



**HAL**  
open science

## Estimation du nombre de blessés graves de la route, au sens MAIS3+, France métropolitaine, 2006-2016

Emmanuelle Amoros, Laura Turbatu, Grégory Soler, Léa Pascal, Blandine Gadegbeku, Amina Ndiaye, Bernard Laumon, Jean-Louis Martin

### ► To cite this version:

Emmanuelle Amoros, Laura Turbatu, Grégory Soler, Léa Pascal, Blandine Gadegbeku, et al.. Estimation du nombre de blessés graves de la route, au sens MAIS3+, France métropolitaine, 2006-2016. [Rapport de recherche] IFSTTAR - Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux. 2019, 105 p. hal-02110133

**HAL Id: hal-02110133**

**<https://hal.science/hal-02110133>**

Submitted on 25 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Estimation du nombre de blessés graves de la route, au sens MAIS3+, France métropolitaine, 2006-2016**

Emmanuelle Amoros, Léa Pascal, Grégory Soler, Laura Turbatu,  
Blandine Gadegbeku, Amina Ndiaye, Bernard Laumon, Jean-louis Martin

Unité Mixte de Recherche Épidémiologique et de Surveillance Transport Travail  
Environnement (UMRESTTE),  
Département Transport, Santé, Sécurité (TS2)  
IFSTTAR  
Université Lyon 1

Convention REG-Trauma IFSTTAR-DSCR n° 2200979046

Janvier 2019

**Résumé :**

En termes de blessés de la route, la Commission Européenne demande dorénavant à ses États Membres d'estimer le nombre de blessés graves selon un critère médical, le MAIS3+ (Maximum Abbreviated Injury Score).

Le Registre du Rhône des victimes d'accidents de la circulation fournit directement ce critère, et permet par modélisations successives, basées sur la coexistence de ce Registre et des données des forces de l'ordre, d'estimer le nombre de blessés graves MAIS3+ au niveau national.

Il est ainsi estimé à 25 400 en 2016, et correspond à 7 fois le nombre de tués en moyenne chaque année (sur 2006-2016). Le nombre de blessés graves à deux-roues motorisé est depuis 2004 supérieur à celui des automobilistes : ils sont respectivement 10 000/an et 6 500/an sur 2013-2016, alors que les tués sont au nombre de 800/an et 1 800/an respectivement. Les piétons et cyclistes blessés MAIS3+ sont tous les deux au nombre de 3 500 /an (sur 2013-2016), alors que les tués sont respectivement 500/an et 150/an.

Le nombre de blessés, tous usagers et toutes gravités, est estimé à 298 300 (2016).

**Mots clés :** sécurité routière ; traumatologie routière ; blessés de la route ; données policières ; Registre médical ; sous-enregistrement, biais, prédiction ; capture-recapture ; projection ; Abbreviated Injury Scale, MAIS3+

RESUME LONG .....	5
<b>1 INTRODUCTION.....</b>	<b>13</b>
<b>2 MATERIEL ET METHODES .....</b>	<b>14</b>
<b>2.1 PRESENTATION DES BASES DE DONNEES.....</b>	<b>14</b>
2.1.1 <i>Le Registre du Rhône .....</i>	14
2.1.2 <i>Les Bulletins d'Analyse des Accidents Corporels (BAAC).....</i>	14
<b>2.2 METHODOLOGIE DE L'EXTRAPOLATION DES BAAC EN 2 ETAPES .....</b>	<b>14</b>
<b>2.3 ÉTAPE 1 : PREDICTION DU NOMBRE DE BLESSES MAIS 3+ SUR LES BLESSES BAAC .....</b>	<b>16</b>
<b>2.4 ÉTAPE 1A : CONSTRUCTION DU MODELE DE PREDICTION DU MAIS .....</b>	<b>16</b>
2.4.1 <i>Données.....</i>	16
2.4.2 <i>Modélisation.....</i>	18
2.4.3 <i>Variables du modèle.....</i>	19
2.4.4 <i>Modèle obtenu .....</i>	19
<b>2.5 ÉTAPE 1B : APPLICATION DU MODELE DE PREDICTION DU MAIS AUX BAAC .....</b>	<b>24</b>
2.5.1 <i>Étape 1B : Application du modèle de prédiction du MAIS aux BAAC du Rhône.....</i>	24
2.5.2 <i>Étape 1B : Application du modèle de prédiction du MAIS aux BAAC France .....</i>	24
<b>2.6 ÉTAPE 2 - PREALABLE : SITUATION DE L'ENREGISTREMENT DANS LE RHONE .....</b>	<b>24</b>
<b>2.7 ÉTAPE 2 : METHODE DE CORRECTION DU SOUS-ENREGISTREMENT ET DE SES BIAIS .....</b>	<b>29</b>
<b>2.8 ÉTAPE 2A : AMELIORATION DU CHAINAGE ENTRE BAAC ET REGISTRE, RHONE .....</b>	<b>30</b>
2.8.1 <i>Redéfinition des effectifs résultant du chaînage .....</i>	33
<b>2.9 ÉTAPE 2B : MODELISATION DU SOUS-ENREGISTREMENT PAR CAPTURE-RECAPTURE.....</b>	<b>35</b>
2.9.1 <i>Synthèse bibliographique sur capture-recapture .....</i>	35
2.9.2 <i>Conditions d'application.....</i>	37
2.9.3 <i>Construction du modèle (logit multinomial).....</i>	40
2.9.4 <i>Modèle d'enregistrement des blessés sur 2006-2012 .....</i>	40
<b>2.10 ÉTAPE 2C : PROJECTION DU DEPARTEMENT DU RHONE A LA FRANCE ENTIERE .....</b>	<b>48</b>
<b>3 RESULTATS .....</b>	<b>50</b>
<b>3.1 NOMBRE DE BLESSES TOUTES CATEGORIES .....</b>	<b>50</b>
3.1.1 <i>Nombre de blessés toutes gravités (MAIS1+).....</i>	50
3.1.2 <i>Nombre de blessés graves MAIS3+.....</i>	51
3.1.3 <i>Proportion de blessés MAIS3+ parmi les blessés MAIS1+.....</i>	52
3.1.4 <i>Ratio Nombre de blesses MAIS3+/Nombre de tués.....</i>	52
<b>3.2 BLESSES SELON LE TYPE D'USAGER.....</b>	<b>53</b>
3.2.1 <i>Nombre de blessés toutes gravités (MAIS1+).....</i>	53
3.2.2 <i>Nombre de blessés graves (MAIS3+) .....</i>	55
3.2.3 <i>Proportion de blessés MAIS3+ pour les blessés MAIS1+.....</i>	58
<b>3.3 BLESSES SELON LE SEXE.....</b>	<b>60</b>
3.3.1 <i>Nombre de blessés toutes gravités (MAIS1+).....</i>	60
3.3.2 <i>Nombre de blessés graves (MAIS3+) .....</i>	61
3.3.3 <i>Proportion de blessés MAIS3+ parmi les blessés MAIS1+.....</i>	62
<b>3.4 BLESSES SELON L'AGE.....</b>	<b>63</b>
3.4.1 <i>Nombre de blessés toutes gravités (MAIS1+).....</i>	63
3.4.2 <i>Nombre de blessés graves (MAIS3+) .....</i>	63
3.4.3 <i>Proportion de blessés MAIS3+ parmi les blessés MAIS1+.....</i>	64
<b>4 DISCUSSION .....</b>	<b>68</b>
<b>4.1 ÉLÉMENTS DE VALIDITE DES RESULTATS .....</b>	<b>68</b>

4.1.1	Validation interne : Gravité selon le Registre du Rhône .....	69
4.1.2	Validation interne : Gravité parmi les blessés non observés .....	69
4.1.3	Validation externe : l'Enquête Nationale Transport et Déplacements (ENTD) .....	70
4.1.4	Validation externe : Effectifs de blessés selon le SDIS= Service Départemental d'Incendie et de Secours.....	71
4.1.5	Validation externe : ratio blessés toutes gravités /tués selon les pays .....	71
4.1.6	Validation externe : ratio blessés graves MAIS3+/tués selon les pays .....	72
<b>4.2</b>	<b>LIMITES .....</b>	<b>73</b>
4.2.1	Manque de fiabilité du nombre estimé d'hospitalisés.....	73
4.2.2	Nombre de blessés hospitalisés à partir du PMSI.....	73
4.2.3	Effets du Registre sur l'enregistrement par les forces de l'ordre du Rhône ?.....	73
<b>5</b>	<b>REMERCIEMENTS .....</b>	<b>75</b>
<b>6</b>	<b>REFERENCES .....</b>	<b>76</b>
<b>7</b>	<b>ANNEXES .....</b>	<b>81</b>
	<b>ANNEXE 1 : ESTIMATION DU NOMBRE DE FAUX POSITIFS ET FAUX NEGATIFS LIE AU CHAINAGE BAAC ET REGISTRE DU RHONE .....</b>	<b>82</b>
	7.1.1 Estimation du nombre de faux positifs.....	82
	7.1.2 Estimation du nombre de faux négatifs .....	86
	<b>ANNEXE 2 : MODELISATION MULTIVARIEE DE L'APPROCHE CAPTURE-RECAPTURE.....</b>	<b>89</b>
	7.1.3 Sous l'hypothèse d'indépendance des deux sources .....	89
	7.1.4 Sous l'hypothèse de dépendance des deux sources.....	92
	<b>ANNEXE 3 : METHODES DE REDRESSEMENT, DANS LE CADRE DE LA THEORIE DES SONDAGES .....</b>	<b>97</b>
	<b>ANNEXE 4 : SYNTHESE BIBLIOGRAPHIQUE SUR LA PROJECTION, DU REGIONAL AU NATIONAL, EN EPIDEMIOLOGIE</b>	<b>99</b>
	<b>ANNEXE 5 - LE CAS PARTICULIER DE LA VARIABLE HOSPITALISE (OUI/NON) .....</b>	<b>103</b>
	<b>ANNEXE 6 : RESUME DES DIFFERENTES APPROCHES TESTEES ET CRITERES DE SELECTION ET DE VALIDATION DE L'APPROCHE. ....</b>	<b>104</b>

## RÉSUMÉ LONG

La Commission Européenne demande à ses États Membres d'estimer annuellement le nombre de blessés graves suite à un accident de la circulation routière, avec le niveau de gravité apprécié à partir de la classification AIS, Abbreviated Injury Scale, échelle traumatologique internationale. L'AIS code la gravité immédiate sur une échelle de 1 à 6 (1=gravité mineure, 2= modérée, 3= sérieuse, 4= sévère, 5 = critique, 6 = au-delà des ressources thérapeutiques). Le blessé grave a été défini comme une victime, non décédée à 30 jours, ayant au moins une blessure de gravité 3 ou plus (MAIS3+).

**Matériel** Les données des forces de l'ordre, disponibles à l'échelle nationale, ont deux inconvénients principaux : un sous-enregistrement des blessés (qui, de plus, dépend de plusieurs facteurs), et une information insuffisante pour apprécier la gravité des blessures (limitée à tué/hospitalisé/non hospitalisé).

Le sous-enregistrement des blessés de la route par les forces de l'ordre est commun à tous les pays. En France, il est estimé que seuls environ 25% des blessés sont enregistrés dans les BAAC (Bulletins d'Analyse d'Accidents de la Circulation). Ce sous-enregistrement est de plus biaisé, c'est-à-dire qu'il dépend de certaines caractéristiques des accidents. Ainsi, les cyclistes blessés sont particulièrement mal enregistrés ; les usagers de 2RM aussi, mais dans une moindre mesure. Les accidents sans tiers sont moins bien enregistrés que les accidents avec tiers, et les accidents les plus graves sont mieux enregistrés que les moins graves.

Dans le département du Rhône, nous disposons depuis 1996 du Registre (médical) des victimes d'accidents de la circulation routière. Il se base sur tous les services hospitaliers, publics et privés, du Rhône ou limitrophes, accueillant des victimes d'accidents de la route (environ 250 services, depuis les SAMU, SMUR, déchocage, réanimation, chirurgie jusqu'aux services de rééducation). En outre, toutes les lésions de toutes les victimes sont codées directement avec l'échelle AIS.

Quand les forces de l'ordre dénombrent environ 2 800 blessés par an (sur 2006-2012 sur le département du Rhône), le Registre en compte 7 400 (et 1 700 sont identifiés communs aux deux).

