00	0.00	0.000	0.103	1E-18	1E-04	103,0
		0.233	0.00	1.00	0.00	0.000	0.233	3E-19	8E-05	233,3
		0.064	0.00	1.00	0.00	0.000	0.064	3E-16	3E-04	64,4
		0.099	0.00	1.00	0.00	0.000	0.099	9E-18	2E-04	99,2
		0.081	0.00	1.00	0.00	0.000	0.081	2E-17	2E-04	81,0
0.015	0.00	1.00	0.00	<b>0.000</b>	<b>0.015</b>	<b>9E+01</b>	1E+00	15,1		

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

VII.2. Scénario 3 : Utilisation légitime du couteau alléguée par la défense

Présentation des résultats de quantification, d'analyse du profil par STRmix et d'évaluation grâce à l'utilisation du BN pour chaque cas reproduit si la proposition H<sub>p</sub> est vraie et si la proposition de la défense H<sub>d3</sub> (impliquant une utilisation légitime du couteau) est vraie.

Bonne Proposition	Vrai agresseur	Quantité totale (en ng)	% POI	% AO	% Bruit de fond	Quantité POI (en ng)	Quantité Non POI (en ng)	LR	LR calibré	Ratio Non POI/ POI
Procureur	6 « Bon donneur »	1,625	1,00	0,00	0,00	1,625	0,000	2E+12	8E+01	0,0
		0,225	1,00	0,00	0,00	0,225	0,000	3E+07	6E+00	0,0
		0,250	1,00	0,00	0,00	0,250	0,000	3E+07	6E+00	0,0
		0,225	1,00	0,00	0,00	0,225	0,000	3E+07	6E+00	0,0
		0,225	1,00	0,00	0,00	0,225	0,000	3E+07	6E+00	0,0
		0,175	1,00	0,00	0,00	0,175	0,000	2E+06	3E+00	0,0
		0,100	1,00	0,00	0,00	0,100	0,000	2E+06	3E+00	0,0
		0,425	1,00	0,00	0,00	0,425	0,000	2E+09	2E+01	0,0
		0,775	1,00	0,00	0,00	0,775	0,000	3E+10	3E+01	0,0
		0,525	0,91	0,00	0,09	0,478	0,000	3E+04	1E+00	0,1
		3,975	1,00	0,00	0,00	3,975	0,000	1E+10	3E+01	0,0
		1,875	0,61	0,00	0,39	1,144	0,000	<b>5E-01</b>	<b>9E-02</b>	0,6
		0,375	0,83	0,00	0,17	0,311	0,000	9E+02	<b>5E-01</b>	0,2
		0,225	0,85	0,00	0,15	0,191	0,000	2E+04	1E+00	0,2
		0,125	0,73	0,00	0,27	0,091	0,000	6E+03	<b>9E-01</b>	0,4
		0,225	0,57	0,00	0,43	0,128	0,000	8E+00	<b>2E-01</b>	0,8
		0,350	0,89	0,00	0,11	0,312	0,000	4E+05	2E+00	0,1
		0,250	1,00	0,00	0,00	0,250	0,000	3E+07	6E+00	0,0
		0,325	0,70	0,00	0,30	0,228	0,000	2E+01	<b>2E-01</b>	0,4
		0,525	0,83	0,00	0,17	0,436	0,000	3E+02	4E-01	0,2
		0,300	1,00	0,00	0,00	0,300	0,000	3E+08	1E+01	0,0
		0,125	0,88	0,00	0,12	0,110	0,000	2E+05	2E+00	0,1
		3,025	0,49	0,00	0,51	1,482	0,000	<b>7E-01</b>	<b>1E-01</b>	1,0
		0,275	1,00	0,00	0,00	0,275	0,000	3E+07	6E+00	0,0
		5,250	1,00	0,00	0,00	5,250	0,000	7E+08	1E+01	0,0
		1,600	0,84	0,00	0,16	1,344	0,000	1E+00	<b>1E-01</b>	0,2
2,550	1,00	0,00	0,00	2,550	0,000	8E+10	4E+01	0,0		
2,475	0,78	0,00	0,22	1,931	0,000	<b>6E-01</b>	<b>1E-01</b>	0,3		
0,001	1,00	0,00	0,00	0,001	0,000	1E+04	1E+00	1,0		
0,100	1,00	0,00	0,00	0,100	0,000	2E+06	3E+00	0,0		
Procureur	2 « Mauvais donneur »	0,025	1,00	0,00	0,00	0,025	0,000	1E+06	3E+00	0,0
		8,522	1,00	0,00	0,00	8,522	0,000	2E+07	6E+00	0,0
		5,443	1,00	0,00	0,00	5,443	0,000	5E+10	4E+01	0,0
		5,257	1,00	0,00	0,00	5,257	0,000	5E+10	4E+01	0,0
		0,175	1,00	0,00	0,00	0,175	0,000	1E+09	1E+01	0,0
		0,300	1,00	0,00	0,00	0,300	0,000	6E+13	2E+02	0,0
0,325	1,00	0,00	0,00	0,325	0,000	6E+13	2E+02	0,0		



Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

		0,100	1.00	0.00	0.00	0.100	0.000	1E+09	1E+01	0,0
		0,150	1.00	0.00	0.00	0.150	0.000	1E+09	1E+01	0,0
		0,100	1.00	0.00	0.00	0.100	0.000	1E+09	1E+01	0,0
		0,025	1.00	0.00	0.00	0.025	0.000	1E+06	3E+00	0,0
		0,075	0.12	0.00	0.88	<b>0.009</b>	<b>0.066</b>	4E+01	<b>3E-01</b>	7,3
		0,200	1.00	0.00	0.00	0.200	0.000	6E+11	7E+01	0,0
		0,000	1.00	0.00	0.00	0.000	0.000	2E+05	2E+00	10,0
		0,100	1.00	0.00	0.00	0.100	0.000	1E+09	1E+01	0,0
		0,025	1.00	0.00	0.00	0.025	0.000	1E+06	3E+00	0,0
		0,025	1.00	0.00	0.00	0.025	0.000	1E+06	3E+00	0,0
		0,175	1.00	0.00	0.00	0.175	0.000	1E+09	1E+01	0,0
		0,000	1.00	0.00	0.00	0.000	0.000	2E+05	2E+00	10,0
		2,278	1.00	0.00	0.00	2.278	0.000	1E+15	4E+02	0,0
		9,986	1.00	0.00	0.00	9.986	0.000	6E+05	3E+00	0,0
		0,110	0.00	0.00	0.00	0.000	0.000	2E+05	2E+00	1,0
		4,807	1.00	0.00	0.00	4.807	0.000	1E+11	4E+01	0,0
Défense	2 « Mauvais donneur »	0,077	0,20	0,13	0,67	0,015	0,010	<b>2E+01</b>	2E-01	4,0
		0,121	0,72	0,08	0,20	0,087	0,010	<b>5E+03</b>	8E-01	0,4
		0,080	0,34	0,37	0,29	0,027	0,030	<b>3E+01</b>	2E-01	1,9
		0,246	0,56	0,27	0,17	0,138	0,066	9E-01	1E-01	0,8
		0,203	0,70	0,17	0,13	0,142	0,034	<b>7E+01</b>	3E-01	0,4
		0,265	0,23	0,14	0,63	0,061	0,037	2E-01	7E-02	3,3
		0,273	0,12	0,77	0,11	0,033	0,210	1E-01	7E-02	7,3
		0,327	0,29	0,36	0,35	0,095	0,118	2E-01	7E-02	2,4
		0,080	0,24	0,34	0,42	0,019	0,027	<b>2E+01</b>	2E-01	3,2
		0,151	0,09	0,77	0,14	0,014	0,116	4E-01	9E-02	10,1
		0,116	0,18	0,23	0,59	0,021	0,027	<b>3E+00</b>	1E-01	4,6
		0,156	0,80	0,16	0,04	0,125	0,025	<b>2E+04</b>	1E+00	0,3
		0,178	0,21	0,24	0,55	0,037	0,043	4E-01	9E-02	3,8
		0,039	0,33	0,40	0,27	0,013	0,015	<b>1E+03</b>	6E-01	2,0
		0,120	0,52	0,13	0,35	0,062	0,016	<b>5E+01</b>	3E-01	0,9
		0,435	0,75	0,17	0,08	0,326	0,074	<b>2E+00</b>	1E-01	0,3
		0,623	0,37	0,56	0,07	0,231	0,349	2E-01	7E-02	1,7
		0,123	0,69	0,25	0,06	0,085	0,031	<b>5E+03</b>	8E-01	0,4
		0,072	0,24	0,13	0,63	0,017	0,009	<b>3E+01</b>	2E-01	3,2
		0,197	0,23	0,49	0,28	0,045	0,097	5E-01	9E-02	3,3
		0,210	0,20	0,26	0,54	0,042	0,055	5E-01	9E-02	4,0
		0,167	0,56	0,36	0,08	0,094	0,060	<b>2E+01</b>	2E-01	0,8
		0,693	0,69	0,31	0,00	0,478	0,215	4E-01	9E-02	0,4
		0,291	0,57	0,43	0,00	0,166	0,125	9E-01	1E-01	0,8
		1,040	0,64	0,35	0,01	0,666	0,365	4E-01	9E-02	0,6
		0,088	0,19	0,81	0,00	0,017	0,071	<b>9E+00</b>	2E-01	4,3
		0,795	0,59	0,15	0,26	0,469	0,119	3E-01	8E-02	0,7
		0,324	0,56	0,10	0,34	0,181	0,032	9E-01	1E-01	0,8
0,086	0,47	0,21	0,32	0,040	0,018	<b>1E+02</b>	3E-01	1,1		
0,104	0,27	0,73	0,00	0,028	0,076	<b>8E+00</b>	2E-01	2,7		
		0,032	0,68	0,32	0,00	0,021	0,010	<b>1E+05</b>	<b>2E+00</b>	0,5