**Méthodes** La coexistence de ces deux sources de données dans le Rhône, permet par une approche capture-recapture, sous certaines conditions, de corriger le sous-enregistrement des BAAC et de ses biais. Le processus utilisé consiste, à partir des données du Rhône, à modéliser la probabilité d'enregistrement dans les BAAC selon tous les facteurs identifiés comme liés au sous-enregistrement, puis d'appliquer ce modèle à toute la France métropolitaine.

Parmi ces différents facteurs, la gravité joue un rôle central. Il reste à la mesurer. La variable hospitalisé (oui/non) présente l'avantage d'être définie dans les deux sources BAAC et Registre, mais n'est pas suffisamment informative ni complètement fiable (18% de discordance entre les deux sources). En plus de cette variable, nous avons voulu utiliser le MAIS. Nous avons donc construit un modèle de prédiction de la probabilité d'être MAIS3+ pour les blessés dans les BAAC. Un tel modèle a pu être construit sur les blessés identifiés à la fois par les BAAC et le Registre, en utilisant les informations des BAAC et la gravité MAIS connue dans le Registre. L'objectif de cette modélisation n'est pas de prédire le MAIS au niveau individuel, mais globalement en termes d'effectifs, selon des profils d'accidents et des caractéristiques des impliqués.

Pour cette étape, nous avons utilisé un modèle multivarié et retenu les variables statistiquement significatives, et dont l'influence est d'ampleur non négligeable sur la probabilité pour une victime d'être MAIS3+. Parmi les variables disponibles dans les BAAC, celles associées à la gravité MAIS3+ et retenues sont les suivantes :

- Lieu d'accident (en/ hors agglomération)

- Type de force de l'ordre renseignant l'accident (CRS/ Gendarmerie / Police) et interaction avec le lieu
- Gravité de l'accident (mortel ou corporel) dans lequel le blessé est impliqué
- Hospitalisé (oui/ non) et interaction avec le type de Force de l'Ordre renseignant l'accident
- Type d'usager (piéton/ cycliste/ motocycliste / automobiliste / usager autre véhicule)
- Implication (oui / non) d'un tiers dans l'accident
- Interaction entre type d'usager, implication d'un tiers et hospitalisé (oui /non)
- Âge en 8 catégories
- Sexe.

Le modèle ainsi construit a ensuite été appliqué aux blessés des BAAC du Rhône et des BAAC nationaux. À noter que l'application aux BAAC nationaux suppose que, pour un type de force de l'ordre et une gravité, un type d'usager, un tiers,... donnés, les pratiques d'enregistrement des informations soient homogènes sur le territoire, notamment le renseignement sur l'hospitalisation.

L'application du modèle de prédiction du MAIS sur les blessés BAAC du Rhône permet de disposer au niveau rhodanien du MAIS pour les blessés des deux sources (BAAC et Registre) et de l'utiliser dans l'étape suivante.

Cette deuxième étape consiste à modéliser la probabilité pour un blessé d'être enregistré dans les BAAC et/ou dans le Registre par une approche dite de capture – recapture. Cette approche nécessite que soient remplies certaines conditions (détaillées dans le rapport), en particulier que les individus aient la même probabilité d'être enregistrés, ou, quand cela n'est pas le cas comme ici, de tenir compte des variables qui influencent cette probabilité

Les variables corrélées à cette probabilité d'enregistrement et retenues sont les suivantes:

- le type de réseau où s'est produit l'accident (autoroute / RN ou RD / voies communale)
- la gravité de l'accident : mortel ou corporel
- l'hospitalisation (oui/non) selon les BAAC ou selon le Registre
- le type de Force de l'ordre en charge de la zone où s'est produit l'accident
- la gravité selon le MAIS (1-2/3+) connu ou prédit
- l'interaction entre les 2 variables ci-dessus,
- le type d'usager (piéton / cycliste / motocycliste / automobiliste / autre)
- l'implication (oui/non) d'un tiers dans l'accident
- l'interaction entre le type d'usager, l'implication d'un tiers et la gravité MAIS
- l'année en quantitatif pour la zone gendarmerie

Ce modèle, construit au niveau du département du Rhône, nous donne le nombre total de blessés (MAIS3+ et MAIS1+) du Rhône et, par suite, les coefficients de correction du sous-enregistrement des BAAC du Rhône et de leur biais. Sous l'hypothèse d'une certaine homogénéité sur le territoire métropolitain des pratiques d'enregistrement des blessés pour un type de force de l'ordre donnée, nous appliquons ensuite ces coefficients aux données des forces de l'ordre sur l'ensemble du territoire métropolitain. Nous obtenons alors des estimations nationales.

**Résultats** Le nombre de blessés, toutes gravités, estimé est de l'ordre de 300 000/an sur les dernières années, alors qu'il est autour de 72 000 selon les BAAC. Les estimations de blessés MAIS3+ sont de l'ordre de 24 000/an sur ces mêmes années. On note une certaine stabilité, après une baisse nette pour les années précédentes. Sur ces années 2013-2016, les blessés MAIS3+ sont au nombre de 10 000/an pour les deux-roues motorisés et 6 500/an pour les véhicules légers. Cela confirme le plus grand nombre de blessés graves chez les usagers de deux-roues motorisés comparés aux automobilistes, ce qui était apparu dès 2003, avec la même méthodologie (Amoros, 2007). Rappelons que les effectifs de tués sont 800/an pour les usagers de deux-roues motorisés et 1 800/an pour les automobilistes. Les effectifs de cyclistes blessés MAIS3+ sont du même ordre de grandeur que les

piétons blessés MAIS3+, autour de 3 500/an entre 2012 et 2016. Pour rappel, les piétons tués sont au nombre de 500/an et les cyclistes 150/an (2013-2015).

Il est à noter que les effectifs de cyclistes blessés se basent sur des effectifs très bas dans les BAAC, entraînant des coefficients de correction très élevés, et ainsi très sensibles au moindre écart d'enregistrement ; ils sont donc à interpréter avec prudence.

Le tableau ci-après donne depuis 2006 l'effectif estimé de blessés toutes gravités, corrigé du sous-enregistrement et des biais associés. Il présente aussi, pour comparaison, les effectifs des blessés toutes gravités selon les BAAC.

*Tableau 1 : effectifs annuels des blessés toutes gravités, selon capture-recapture ou selon les BAAC, France*

Année	Blessés MAIS1+ Estimations CRC	Blessés toutes gravités Effectifs BAAC*
2006	360 522	102 204
2007	382 189	103 268
2008	364 415	93 853
2009	361 479	91 002
2010	338 282	84 698
2011	327 114	81 399
2012	307 910	75 932
2013	274 263	70 607
2014	291 693	73 048
2015	284 254	70 818
2016	298 293	72 645

CRC = capture-recapture sur BAAC et Registre

\* données BAAC : de possibles indemnes ont été exclus parmi les blessés classés non graves

Comme attendu, les estimations sont bien plus élevées que les chiffres des forces de l'ordre, d'un facteur multiplicatif de 3,6-4 environ.

L'ordre de grandeur des estimations des blessés MAIS1+ par capture recapture est validé par des données provenant de sources telles que l'Enquête Nationale Transports et Déplacements (ENTD), les chiffres des Services d'Incendie et de Secours, ainsi que le ratio du nombre de blessés toutes gravités par le nombre de tués, dans les différents pays européens (chiffres dans la partie Discussion).

Les tableaux ci-après donnent l'effectif de blessés graves (MAIS3+), estimé par capture-recapture et, pour rappel, l'effectif de blessés hospitalisés d'après les BAAC. Ces deux critères sont différents. Un blessé hospitalisé n'a qu'une probabilité de 50% d'être MAIS3+, et de surcroît un blessé MAIS3+ n'est pas toujours hospitalisé (5% d'exception) mais l'hospitalisation (oui/non) est le seul critère de gravité qui est présent dans les BAAC, permettant ainsi une comparaison.



**Tableau 2 : effectifs annuels des blessés MAIS3+ selon capture-recapture, et des blessés hospitalisés selon les BAAC, France**

Année	Blessés MAIS3+ estimations par CRC	Blessés hospitalisés BAAC*
2006	32 037	40 664
2007	32 036	38 619
2008	30 231	34 965
2009	29 509	33 325
2010	27 228	30 418
2011	26 754	29 677
2012	24 542	27 132
2013	23 291	25 966
2014	24 592	26 635
2015	24 273	26 595
2016	25 401	27 187

CRC= capture-recapture sur BAAC et Registre

\* données BAAC : de possibles indemnes ont été exclus parmi les blessés classés non graves

En termes d'ordre de grandeur, les deux effectifs sont proches, alors qu'ils ne devraient pas l'être. En effet, étant donné que, dans le Registre du Rhône, on observe que, parmi les hospitalisés, 45% sont MAIS3+, cela donnerait en partant de l'estimation des MAIS3+, un effectif de 54 000 hospitalisés en 2015 (et non 27 000 comme l'indiquent les BAAC). Cela corrobore le sous-enregistrement des BAAC.

Les tableaux ci-après donnent la répartition des blessés (MAIS1+, MAIS3+) selon le type d'utilisateur.

**Tableau 3 : effectifs annuels des blessés toutes gravités, selon capture-recapture selon le type d'utilisateur, France**

Année	2RM	Vélo	VL	Piéton	Autre*
2006	115 518	61 536	144 983	27 853	10 632
2007	124 623	73 673	145 662	28 288	9 942
2008	120 460	77 327	129 829	27 212	9 588
2009	119 377	75 510	130 320	26 788	9 483
2010	107 198	72 878	122 839	26 125	9 242
2011	103 203	75 576	115 229	25 406	7 700
2012	95 564	68 301	110 686	24 563	8 795
2013	84 711	54 735	103 214	23 740	7 863
2014	87 511	61 840	110 581	24 393	7 368
2015	81 423	61 710	109 956	23 236	7 929
2016	80 895	62 793	123 210	23 220	8 175

\*Pour rappel ; Autre : VU (véhicule utilitaire), TC/Train (TC : Transport en commun), PL (Poids lourd)

Les ordres de grandeur sont différents des effectifs des BAAC.

Selon capture-recapture, les blessés toutes gravités à deux-roues motorisés sont quatre fois plus nombreux que les chiffres des BAAC, 15 fois plus nombreux pour les blessés cyclistes. Par rapport aux BAAC, capture-recapture donne 3,4 fois plus de blessés pour les automobilistes, 2,2 pour les piétons et 2,3 pour les usagers autres (sur 2013-2015).

Les estimations par type d'utilisateur donnent un classement différent selon les sources ou méthodes. Selon les BAAC, les blessés toutes gravités concernent d'abord les automobilistes, puis les usagers de

deux-roues motorisés, puis les piétons, et enfin les cyclistes, alors que l'estimation par capture-recapture indique toujours que le nombre de blessés automobilistes est plus élevé que celui des deux-roues motorisés, mais que le nombre de blessés cyclistes toutes gravités est bien au-dessus de celui des piétons (presque trois fois plus élevé).

Les évolutions sont les mêmes que celles des BAAC : cela vient de la façon dont nous avons construit le modèle, lui-même basé sur la comparaison des évolutions des effectifs des BAAC et du Registre dans le département du Rhône.

*Tableau 4 : effectifs annuels des blessés MAIS3+ estimés selon capture-recapture, selon le type d'usager, France*

Année	2RM	Vélo	VL	Piéton	Autre
2006	13 371	3 503	9 615	4 267	1 281
2007	14 165	3 647	8 972	4 205	1 046
2008	13 606	3 727	7 855	4 012	1 030
2009	13 295	3 679	7 686	3 876	973
2010	11 778	3 613	7 166	3 736	935
2011	11 602	3 636	7 004	3 722	789
2012	10 515	3 162	6 487	3 483	894
2013	9 739	3 039	6 160	3 462	890
2014	9 966	3 802	6 459	3 556	810
2015	9 845	3 534	6 533	3 551	810
2016	9 738	4 216	7 127	3 520	800

Le critère MAIS3+ corrigé du sous-enregistrement ne donne pas le même classement en termes d'effectifs, entre les types d'usagers, que le critère hospitalisation des BAAC. En effet, les chiffres d'hospitalisés des BAAC donnent plus de blessés hospitalisés automobilistes que de blessés hospitalisés usagers de deux-roues motorisés alors qu'à l'inverse, selon le critère MAIS3+ corrigé, les usagers de deux-roues motorisés sont plus nombreux à être blessé MAIS3+ que les automobilistes. Rappelons que le nombre de tués est de 800/an pour les usagers de deux-roues motorisés et de 1 800/an pour les automobilistes. En termes de kilomètres parcourus, les usagers de deux-roues motorisés représentent 2% du trafic alors que les automobilistes en représentent 70%.

Pour les piétons et les cyclistes, les BAAC enregistrent de 4 000 à 5 000 blessés hospitalisés chez les piétons et 1 500 à 1 700 chez les cyclistes. Le critère MAIS3+ et surtout la correction du sous-enregistrement et de ses biais indique qu'il y a quasiment autant de blessés graves piétons et de blessés graves cyclistes (environ 3 500 à 4 000).

Rappelons que les piétons tués sont au nombre de 500/an et les cyclistes 150/an (sur 2013-2015). En termes de nombre de trajets, les piétons sont à 23% de part modale et les cyclistes à 2-3%.

Le graphique ci-après illustre les ordres de grandeur.

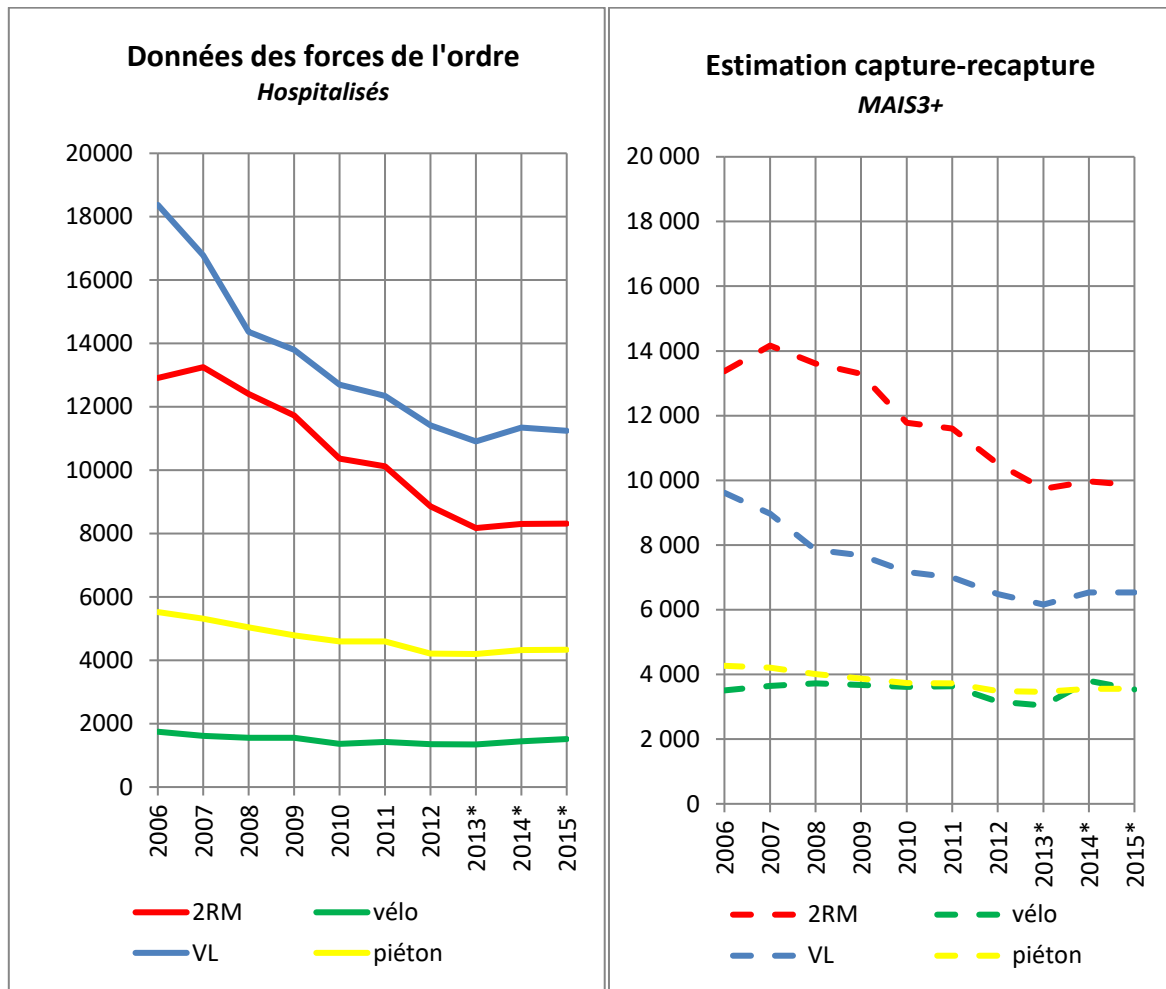


Figure 1: Nombre de blessés graves selon le type d'usager, 2006-2015

Le tableau suivant donne la proportion de blessés MAIS3+ parmi les blessés MAIS1+, selon le mode de transport et selon qu'un tiers est impliqué ou non dans l'accident.

*Tableau 5 : Effectifs annuels moyens estimés par capture-recapture des blessés MAIS3+, des blessés MAIS1+ et ratio MAIS3+/MAIS1+ selon le type d'utilisateur et l'absence/présence de tiers, 2006-2012, France*

Type d'utilisateur	Tiers	MAIS3+	MAIS1+	Ratio (%) MAIS3+/MAIS1+ (CRC France)
Occupants de voiture	Oui	3 888	85 344	4,6
	Non	3 938	43 163	9,1
		7 826	128 507	6,1
Usagers de 2RM	Oui	6 357	45 059	14,1
	Non	6 262	67 218	9,3
		12 619	112 278	11,2
Cyclistes	Oui	1 133	10 716	10,6
	Non	2 434	61 398	4,0*
		3 567	72 115	4,9
Piéton	Oui	3 900	26 605	14,7
Autre	Oui	385	4 700	8,2
	Non	608	4 640	13,1
		993	9 340	10,6

CRC = capture-recapture sur BAAC et Registre

\*Sous enregistrement des blessés graves sans doute insuffisamment corrigé chez les cyclistes.

Pour les accidents sans tiers, les blessés à deux-roues motorisés et les blessés automobilistes connaissent la même proportion de blessés graves (autour de 9%). Elle est plus élevée chez les blessés autres (VU, PL, bus, ...)

Pour les accidents avec tiers (qui est dans la grande majorité des cas un usager motorisé), la proportion de blessés graves est la plus faible chez les automobilistes (4,6%) puis chez les usagers « autres », puis chez les cyclistes (10,6%), puis chez les motocyclistes et piétons (respectivement 14,1% et 14,7%).

Les résultats suivants montrent que la gravité des blessés augmente avec l'âge, et que les hommes sont plus souvent blessés gravement que les femmes. Le sex-ratio est de 2,1 pour les blessés MAIS1+ et 3,3 pour les blessés MAIS3+, ce qui semble indiquer une prise de risques plus grande chez les hommes.

*Tableau 6: Effectifs annuels moyens estimés par capture-recapture des blessés MAIS3+, des blessés MAIS1+ et ratio MAIS3+/MAIS1+ selon l'âge et selon le sexe, 2006-2012, France*

	MAIS3+	MAIS1+	Ratio (%) MAIS3+/MAIS1+ (CRC France)
<b>Age (en années)</b>			
00-09	700	11 071	6,3
10-14	1 165	16 172	7,2
15-19	4 470	53 206	8,4
20-29	6 825	93 006	7,3
30-54	9 804	126 869	7,7
55-64	2 649	25 643	10,3
65-74	1 358	12 083	11,2
75+	1 934	10 794	17,9
<b>Sexe</b>			
Féminin	7 255	123 019	5,9
Masculin	21 650	225 825	9,6

CRC = capture-recapture sur BAAC et Registre

**Discussion** L'estimation nationale du nombre de blessés toutes gravités obtenue peut être comparée à deux autres estimations : la première obtenue à partir de l'Enquête Nationale Transports et Déplacements, réalisée tous les 10 ans sur 20 000 personnes, est de 470 000 blessés toutes gravités en 2006, versus 360 000 pour notre estimation et de 416 000 en 2007, versus 380 000. La deuxième, effectuée par les services d'incendie et de secours est de 270 000 en moyenne sur 2012-2015, versus 280 000 pour notre estimation.

Par ailleurs, le ratio nombre de blessés toutes gravités/nombre de tués que nous estimons autour de 60-70 blessés toutes gravités pour un tué est similaire aux ratios les plus élevés des autres pays européens tels que le Royaume Uni ou l'Allemagne.

Enfin, le ratio nombre de blessés graves/nombre de tués est estimé à 7 blessés graves pour un tué. Les autres pays européens (Allemagne, Espagne, Italie, Suède, Belgique) sont plutôt autour de 4, le Royaume Uni à 2,5 et les Pays-Bas à 12. Ces différences s'expliquent en partie par les méthodes utilisées qui, dans les premiers pays cités, se basent uniquement sur les données hospitalières, alors que les Pays-Bas utilisent une méthode proche de capture-recapture. Ces différences peuvent aussi s'expliquer par les différences de répartition des modes de transport entre les pays alors qu'ils ont des ratios très contrastés : 24 blessés graves pour un tué chez les cyclistes d'après nos estimations 2014, mais 4 chez les automobilistes. Le ratio (global) très élevé aux Pays-Bas provient sans doute du fait que les cyclistes y sont très nombreux et que les ratios MAIS3+/tués par type d'utilisateur sont les mêmes que les nôtres.

## 1 INTRODUCTION

En France, comme dans la plupart des pays développés, le problème de santé publique posé par les accidents de la circulation routière est surtout évalué en termes de mortalité, et la communication institutionnelle sur le sujet porte essentiellement sur le nombre de tués et son évolution selon une certaine périodicité. La mortalité routière est évidemment à considérer en priorité, mais la morbidité doit également être estimée pour prendre en compte l'ensemble des conséquences sanitaires des accidents de la route. Or si le décompte du nombre de tués est effectué de manière satisfaisante par les forces de l'ordre, le décompte des blessés et du niveau de gravité de leur traumatologie est toujours plus délicat. La première raison est que le nombre de blessés, qui sont recensés par les forces de l'ordre (en France et ailleurs) est sous-estimé (Amoros et al., 2006). Cette sous-estimation provient d'une sous-déclaration par les blessés eux-mêmes et d'un sous-enregistrement par les forces de l'ordre. Ce sous-enregistrement est en outre variable selon les caractéristiques de l'accident. Deuxièmement, le niveau de gravité des atteintes du blessé est délicat à déterminer, grandement dépendant de la source des données et de l'échelle de gravité utilisée. Ainsi la notion de blessé grave nécessite une définition claire pour être véritablement opérationnelle.

La notion de blessé hospitalisé, adoptée dans les données des forces de l'ordre en France (BAAC), est assez claire et adoptée dans la plupart des pays européens. Cependant, ce critère n'est pas satisfaisant pour mesurer la gravité. Depuis 2014, un consensus à la Commission Européenne a abouti à considérer un blessé grave à partir de la codification internationale des traumatismes Abbreviated Injury Scale (AIS). La classification AIS permet de coder chacune des lésions du blessé, de 1 à 6, 1 pour blessure légère et 6, au-delà de tout traitement possible. La gravité du blessé est définie comme sérieuse lorsque l'AIS codant la blessure la plus grave est supérieure ou égale à 3 (MAIS3+, Maximum AIS  $\geq$  3).

La plupart des pays européens qui fournissent le nombre de victimes MAIS3+ à la Commission utilisent des données de santé (en général hospitalières) qui permettent d'accéder au niveau de traumatisme à travers l'utilisation de la Classification Internationale des Maladies (CIM version 9 ou 10). Celle-ci est utilisée pour le codage de l'ensemble des pathologies, et donne la possibilité pour les traumatismes de renseigner leur cause "externe" (c'est-à-dire si c'est accidentel ou intentionnel, et quel type d'accident (route, vie courante, ...)). Le passage de la classification CIM à l'échelle AIS se fait ensuite grâce à divers algorithmes qui permettent d'évaluer un niveau MAIS3+ (en oui/non), mais pas toujours pour tous les codes.