Evaluation des résultats ADN considérant des propositions au niveau de l'activité  
2019

Défense	6 « Bon donneur »	0,085	0,24	0,27	0,49	0,020	0,023	<b>4E+01</b>	3E-01	3,2
		0,088	0,35	0,14	0,51	0,031	0,012	<b>1E+02</b>	3E-01	1,9
		0,082	0,63	0,23	0,14	0,052	0,019	<b>4E+04</b>	1E+00	0,6
		0,050	0,79	0,06	0,15	0,039	0,003	<b>3E+05</b>	<b>2E+00</b>	0,3
		0,350	0,84	0,16	0,00	0,294	0,056	<b>4E+04</b>	1E+00	0,2
		0,187	0,73	0,17	0,10	0,137	0,032	<b>2E+03</b>	7E-01	0,4
		0,084	0,58	0,13	0,29	0,049	0,011	<b>3E+04</b>	1E+00	0,7
		0,189	0,91	0,02	0,07	0,172	0,004	<b>1E+08</b>	<b>9E+00</b>	0,1
		0,428	0,23	0,67	0,10	0,098	0,286	1E-04	1E-02	3,3
		0,166	0,48	0,45	0,07	0,080	0,075	<b>2E+01</b>	2E-01	1,1
		0,225	0,61	0,27	0,12	0,137	0,061	<b>6E+01</b>	3E-01	0,6
		0,181	0,52	0,48	0,00	0,094	0,087	<b>3E+01</b>	2E-01	0,9
		0,490	0,60	0,12	0,28	0,294	0,059	4E-01	9E-02	0,7
		0,253	0,70	0,22	0,08	0,177	0,056	<b>2E+02</b>	4E-01	0,4
		0,161	0,82	0,18	0,00	0,132	0,029	<b>4E+07</b>	<b>7E+00</b>	0,2
		0,109	0,33	0,66	0,01	0,036	0,072	<b>2E+01</b>	2E-01	2,0
		0,457	0,68	0,29	0,03	0,311	0,133	9E-01	1E-01	0,5
		0,535	0,65	0,19	0,16	0,348	0,102	9E-01	1E-01	0,5
		0,111	0,64	0,22	0,14	0,071	0,024	<b>3E+03</b>	7E-01	0,6
		0,207	0,65	0,13	0,22	0,135	0,027	<b>2E+02</b>	4E-01	0,5
		0,134	0,73	0,13	0,14	0,098	0,017	<b>2E+05</b>	<b>2E+00</b>	0,4
		0,112	0,52	0,34	0,14	0,058	0,038	<b>2E+02</b>	4E-01	0,9
		0,122	0,52	0,30	0,18	0,063	0,036	<b>3E+02</b>	4E-01	0,9
		0,078	0,35	0,29	0,36	0,027	0,023	<b>1E+02</b>	3E-01	1,9
		0,563	0,71	0,14	0,15	0,400	0,079	9E-01	1E-01	0,4
		0,593	0,66	0,31	0,03	0,391	0,184	4E-03	3E-02	0,5
		0,363	0,22	0,47	0,31	0,080	0,170	1E-03	2E-02	3,5
		0,115	0,60	0,18	0,22	0,069	0,021	<b>2E+03</b>	7E-01	0,7
		0,177	0,45	0,10	0,45	0,080	0,018	<b>4E+00</b>	2E-01	1,2