Pour la France, la même méthode est envisagée à partir des données hospitalières (estimations en cours). Mais une autre méthode est possible en raison de l'existence depuis de nombreuses années (1996) du Registre du Rhône, dans lequel on dispose du codage des traumatismes dans l'échelle AIS pour chaque victime considérée comme blessée dès qu'elle présente au moins une lésion de niveau 1. Le nombre de blessés MAIS3+ est donc mesurable directement à partir de ce Registre. De plus, ce Registre qui se base sur l'ensemble des services hospitaliers publics et privés du Rhône (et limitrophes) susceptibles d'accueillir des victimes de la route est bien plus complet : alors que les forces de l'ordre du Rhône comptent 2 800 blessés/an, le Registre en décompte 7 400/an. Un chaînage entre les deux sources permet d'identifier 1 700 blessés communs et un plus grand nombre encore (6 800), qui ne sont connus que dans une source. Cette défaillance indique que des blessés peuvent échapper aux deux recensements. Une méthode, capture-recapture, permet à partir de l'existence de deux sources (ou plus), et sous certaines conditions, d'estimer le nombre de cas échappant aux deux sources et ainsi le nombre total de blessés dans le département du Rhône. De là, en comparant ce bilan complet et les données des forces de l'ordre (BAAC) du Rhône, on estime des coefficients correcteurs du sous-enregistrement des BAAC et des biais associés. Ces coefficients correcteurs sont ensuite appliqués, aux données nationales des forces de l'ordre, et nous obtenons ainsi le bilan national de la morbidité

routière. L'hypothèse sous-jacente est celle d'une certaine homogénéité pour une force de l'ordre donnée dans ses pratiques d'enregistrement des blessés de la route, sur le territoire français.

Les modèles sont construits sur la période 2006-2012 et sont appliqués à la période 2006-2016.

## **2 MATERIEL ET METHODES**

### **2.1 Présentation des bases de données**

#### **2.1.1 Le Registre du Rhône**

Créé en 1995, Le Registre du Rhône enregistre les victimes d'accident corporel de la circulation routière survenant dans le département du Rhône. Pour être répertorié dans le Registre, l'accident doit provoquer au moins une victime, survenir sur une voie publique ou privée, impliquer un véhicule, motorisé ou non, en mouvement. Par victime est entendue toute personne présentant au moins une lésion au sens de l'échelle AIS, nécessitant ou non un soin médical. Les informations sont fournies par les différentes structures sanitaires prenant en charge les blessés de la route, c'est-à-dire les hôpitaux privés et publics du département et limitrophes. Ainsi, 247 services de soins participent au recueil des données, allant de la prise en charge pré-hospitalière aux services de réadaptation fonctionnelle. Les données concernent l'identification de la victime (nom, prénom, sexe, date de naissance), les informations administratives (adresse, notion d'accident de travail), les caractéristiques accidentologiques (date, heure, lieu, type de véhicule, type d'usager impliqué, antagoniste, la position sur ou dans le véhicule, les équipements de sécurité : ceinture, casque...), le bilan lésionnel complet (dont l'hospitalisation ou non), et le devenir du patient. Le bilan lésionnel complet est basé sur l'ensemble des diagnostics établis par les différentes unités de soins ayant pris en charge la victime. Dans le Registre, le bilan lésionnel est codé en AIS, mais aussi en IIS (Injury Impairment Scale), permettant de mesurer les séquelles.

Les tués sur le coup ou ultérieurement sont recensés dans le Registre, notamment grâce à l'institut médico-légal de Lyon. Ils sont exclus de l'estimation du nombre de blessés graves.

#### **2.1.2 Les Bulletins d'Analyse des Accidents Corporels (BAAC)**

Les forces de l'ordre françaises doivent rédiger un procès-verbal pour chaque accident corporel de la circulation routière. Par définition, un accident corporel (mortel ou non) de la circulation routière doit provoquer la mort ou la blessure d'au moins une personne, survenir sur une voie publique, impliquer un véhicule en mouvement, en excluant les suicides et homicides volontaires. L'ensemble des caractéristiques accidentologiques sont également répertoriées. Les forces de l'ordre classifient chaque personne impliquée dans l'accident comme indemne, blessé léger, blessé hospitalisé, ou tué, si la mort est survenue dans les trente jours après l'accident.

Les tués et les indemnes sont exclus de l'approche d'estimation du nombre de blessés graves.

### **2.2 Méthodologie de l'extrapolation des BAAC en 2 étapes**

La coexistence des BAAC et du Registre du Rhône permet, par une approche capture-recapture et sous certaines conditions, d'estimer le nombre total de blessés dans le Rhône et ainsi les coefficients correcteurs du sous-enregistrement des BAAC et de leur biais.

Cela constitue la 2<sup>ème</sup> étape dans la méthode mise en place pour estimer le nombre national de blessés graves. La première étape concerne la prédiction, parmi les blessés BAAC (non encore corrigés) de la proportion des blessés MAIS3+. Pour atteindre cet objectif nous nous basons sur les blessés

communs au Registre et aux BAAC du Rhône, où nous disposons ainsi à la fois de la gravité MAIS des blessés par le Registre, et des caractéristiques de l'accident par les BAAC. Nous construisons (étape 1A) un modèle de prédiction du MAIS sur ces blessés, puis (étape 1B) nous l'appliquons aux BAAC rhodaniens et nationaux. L'hypothèse sous-jacente est que, pour une force de l'ordre donnée, les pratiques d'enregistrement des blessés sont relativement homogènes sur le territoire français.

La figure ci-après indique les différentes étapes et les couleurs indiquent s'il s'agit des données rhodaniennes (en rose) ou des données nationales (en bleu).

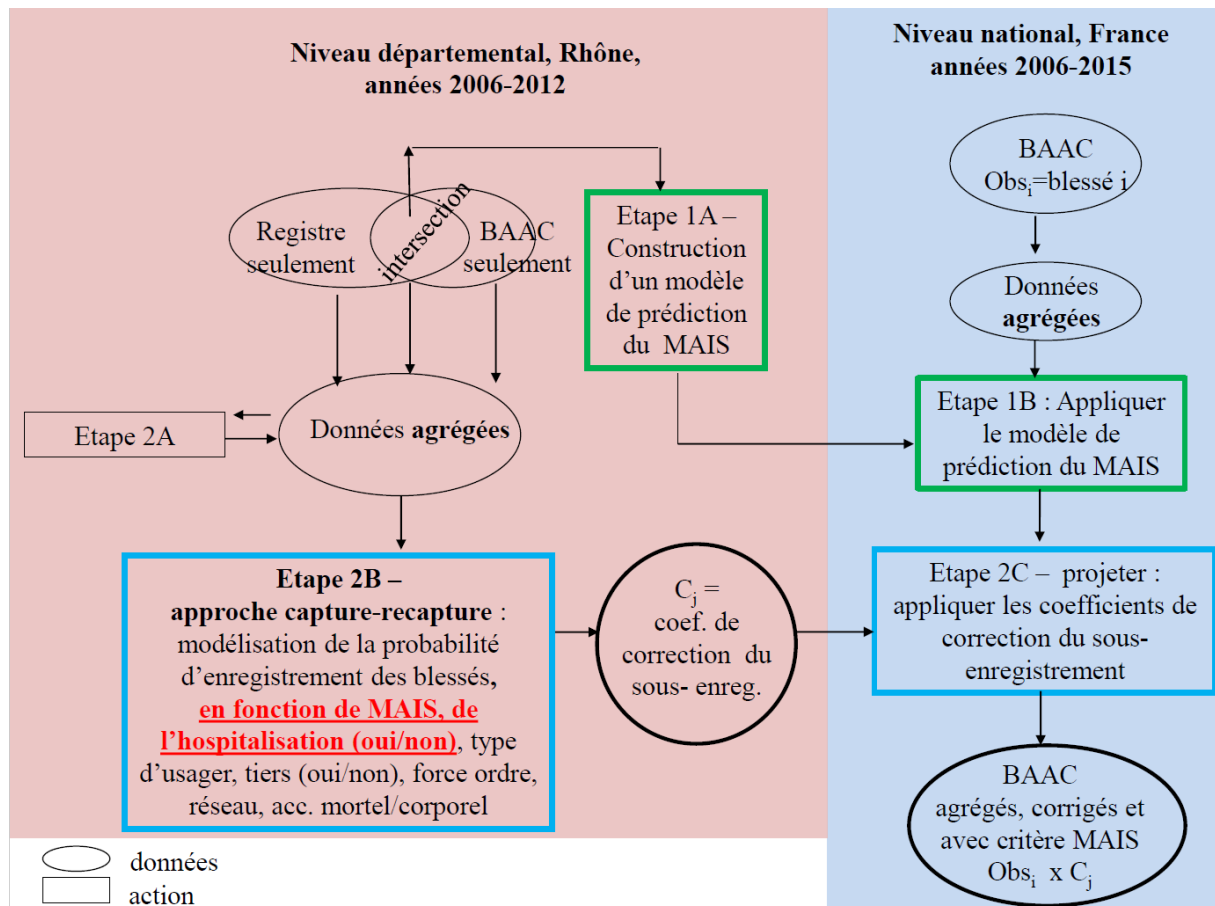


Figure 2 : description des étapes de l'extrapolation

L'étape 2A est un préambule à l'étape capture-recapture : elle consiste à améliorer le résultat du chaînage entre BAAC et Registre, rhodaniens, pour s'approcher de la condition n°2 de capture-recapture, c'est-à-dire l'identification parfaite des individus communs aux deux sources. Nous estimons le nombre de faux positifs et faux négatifs (définis en regard du chaînage) et nous corrigeons le résultat du chaînage à l'aide de ces effectifs.

L'étape 2B est l'étape de capture-recapture elle-même, au niveau rhodanien : estimer le nombre total de blessés de la route, directement par un modèle multivarié, et en déduire les coefficients multiplicateurs de correction (du sous-enregistrement et de ses biais) entre les BAAC, et le sur-ensemble exhaustif obtenu.

L'étape 2C est l'étape de projection à l'échelle nationale : appliquer ces coefficients de correction (estimés au niveau rhodanien) aux BAAC nationaux, sous certaines conditions (cf. 2.10), et permettre ainsi une estimation du bilan national des blessés de la route. Il est à noter que cette projection n'est



pas une projection de l'accidentalité du Rhône sur la France, mais une projection des pratiques d'enregistrement des blessés par les forces de l'ordre du Rhône à la France. L'hypothèse faite est celle d'une certaine homogénéité, pour un type de force de l'ordre donné, des pratiques d'enregistrement des blessés de la route, sur le territoire français.

### 2.3 Étape 1 : Prédiction du nombre de blessés MAIS 3+ sur les blessés BAAC

Il ne s'agit pas de prédire le niveau de MAIS à l'échelle individuelle de chaque blessé, mais à un niveau agrégé, en termes d'effectifs selon les groupes définis par diverses caractéristiques du blessé et/ou de l'accident et qui sont renseignées dans les BAAC. Le niveau d'agrégation est défini par les variables utilisées ultérieurement dans le modèle de sous-enregistrement (étape 2) et par toute autre variable d'intérêt telle que l'âge ou le sexe.

### 2.4 Étape 1A : construction du modèle de prédiction du MAIS

Nous construisons le modèle de prédiction du MAIS3+ sur les blessés identifiés comme communs aux BAAC et au Registre du Rhône, où nous disposons ainsi du MAIS3+ par le Registre, et des variables BAAC dont certaines vont être choisies comme variables prédictives, puisque le but est ensuite d'appliquer le modèle de prédiction ainsi construit, aux données BAAC France entière.

La principale variable de prédiction du MAIS est le critère BAAC de gravité du blessé, à savoir depuis 2005 le critère hospitalisé/non-hospitalisé.

#### 2.4.1 Données

Nous travaillons sur les blessés identifiés comme communs aux BAAC et au Registre du Rhône. Sur la période 2006-2012, on dénombre 12 421 blessés communs, soit une moyenne de 1 774 blessés par an.

Les observations pour lesquelles le MAIS est manquant sont exclues de la construction du modèle de prédiction (étant donné qu'il s'agit de la variable à prédire). Elles représentent 1,15 % des observations. Restent ainsi 12 308 observations.

Les variables à inclure dans le modèle peuvent présenter des valeurs manquantes. Elles sont soit imputées (imputation simple, et d'après une distribution multinomiale), soit regroupées avec la modalité « autre » quand elle existe.

Afin de construire le modèle de prédiction du MAIS3+, il est utile d'explorer la proportion de blessés MAIS3+ parmi les blessés identifiés comme communs aux BAAC et au Registre, selon différentes variables concernant l'accident ou le blessé.

La variable « hospitalisé » (oui/non) disponible dans les BAAC est très corrélée à la gravité MAIS3+. En effet, parmi les blessés classés « hospitalisés » par les forces de l'ordre, 39,6% sont MAIS3+, alors que parmi les blessés classés « non-hospitalisés » par les forces de l'ordre, 3,1% sont MAIS3+.

On note cependant des variations selon les types de force de l'ordre (cf. tableau ci-après).

*Tableau 7 : Proportion de blessés MAIS3+ selon le type de force de l'ordre parmi les blessés identifiés comme communs aux BAAC et au Registre, 2006-2012*

Type de force de l'ordre	Hospitalisation selon BAAC	Proportion de blessés MAIS3+ (%)
CRS	Blessé hospitalisé	30,2
	Blessé non hospitalisé	0,8
Gendarmerie	Blessé hospitalisé	46,5
	Blessé non hospitalisé	3,7
Police	Blessé hospitalisé	40,0
	Blessé non hospitalisé	3,6

Les variations observées entre forces de l'ordre peuvent en partie (mais en partie seulement) s'expliquer par le fait que les trois types de forces de l'ordre couvrent des réseaux différents où les gravités diffèrent, couplés à des zones plus ou moins urbaines ou rurales, où les gravités diffèrent aussi.

Le tableau ci-après donne la proportion de blessés MAIS3+, en fonction du type d'utilisateur, de la présence/absence de tiers et du classement hospitalisé (oui/non) selon les forces de l'ordre.

Les proportions de MAIS3+ selon ces variables peuvent sembler curieuses. Par exemple, pour les cyclistes, la proportion de MAIS3+ parmi ceux classés hospitalisés (par les forces de l'ordre) est plus élevée pour les cyclistes sans tiers (56 %) que pour les cyclistes avec tiers (43,9 %) alors que l'on sait (d'après le Registre) que les cyclistes blessés avec tiers sont plus gravement atteints que les cyclistes blessés sans tiers ; cela se voit dans le tableau suivant. On s'attend d'ailleurs à cela, puisque l'énergie du choc dépend de la masse du/des impliqués et de la vitesse du/des impliqués et on sait que l'antagoniste des cyclistes est généralement un véhicule motorisé.

Cette apparente discordance s'explique par le fait que, dans le tableau ci-dessous, il s'agit des cyclistes enregistrés par les forces de l'ordre : pour que les cyclistes blessés sans tiers soient enregistrés par les forces de l'ordre, il faut qu'ils soient assez graves (cf. partie 2.6 « situation de l'enregistrement dans le Rhône).

Il en est de même pour les usagers de deux-roues motorisés.

En revanche (tableau suivant), pour les automobilistes et les usagers « autres », la proportion plus élevée de blessés MAIS3+ chez les hospitalisés, dans un accident sans tiers par rapport à un accident avec tiers, correspond à la gravité réelle observée.

Les proportions de MAIS3+ chez les blessés classés non-hospitalisés par les forces de l'ordre sont très variables. Cela est sans doute lié aux petits effectifs de MAIS3+ dans ces catégories.

**Tableau 8 : Proportion de blessés MAIS3+ selon le type d'usager, la présence ou l'absence de tiers et l'hospitalisation du blessé parmi les blessés identifiés comme communs aux BAAC et au Registre, 2006-2012**

Type usager (BAAC)	Tiers (BAAC)	Hospitalisation : classement selon BAAC	Proportion de blessés MAIS3+ (%)
2RM	Avec tiers	Blessé hospitalisé	49,0
		Blessé non hospitalisé	4,9
	Sans tiers	Blessé hospitalisé	56,6
		Blessé non hospitalisé	9,4
Vélo	Avec tiers	Blessé hospitalisé	43,9
		Blessé non hospitalisé	2,7
	Sans tiers	Blessé hospitalisé	56,3
		Blessé non hospitalisé	0,1
VL	Avec tiers	Blessé hospitalisé	27,3
		Blessé non hospitalisé	1,0
	Sans tiers	Blessé hospitalisé	36,5
		Blessé non hospitalisé	2,4
Piéton	Avec tiers	Blessé hospitalisé	47,1
		Blessé non hospitalisé	7,2
Autre	Avec tiers	Blessé hospitalisé	30,8
		Blessé non hospitalisé	0,6
	Sans tiers	Blessé hospitalisé	45,8
		Blessé non hospitalisé	0,9

Le tableau ci-après indique la proportion de MAIS3+ selon le sexe. La différence entre hommes et femmes est conforme à la réalité.

**Tableau 9 : Proportion de blessés MAIS3+ selon le sexe et l'hospitalisation du blessé parmi les blessés identifiés comme communs aux BAAC et au Registre, 2006-2012**

Sexe du blessé	Hospitalisation selon BAAC	Proportion de blessés MAIS3+ (%)
Féminin	Blessé hospitalisé	33,8
	Blessé non hospitalisé	2,2
Masculin	Blessé hospitalisé	45,0
	Blessé non hospitalisé	3,7

#### 2.4.2 Modélisation

La variable à prédire est le MAIS en trois catégories : MAIS 1, MAIS 2 et MAIS 3+ ; c'est une variable catégorielle ordinaire. Le modèle multinomial de logit cumulé est celui qui est a priori le plus adapté. Il faut cependant, pour pouvoir l'utiliser, valider ses conditions d'application. Il s'agit de vérifier que les variables prédictives du modèle agissent de la même manière sur la variable MAIS quel que soit son niveau (1, 2 ou 3+), autrement dit que les coefficients des variables prédictives du modèle soient les mêmes quel que le soit le niveau du MAIS étudié. On effectue pour cela un test des odds proportionnels. La p-value du test étant inférieure à 0.0001, l'hypothèse des odds proportionnels ne peut pas être retenue. On effectue donc deux modèles logistiques séparés : MAIS1,2 vs MAIS3+ et MAIS1 vs MAIS2+. Les deux modèles s'écrivent :

$$\begin{aligned} \text{Logit } P_g &= \ln P(\text{MAIS} \geq g) / P(\text{MAIS} < g), \text{ pour } g=2 \text{ et } 3 \text{ (valeurs seuils)} \\ &= \beta_{g0} + \sum_k \beta_{gk} X_{gk} \text{ où } X_k \text{ sont les variables prédictives} \end{aligned}$$

### 2.4.3 Variables du modèle

Le modèle de prédiction du MAIS retenu contient les variables suivantes (des BAAC) :

- type de force de l'ordre (CRS/gendarmerie/Police) renseignant l'accident
- lieu de l'accident (En agglomération/Hors agglomération)
- interaction entre le lieu de l'accident et le type de force de l'ordre
- gravité de l'accident (mortel/corporel) dans lequel le blessé est impliqué
- type d'utilisateur (automobiliste/usager de 2RM/cycliste/piéton/autre)
- présence/absence d'un tiers dans l'accident (oui/non)
- type d'antagoniste (2RM/Piéton-Cycliste/Véhicule Lourd/Voiture-Autre/Véhicule utilitaire)
- hospitalisation selon les BAAC (blessé hospitalisé/blessé non-hospitalisé)
- interaction entre le type de force de l'ordre et l'hospitalisation selon les BAAC
- interaction entre le type d'utilisateur, la présence de tiers et l'hospitalisation selon les BAAC
- âge en catégories (0-9 ans/10-14/15-19/20-29/30-54/55-64/65-74/75 ans et +)
- sexe (Masculin/Féminin).

Le choix de ces variables ne s'est pas fait suivant un critère statistique, mais suivant notre connaissance de leur corrélation à la gravité des blessés et/ou à la gravité de l'accident. La gravité de l'accident est notamment liée à l'énergie cinétique dégagée pendant l'accident, dépendante de la masse des véhicules et de leur vitesse lors de l'accident. La masse et la vitesse n'étant pas disponibles, nous avons choisi des variables approchantes. La masse est liée au type d'utilisateur (voiture, 2RM, vélo, ...) et au type d'antagoniste (vélo, voiture, ...). La vitesse est liée au lieu (en/hors agglomération) et indirectement au type de force de l'ordre (par le biais du réseau). Ces deux variables sont liées, et nous traduisons cela par une interaction.

Pour les variables choisies, la proportion de valeurs manquantes est faible (moins de 5%). Elles sont soit regroupées avec la modalité « autre » quand elle existe, soit remplacées par des valeurs imputées (c'est le cas des variables âge et sexe).

### 2.4.4 Modèle obtenu

Les paramètres des modèles logistiques sont donnés ci-après

**Tableau 10 : modèle logistique MAIS3+ vs MAIS1-2 multivarié,  
construit sur 12 308 blessés communs au Registre du Rhône et aux BAAC, 2006-2012**

Variable	beta	OR	IC à 95%	
Ordonnée à l'origine	-4,65			
Lieu accident *type de force de l'ordre				
(Hors agglo) CRS	-1,21	0,30	0,15	0,57
En agglo, Gendarmerie ( <i>réf</i> )	0	1	-	-
En agglo, Police	-0,17	0,84	0,55	1,28
Hors agglo, Gendarmerie	0,41	1,51	1,24	1,84
Hors agglo, Police	-0,17	0,84	0,43	1,65
Gravité de l'accident				
Corporel ( <i>réf</i> )	0,00	1,00		
Mortel	0,55	1,73	1,16	2,60
Force de l'ordre* gravité BAAC				
CRS, BH	4,72	112	31,4	401
CRS, BNH	-5,93	0,003	-	-
Gendarmerie, BH	3,81	45	13,6	149
Gendarmerie, BNH ( <i>réf</i> )	0	1	-	-
Police, BH	3,77	43	135	140
Police, BNH	-4,11	0,02	-	-
Type d'usager * tiers (avec /sans) *gravité BAAC				
2RM avec tiers, BH	0,53	1,70	0,56	5,18
2RM avec tiers, BNH	1,54	4,68	3,07	7,14
2RM sans tiers, BH	0,58	1,78	1,03	3,08
2RM sans tiers, BNH	2,25	9,44	2,93	30,40
Vélo BH	0,27	1,32	0,45	3,87
Vélo BNH	0,85	2,35	1,17	4,72
VL avec tiers, BH	-0,56	0,57	0,18	1,74
VL avec tiers, BNH ( <i>réf</i> )	0	1	-	-
VL sans tiers, BH	-0,25	0,78	0,46	1,32
VL sans tiers, BNH	0,96	2,62	0,83	8,22
Piéton avec tiers, BH	0,50	1,65	0,54	5,06
Piéton avec tiers, BNH	1,94	6,97	4,54	10,7
Autre avec tiers, BH	-0,68	0,50	0,15	1,66
Autre avec tiers, BNH	-1,47	0,23	0,03	1,67
Autre sans tiers, BH	11,97	-	-	-
Autre sans tiers, BNH	0,98	2,66	0,56	12,7
Type d'antagoniste				
2RM	-0,15	0,86	0,62	1,20
Piéton ou Vélo	-0,06	0,94	0,55	1,61
Véhicule utilitaire	0,33	1,39	1,09	1,76
Aucun	0,23	1,26	0,47	3,38
Véhicule Lourd	0,52	1,68	1,35	2,10
Voiture ou Autre ( <i>réf</i> )	0	1	-	-
Age				
0-9 ans	-0,29	0,79	0,56	1,11
10-14	-0,08	0,92	0,67	1,27
15-19	-0,10	0,90	0,75	1,09
20-29	-0,06	0,94	0,81	1,09
30-54 ( <i>réf</i> )	0,00	1,00	-	-
55-64	0,37	1,45	1,15	1,82
65-74	0,27	1,31	0,98	1,77
75+	0,74	2,10	1,61	2,75
Sexe				
Féminin ( <i>réf</i> )	0	1	-	-
Masculin	0,37	1,44	1,26	1,65

BH= Blessé hospitalisé, BNH= Blessés léger (non hospitalisé), 2RM = deux-roues motorisé, réf = catégorie de référence

**Tableau 11 : modèle logistique MAIS2+ vs MAIS1 multivarié,  
construit sur 12 308 blessés communs au Registre du Rhône et aux BAAC, 2006-2012**

Variable	beta	OR	IC à 95%	
Ordonnée à l'origine	-2,02			
Lieu accident *type de force de l'ordre				
(Hors agglo) CRS	-0,30	0,74	0,60	0,95
En agglo, Gendarmerie (réf)	0	1	.-	.-
En agglo, Police	-0,35	0,71	0,57	0,88
Hors agglo, Gendarmerie	0,33	1,40	1,15	1,70
Hors agglo, Police	-0,26	0,77	0,49	1,23
Gravité de l'accident				
Corporel (réf)	0,00	1,00	.	.
Mortel	0,08	1,09	0,76	1,55
Force de l'ordre* gravité BAAC				
CRS, BH	1,41	4,12	1,18	14,35
CRS, BNH	-1,71	0,18	-	-
Gendarmerie, BH	1,42	4,14	1,20	14,35
Gendarmerie, BNH (réf)	0	1	-	-
Police, BH	1,42	4,12	1,19	14,23
Police, BNH	-2,02	0,13	-	-
Type d'utilisateur * tiers (avec /sans) *gravité BAAC				
2RM avec tiers, BH	2,18	8,88	2,57	30,6
2RM avec tiers, BNH	1,32	3,76	3,19	4,44
2RM sans tiers, BH	0,99	2,70	1,39	5,23
2RM sans tiers, BNH	0,19	1,21	0,39	3,77
Vélo, BH	1,74	5,68	1,62	19,8
Vélo, BNH	1,17	3,22	2,51	4,15
VL avec tiers, BH	0,63	1,88	0,55	6,47
VL avec tiers, BNH (réf)	0	1	-	-
VL sans tiers, BH	-0,08	0,93	0,51	1,67
VL sans tiers, BNH	-0,80	0,45	0,15	1,37
Piéton avec tiers, BH	2,03	7,65	2,20	26,6
Piéton avec tiers, BNH	1,45	4,26	3,54	5,12
Autre avec tiers, BH	0,37	1,45	0,40	5,30
Autre avec tiers, BNH	-0,25	0,78	0,54	1,12
Autre sans tiers, BH	-3,63	0,03	.-	.-
Autre sans tiers, BNH	-0,83	0,44	0,13	1,45
Type d'antagoniste				
2RM	-0,35	0,71	0,55	0,91
Piéton ou Vélo	0,04	1,04	0,71	1,53
Véhicule utilitaire	-0,09	0,91	0,75	1,10
Aucun	1,45	4,25	1,43	12,7
Véhicule Lourd	0,35	1,42	1,19	1,69
Voiture ou Autre (réf)	0	1	.-	.-
Age				
0-9 ans	-0,50	0,60	0,47	0,78
10-14	-0,08	0,81	0,63	1,04
15-19	-0,16	0,85	0,73	0,99
20-29	-0,18	0,83	0,75	0,93
30-54 (réf)	0	1	-	-
55-64	0,36	1,44	1,20	1,72
65-74	0,42	1,52	1,21	1,92
75+	0,61	1,85	1,45	2,35
Sexe				
Féminin (réf)	0,00	1,00	.	.
Masculin	0,35	1,41	1,28	1,56

BH= Blessé hospitalisé, BNH= Blessés léger (non hospitalisé), 2RM = deux-roues motorisé, réf= catégorie de référence

Pour les variables du modèle qui jouent seulement en effet simple sur la probabilité du blessé d'être MAIS3+, à savoir les variables gravité de l'accident (mortel/corporel), type d'antagoniste, âge, sexe, les odds-ratios peuvent être lus directement dans le tableau. Exemple l'odds-ratio des 75 ans et + d'être MAIS3+ est de 2,1 par rapport à la catégorie de référence les 30-54 ans, toutes choses égales par ailleurs, c'est-à-dire que c'est l'effet propre de l'âge, les autres effets étant pris en compte par le modèle.

L'odds-ratio est un rapport de cotes, c'est-à-dire le ratio entre deux cotes, au sens utilisé dans les paris hippiques : la cote est la probabilité de gagner, divisée par la probabilité de perdre. La cote ici est la probabilité d'être blessé gravement au sens MAIS, c'est-à-dire être un blessé MAIS3+ contre la probabilité d'être blessé légèrement au sens MAIS, c'est-à-dire être un blessé MAIS1,2. Le ratio entre deux cotes est donc ici le ratio entre la cote d'un groupe de blessés (ex : blessés âgés de plus de 75 ans) et la cote du groupe de blessé de référence (ex : blessés dont l'âge est compris entre 30 et 54 ans). Par exemple, l'odds-ratio correspondant aux blessés de plus de 75 ans est estimé à 2,10. Les blessés de 75 ans et plus sont donc beaucoup plus souvent blessés MAIS3+ que les blessés dont l'âge est compris entre 30 et 54 ans, toutes choses égales par ailleurs, notamment en tenant déjà compte de la catégorie hospitalisé (oui/non) attribuée dans les BAAC.

Pour les variables qui sont en interaction avec d'autres, il faut des calculs supplémentaires pour obtenir l'odds-ratio entre deux profils de blessés. Ceux-ci sont donnés dans le tableau ci-après.

*Tableau 12 : odds-ratios entre différents profils de blessés, associés à la probabilité qu'ils soient MAIS3+*

Profils de blessés :	Comparaison	OR
Accidentés en zone gendarmerie	Hors agglomération vs en agglomération	1,51
Accidentés en zone police	Hors agglomération vs en agglomération	1,001
Véhicule léger, blessé hospitalisé	Avec tiers vs sans tiers	0,73
Véhicule léger, blessé non hospitalisé	Avec tiers vs sans tiers	0,38
2RM, blessé hospitalisé	Avec tiers vs sans tiers	0,96
2RM, blessé non hospitalisé	Avec tiers vs sans tiers	0,50
Vélo, blessé hospitalisé	Avec ou sans tiers regroupés	-
Vélo, blessé non hospitalisé	Avec ou sans tiers regroupés	-
Piéton, blessé hospitalisé	Tous avec tiers (définition)	-
Piéton, blessé non hospitalisé	Tous avec tiers (définition)	-
Autre, blessé hospitalisé	Avec tiers vs sans tiers	0,000003
Autre, blessé non hospitalisé	Avec tiers vs sans tiers	0,086
Véhicule léger avec tiers, zone gendarmerie	BH vs BNH	45
Véhicule léger avec tiers, zone police	BH vs BNH	1510
Véhicule léger avec tiers, zone CRS	BH vs BNH	24000

BH= Blessé hospitalisé, BNH= Blessés léger (non hospitalisé), 2RM = deux-roues motorisé.

L'odds-ratio entre blessés en ou hors agglomération par rapport à la probabilité d'être MAIS3+ dépend du type de force de l'ordre, et est valable à l'intérieur d'un profil de blessés, défini en termes de mode de transport, accident avec ou sans tiers, et classement hospitalisé (oui/non) dans les BAAC. L'odds-ratio vaut 1,51 en zone gendarmerie pour la comparaison hors agglomération versus en agglomération. Cela correspond bien à une gravité moyenne plus forte hors agglomération qu'en agglomération, avec des vitesses moyennes plus élevées. Pour la zone police, l'odds-ratio est de 1, n'indiquant pas de différence de probabilité d'être MAIS3+ selon ce critère, toutes choses égales par ailleurs.

Nous obtenons aussi les odds-ratios associés à la probabilité d'être MAIS3+ entre les blessés dans un accident avec tiers versus sans tiers. Ils dépendent du mode de transport et du classement blessé hospitalisé/non hospitalisé. On note notamment, chez les véhicules légers et les deux-roues motorisés

que l'odds-ratio des blessés dans un accident, avec tiers versus sans tiers est plus loin de 1 (où 1 est l'absence d'effet) chez les blessés classés non hospitalisés que chez ceux classés hospitalisés. Autrement dit, chez les blessés classés non hospitalisés, la présence ou absence de tiers joue plus sur la gravité MAIS3+ que chez les blessés classés hospitalisés.

Les résultats sont curieux pour les modes de transport « autres » (véhicules utilitaires, camions, autocars, ...). Cela provient sans doute de petits effectifs de certains profils de blessés.

Nous avons aussi estimé les odds-ratios associés à la probabilité d'être MAIS3+ des blessés classés hospitalisés versus non hospitalisés selon les BAAC pour les différents modes de transport. Pour la catégorie des véhicules légers dans un accident avec tiers, l'odds-ratio est de 45 en zone gendarmerie, 1510 en zone Police et 24000 en zone CRS. Ces odds-ratios très élevés, surtout en zone police et CRS s'expliquent par deux éléments.

Premièrement, les variables hospitalisé (oui/non) et MAIS3+ (oui/non) sont corrélées. Parmi les blessés hospitalisés, une grande partie sont MAIS3+ (la moitié d'après le Registre du Rhône) et surtout, parmi les blessés non hospitalisés, rares sont ceux qui sont MAIS3+.

Deuxièmement, la variable « hospitalisé » des BAAC, n'est pas fiable à 100 % (elle n'est pas en correspondance parfaite avec la variable hospitalisé (oui/non) du Registre, chez les blessés communs aux deux sources) et sa qualité semble dépendre du type de force de l'ordre.

La gendarmerie enregistre peu les blessés légers (cf. Tableau 15) et l'odds-ratio de 45 (entre blessé hospitalisé et blessé non hospitalisé), bien moindre qu'en zone police ou CRS, indique que, parmi les blessés que la gendarmerie enregistre comme non hospitalisés, un nombre non-négligeable sont en fait MAIS3+. Dit autrement, les blessés légers que la gendarmerie choisit d'enregistrer comme tels ne sont pas si légers que ça. Les odds-ratios très élevés en zone police ou CRS indiquent une bien plus grande dichotomisation entre blessés hospitalisés et blessés non hospitalisés, ce qui est plus conforme à la réalité.

Ce modèle permet surtout d'estimer la probabilité pour un blessé de la route d'être MAIS3+ pour chacun des profils d'accident. Par exemple, le tableau ci-dessous donne la probabilité d'être un blessé MAIS3+ selon l'hospitalisation (oui/non) du blessé, le type d'usager et la présence et l'absence de tiers (et pour des valeurs données des autres variables du modèle).

**Tableau 13 : Probabilité pour un accidenté de la route d'être blessé MAIS3+ selon l'hospitalisation du blessé, le type d'usager et la présence ou l'absence de tiers, pour les hommes entre 30 et 59 ans, avec antagoniste automobiliste quand il y a la présence de tiers, pour un accident non mortel, en agglomération et en zone Police.**

Type d'usager	Présence de tiers	Hospitalisation	P(MAIS3+)
2RM	avec tiers	Hospitalisé	0,46
2RM	avec tiers	Non hospitalisé	0,05
2RM	sans tiers	Hospitalisé	0,53
2RM	sans tiers	Non hospitalisé	0,12
Vélo	avec tiers	Hospitalisé	0,40
Vélo	avec tiers	Non hospitalisé	0,03
Vélo	sans tiers	Hospitalisé	0,45
Vélo	sans tiers	Non hospitalisé	0,03
VL	avec tiers	Hospitalisé	0,22
VL	avec tiers	Non hospitalisé	0,01
VL	sans tiers	Hospitalisé	0,33
VL	sans tiers	Non hospitalisé	0,04
Piéton	avec tiers	Hospitalisé	0,45
Piéton	avec tiers	Non hospitalisé	0,07
Autre	avec tiers	Hospitalisé	0,20
Autre	avec tiers	Non hospitalisé	0,00001
Autre	sans tiers	Hospitalisé	0,39
Autre	sans tiers	Non hospitalisé	0,04



Ce sont ces probabilités qui vont être appliquées aux données BAAC.

## 2.5 Étape 1B : Application du modèle de prédiction du MAIS aux BAAC

Le modèle de prédiction du MAIS est appliqué aux données des forces de l'ordre. Plus précisément, le modèle est appliqué à des sous-groupes formés selon tous les profils d'accidents possibles. Ces profils sont définis par la combinaison des variables explicatives incluses dans le modèle. Pour chaque profil, les probabilités prédites associées sont appliquées à l'effectif des blessés du profil, donnant ainsi les effectifs prédits de blessés MAIS 1, MAIS 2, MAIS 3+.

### 2.5.1 Étape 1B : Application du modèle de prédiction du MAIS aux BAAC du Rhône

Le modèle de prédiction du MAIS est dans un premier temps appliqué aux données 2006-2012 des forces de l'ordre du Rhône, « BAAC seulement », afin de préparer l'étape 2 de capture-recapture. Le modèle de capture-recapture (qui doit être basé sur l'ensemble des enregistrements des données des forces de l'ordre et du Registre, sans exclusion) pourra alors inclure la variable MAIS. En effet, cette variable, même prédite, est indispensable au modèle capture-recapture car elle joue fortement sur la probabilité d'enregistrement des blessés.

### 2.5.2 Étape 1B : Application du modèle de prédiction du MAIS aux BAAC France

Le modèle de prédiction du MAIS est appliqué aux données nationales 2006-2015 des forces de l'ordre, afin de préparer l'étape 2B de projection. En effet, lors de la projection proprement dite, les coefficients de correction du sous-enregistrement, définis notamment selon trois niveaux du MAIS, pourront être alors appliqués aux données nationales des forces de l'ordre.


## 2.6 Étape 2 - préalable : Situation de l'enregistrement dans le Rhône

L'étape 2 de capture-recapture consiste, à partir de l'existence des deux sources d'enregistrement, à estimer le nombre de blessés non-enregistrés quelle que soit la source (forces de l'ordre ou Registre), et ainsi le nombre de blessés total.

Sur la période 2006-2012, les BAAC du Rhône ont enregistré en moyenne 2 800 blessés par an, et le Registre 7 400. Les blessés identifiés comme communs aux deux sources (« reliés ») sont au nombre de 1 700 par an. Le tableau ci-après reprend cela

*Tableau 14 : répartition du nombre de blessés du Rhône, selon la source (BAAC ou Registre, effectifs annuels moyen, 2006-2012*

		Dans le Registre ?		Total
		Oui	Non	
Dans les BAAC ?	Oui	1 700	1 100	2 800
	Non	5 700	n	
Total		7 400		8 500 + n

 = ensemble des blessés recensés, « BAAC union Registre »

L'ensemble des blessés observés est constitué des blessés identifiés par l'une et/ou l'autre source ; en langage mathématique ensembliste, on parle de BAAC union Registre (en grisé). L'ensemble des blessés recensés est donc de 8 500 blessés/an. Cela correspond à 3 fois plus que les BAAC.

Les tableaux ci-après indiquent la situation de l'enregistrement des blessés de la route dans les BAAC et dans le Registre au niveau du Rhône, en les déclinant selon les variables qui jouent le plus sur la probabilité d'enregistrement. Le ratio Ensemble/BAAC, qui est le rapport entre le nombre de blessés dans « BAAC union Registre » et le nombre de blessés dans les BAAC, est une première approximation des coefficients de sous-enregistrement que nous obtiendrons par le modèle de capture recapture.

À noter que pour les blessés identifiés par le Registre seulement, la force de l'ordre théoriquement en charge d'enregistrer tel ou tel blessé est la force de l'ordre déterminée en fonction de l'adresse où a eu lieu l'accident en termes de commune et type de réseau.

Le tableau ci-après montre que la propension des forces de l'ordre à enregistrer les blessés n'est pas la même d'un type de force de l'ordre à l'autre.

*Tableau 15 : effectifs annuels moyens des blessés selon la source et le type de force de l'ordre, Rhône, 2006-2012*

	BAAC	Registre	Reliés	Ensemble	Ratio (Ens/BAAC)
CRS	524	673	305	892	1,7
Gendarmerie	549	2 690	355	2884	5,3
Police	1 712	4 054	1 046	4 720	2,8
Total	2 785	7 417	1 706	8 496	3,1

Ces disparités entre types de forces de l'ordre, recouvrent aussi des disparités selon le classement hospitalisé (oui/non). Ainsi, comme le montre le tableau ci-après, les taux d'enregistrement diffèrent peu entre les forces de l'ordre chez les blessés classés hospitalisés (par les forces de l'ordre), mais sont très différents chez les blessés classés non hospitalisés (par les forces de l'ordre). La prise en compte de ce facteur dans le modèle d'enregistrement des blessés est donc indispensable.

*Tableau 16 : effectifs annuels moyens des blessés selon la source, le type de force de l'ordre et la gravité des blessés, Rhône, 2006-2012*

	BAAC	Registre	reliés	Ensemble	Ratio (E/B)
CRS, bl. hospitalisé	104	94	71	128	1,2
CRS, bl. non hosp.	420	579	234	765	1,8
Gend., bl. hospitalisé	345	527	244	627	1,8
Gend., bl non hosp.	204	2 163	110	2 257	11,0
Police, bl. hospitalisé	419	517	301	635	1,5
Police, bl. non hosp.	1 293	3 538	745	4 085	3,2
Total	2 785	7 417	1 706	8 496	3,1

La probabilité d'enregistrement dans les BAAC dépend aussi du type d'usager et de l'existence ou non d'un tiers dans l'accident. Le tableau ci-après donne les effectifs selon les BAAC, le Registre, et l'ensemble.

Les accidents sans tiers sont bien moins enregistrés dans les BAAC que les accidents avec tiers, et l'écart dépend du type d'usager. On note le cas très particulier des cyclistes blessés dans un accident sans tiers : ils sont très rarement enregistrés dans les BAAC (6/an dans les BAAC versus 855/an dans le Registre).

On comprend bien que les accidents avec tiers soient mieux déclarés que les accidents sans tiers : il y a un enjeu supplémentaire entre deux protagonistes (qui est responsable et qui ou quelle assurance paiera les frais).

*Tableau 17 : effectifs annuels moyens des blessés selon la source, et selon le type d'usager et la présence d'un tiers, Rhône, 2006-2012*

	BAAC	Registre	Reliés	Ensemble	Ratio (E/B)
2RM, avec tiers	582	918	390	1110	1,9
2RM, sans tiers	100	1050	65	1085	10,9
Vélo, avec tiers	138	301	88	351	2,5
Vélo, sans tiers	6	855	4	857	136,3
VL, avec tiers	1 072	2 328	619	2 780	2,6
VL, sans tiers	273	1 012	177	1 108	4,1
Piéton, avec tiers	449	705	277	877	2,0
Autre, avec tiers	117	114	61	170	1,4
Autre, sans tiers	48	134	24	158	3,3
Total	2 785	7 417	1 706	8 496	3,1

La probabilité d'enregistrement des blessés dans les BAAC dépend du type de réseau, pour certaines forces de l'ordre. Cela ne semble pas jouer en zone CRS, car ils n'ont en charge que des accidents sur autoroutes ou voies rapides. Cela joue peu en zone police, mais en revanche, en zone gendarmerie, les blessés sur voies communales sont bien moins enregistrés que les blessés sur les autres réseaux, c'est-à-dire les autoroutes, les routes nationales (devenues départementales en 2006) et donc l'ensemble des départementales.

*Tableau 18 : effectifs annuels moyens des blessés selon la source, et selon la force de l'ordre et le réseau, Rhône, 2006-2012*

	BAAC	Registre	reliés	Ensemble	Ratio (E/B)
CRS AR	334	432	196	571	1,7
CRS RN-RD	177	236	104	309	1,7
Gend AR	27	44	17	55	2,0
Gend RN-RD	383	1 190	254	1 318	3,4
Gend VC	123	1 388	75	1 436	11,7
Gend autre	16	68	9	75	4,6
Police RN-RD	117	212	72	257	2,2
Police VC	1 553	3 755	949	4 359	2,8
Police autre	38	85	23	100	2,6
Total	2 785	7 417	1 706	8 496	3,1

AR= autoroute, RN-RD= routes nationales (devenues départementales en 2006) et Routes départementales, VC= Voies Communales

La probabilité d'enregistrement des blessés dans les BAAC dépend de la gravité du blessé qui peut être dorénavant mesurée par la variable MAIS prédite. Elle dépend en même temps du type de force de l'ordre. Comme le tableau ci-après le montre l'écart d'enregistrement des blessés légers et graves au sens MAIS est faible en zone CRS, du simple au double en zone police, et très important en zone gendarmerie.

**Tableau 19 : effectifs annuels moyens des blessés selon la source, et selon la force de l'ordre et la gravité de la blessure prédite, Rhône, 2006-2012**

	BAAC	Registre	reliés	Ensemble	Ratio (E/B)
CRS MAIS1,2	491	642	293	840	1,7
CRS MAIS3+	34	31	23	42	1,2
Gend MAIS1,2	403	2 468	262	2 609	6,5
Gend MAIS3+	167	220	119	268	1,6
Police MAIS1,2	1 502	3 812	929	4 385	2,9
Police MAIS3+	212	248	149	311	1,5
Total	2 809	7 421	1 775	8 455	3,0

La probabilité d'enregistrement dans les BAAC dépend aussi, à la fois de la gravité MAIS prédite et du type d'utilisateur (tableau ci-après). Les nombres de blessés MAIS2+ sont à multiplier par des coefficients variant de 1,8 (usagers autres) à 9,4 (cyclistes). Les nombres de blessés MAIS3+ sont à multiplier par des coefficients variant de 1,3 à 3,9.

**Tableau 20 : effectifs annuels moyens des blessés selon la source, et selon le type d'utilisateur et la gravité de la blessure prédite, Rhône, 2006-2012**

	BAAC	Registre	reliés	Ensemble	Ratio (E/B)
2RM MAIS1,2	515	1 803	340	1 978	3,8
2RM MAIS3+	168	202	123	247	1,5
Vélo MAIS1,2	119	1 075	76	1 118	9,4
Vélo MAIS3+	22	79	16	85	3,9
VL MAIS1,2	1 266	3 227	775	3 718	2,9
VL MAIS3+	110	114	77	147	1,3
Piéton MAIS1,2	346	615	213	748	2,2
Piéton MAIS3+	98	91	66	123	1,3
Autre MAIS1,2	149	201	80	270	1,8
Autre MAIS3+	14	13	9	18	1,3
Total	2 807	7 420	1 775	8 452	3,0

Comme nous avons vu que la probabilité d'enregistrement dans les BAAC dépend aussi de la présence/absence d'un tiers, nous croisons aussi avec cette variable (ci-après).

*Tableau 21 : effectifs annuels moyens des blessés selon la source, et selon la gravité de la blessure prédite, le type d'usager et la présence ou non de tiers, Rhône, 2006-2012*

	BAAC	Registre	reliés	Ensemble	Ratio (E/B)
2RM Avec tiers MAIS1,2	450	801	299	952	2,1
2RM Avec tiers MAIS3+	132	122	97	157	1,2
2RM Sans tiers MAIS1,2	65	1 002	41	1 026	15,8
2RM Sans tiers MAIS3+	36	80	26	90	2,5
Vélo Avec tiers MAIS1,2	115	277	74	318	2,8
Vélo Avec tiers MAIS3+	21	24	15	30	1,4
Vélo Sans tiers MAIS1,2	5	798	3	800	160,0
Vélo Sans tiers MAIS3+	2	55	1	56	28,0
VL Avec tiers MAIS1,2	1 030	2 268	619	2 679	2,6
VL Avec tiers MAIS3+	71	63	48	86	1,2
VL Sans tiers MAIS1,2	236	959	156	1 039	4,4
VL Sans tiers MAIS3+	41	51	30	62	1,5
Piéton Avec tiers MAIS1,2	346	615	213	748	2,2
Piéton Avec tiers MAIS3+	98	91	66	123	1,3
Autre Avec tiers MAIS1,2	109	106	60	155	1,4
Autre Avec tiers MAIS3+	6	5	4	7	1,2
Autre Sans tiers MAIS1,2	40	95	20	115	2,9
Autre Sans tiers MAIS3+	8	8	5	11	1,4
Total	2 811	7 420	1 777	8 454	3,0

Nous constatons que, même pour les blessés graves (MAIS3+ prédits), l'enregistrement dépend à la fois du type d'usager et de la présence/absence d'un tiers ; de plus les écarts d'enregistrement sont variables. Ils varient de manière mesurée chez les piétons, usagers « autres » et automobilistes. Ils varient plus amplement chez les usagers de deux-roues motorisés, et très fortement chez les cyclistes. Toutes ces variables qui jouent en univarié, bivarié, et « trivarié » sur la probabilité d'enregistrement dans les BAAC seront considérées pour l'inclusion dans le modèle multivarié (cf. point 2.9.4).


## 2.7 Étape 2 : méthode de correction du sous-enregistrement et de ses biais

L'approche de capture-recapture est d'abord utilisée au niveau du département du Rhône, où les victimes de la route sont enregistrées par deux sources : d'une part par les forces de l'ordre, et d'autre part par le Registre des victimes d'accidents de la route dans le Rhône. Le chaînage entre les 2 sources, c'est-à-dire l'identification des victimes communes aux BAAC et au Registre, permet aussi de voir que certaines victimes ne sont enregistrées que par une seule source. Cette défaillance dans l'enregistrement de l'une ou de l'autre source, indique que des victimes peuvent échapper aux deux enregistrements.

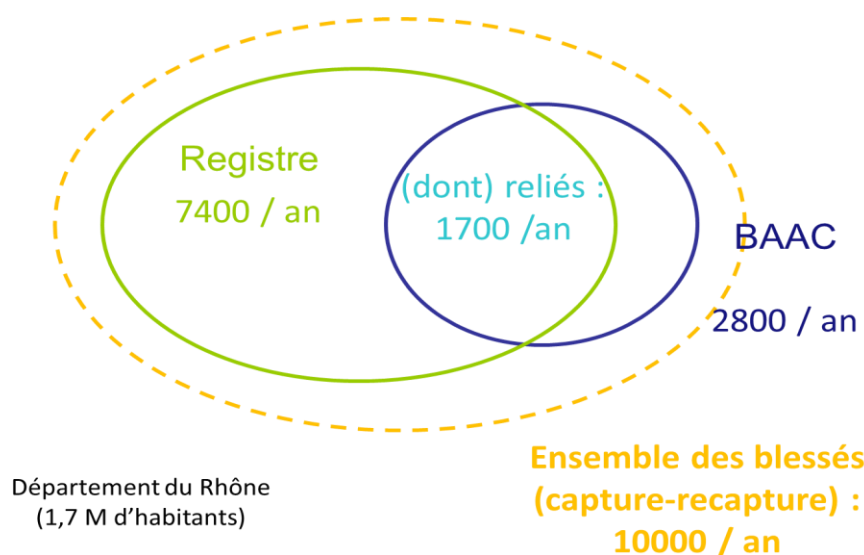
Cela se conçoit encore mieux avec le tableau ci-dessous :

*Tableau 22 : répartition du nombre de blessés du Rhône, selon la source (BAAC ou Registre, effectifs annuels moyen, 2006-2012*

		Dans le Registre ?		Total
		Oui	Non	
Dans les BAAC ?	Oui	1 700	1 100	2 800
	Non	5 700	n	
Total		7 400		8 500 + n

 = ensemble des blessés recensés, « BAAC union Registre »

La méthode de capture-recapture, peut alors, sous certaines conditions, estimer le nombre de personnes échappant aux deux recensements. On estime ainsi le nombre réel de victimes de la route dans le Rhône (cf. figure ci-dessous : sur-ensemble orange).



*Figure 3 : capture-recapture, département du Rhône – Effectifs annuels moyens 2006-2012*

Cette estimation de l'effectif du sur-ensemble se fait directement, par modélisation. Nous pouvons alors en déduire des coefficients de correction entre les effectifs des BAAC du Rhône et l'effectif du sur-ensemble total.

Intervient alors la méthode de projection : nous appliquons les coefficients de correction des BAAC, estimés sur le Rhône, aux BAAC nationaux (figure ci-dessous). Cela ne peut se faire que sous l'hypothèse d'homogénéité des pratiques d'enregistrement des blessés de la route pour un type de force de l'ordre donné.

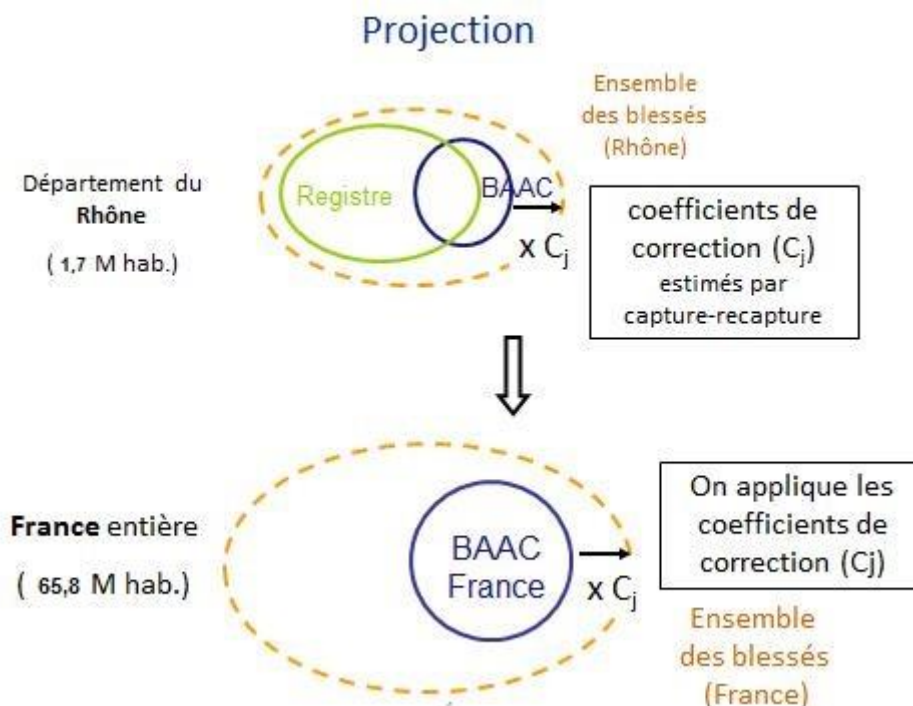


Figure 4: projection, du Rhône vers la France métropolitaine (effectifs de population en 2014)

Nous ne pouvons pas être certains de cette homogénéité des pratiques d'enregistrement des forces de l'ordre. Nous avons exploré le taux de remplissage de certaines variables des BAAC, notamment « sensibles », comme le port de la ceinture ou du casque, et nous avons noté un certain gradient Nord-Sud : les taux de valeur manquante les plus élevés se trouvent dans les départements du Sud de la France, et les moins élevés dans le Nord de la France. Il ne s'agit pas exactement du même processus (complétion d'une variable versus complétion d'un procès-verbal) mais c'est un indice en faveur du département du Rhône, élégamment situé entre le Sud et le Nord de la France : on peut donc penser qu'il représente un bon taux moyen du sous-enregistrement.

## 2.8 Étape 2A : Amélioration du chaînage entre BAAC et Registre, Rhône

Avant de modéliser le sous-enregistrement par l'approche capture-recapture, étape 2, il est nécessaire de nous approcher au plus près d'une des conditions de cette approche, à savoir, l'identification parfaite des sujets communs aux deux sources, ici, BAAC et Registre rhodaniens.

Cela étant difficilement atteignable, et difficilement mesurable, nous estimons de façon aussi fiable que possible, la répartition des blessés selon leur source d'enregistrement, en distinguant :

- les blessés communs aux BAAC et au Registre, rhodaniens, notés « intersection BAAC-Registre »,
- les blessés présents seulement dans les BAAC, notés « BAAC seulement »,
- les blessés présents seulement dans le Registre, notés « Registre seulement ».

Nous travaillons en termes d'effectifs (de blessés), i.e. de données agrégées, et non sur les blessés eux-mêmes, i.e. les données individuelles.



Département du Rhône  
(1,7 M d'habitants)

Figure 5 : répartition des blessés recensés par le Registre et par les BAAC, Rhône, effectifs annuels moyens sur 2006-2012

Lors de la procédure de rapprochement ou chaînage (record-linkage), on compare les victimes des BAAC (Rhône) à celles du Registre en formant des paires constituées d'un enregistrement d'une victime des BAAC et d'un enregistrement d'une victime du Registre. Ces paires peuvent être classées de la façon suivante, dans le tableau ci-après.

Tableau 23 : description d'un résultat de chaînage entre deux fichiers d'enregistrement (unité=une paire d'enregistrements)

décision :	situation réelle :	
	même individu (Matched)	individus différents (Unmatched)
rapprochement	vrais positifs	faux positifs
non-rapprochement	faux négatifs	vrais négatifs

Positif/négatif se réfère au résultat de la décision : positif signifie « rapproché » ou « relié » ou « chaîné », et négatif signifie non-rapproché. Faux et vrai qualifient la décision par rapport à la réalité (mais non connue). Dans la terminologie du record-linkage, l'ensemble des paires d'enregistrements correspondant (réellement) au même individu est noté M pour Matched, et l'ensemble des paires d'enregistrements correspondant (réellement) à des individus distincts est noté U pour Unmatched. Matched et Unmatched constituent ensemble le gold standard. Enfin, « relié » se traduit par « linked ».

Les faux positifs sont les paires d'enregistrements reliées à tort, car ne correspondant pas à la même victime. Les faux négatifs sont les paires non-reliées, à tort, c'est-à-dire deux enregistrements non reliés alors qu'ils correspondent à la même victime (Clark, 2004).

Nous allons améliorer le résultat du chaînage par l'estimation du nombre de faux positifs et de faux négatifs. Les calculs et les hypothèses faits sont détaillés en annexe.



Le tableau ci-dessous donne un premier résultat : l'effectif de faux positifs, c'est-à-dire, parmi les blessés rapprochés entre Registre et BAAC, le nombre de ceux qui sont rapprochés à tort, c'est-à-dire qui ne correspondent pas à la même personne.

*Tableau 24 : Proportion sur U = unmatched de faux positifs et leurs effectifs ; résultats du chaînage Registre et BAAC, Rhône, par année, sur 1996-2004 et 2006-2012*

critère	1996	1997	1998	1999	2000	2001	2002	2003	2004
P(FP)	2,0 E-6	2,1 E-6	2,1 E-6	2,0 E-6	2,1 E-6	2,1 E-6	1,7 E-6	1,5 E-6	1,4 E-6
# FP	92	90	107	105	94	92	56	38	34

critère	2006	2007	2008	2009	2010	2011	2012
P(FP)	1,5 E-6	1,5 E-6	1,4 E-6	1,4 E-6	1,4 E-6	1,3 E-6	1,3 E-6
# FP	32,8	33,7	25,3	30,3	29,1	24,2	26,6

On constate que les proportions et les effectifs de faux positifs sont plus faibles à partir de 2002, ce qui est cohérent avec le fait que la stratégie de chaînage est devenue à cette date plus restrictive.

Le tableau suivant donne le nombre de faux négatifs, c'est-à-dire, parmi les blessés non rapprochés, l'effectif de ceux non rapprochés à tort, qui sont en fait communs au BAAC et au Registre : ce sont deux enregistrements de la même victime.

*Tableau 25 : proportions sur M (=Matched), de faux négatifs, et leur effectif ; résultats du chaînage Registre et BAAC, Rhône, par année, 1996-2004 et 2006-2012*

critère	1996	1997	1998	1999	2000	2001	2002	2003	2004
P(FN)	15,1%	13,9%	13,0%	13,1%	13,3%	12,9%	17,6%	17,6%	17,6%
# FN	505	408	457	445	410	403	417	393	369

critère	2006	2007	2008	2009	2010	2011	2012
P(FN)	17,60%	17,60%	17,60%	17,60%	17,60%	17,60%	17,60%
# FN	375	361	332	346	379	354	358

On note que les proportions de faux négatifs sont plus élevées à partir de 2002, ce qui est attendu pour une stratégie de chaînage plus restrictive.

En termes d'effectifs, on n'observe pas cela, car le nombre réel de blessés a diminué (baisse visible que ce soit sur le Rhône ou la France, dans le Registre et dans les BAAC).

### 2.8.1 Redéfinition des effectifs résultant du chaînage

Les effectifs estimés de faux négatifs et de faux positifs permettent d'améliorer le résultat du chaînage. Plus précisément cela permet de redéfinir les tailles de chaque sous-ensemble : « BAAC seulement », « Registre seulement » et « intersection BAAC-Registre » (cf. figure ci-après).

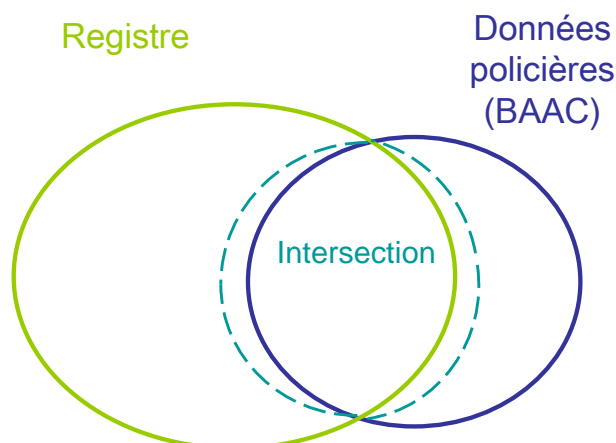


Figure 6 : redéfinition de la taille de l'intersection BAAC-Registre, Rhône

L'intersection BAAC-Registre doit être diminuée de l'effectif des faux positifs et augmentée de l'effectif des faux négatifs. Les sous-ensembles « BAAC seulement », « Registre seulement » doivent être chacun augmentés de l'effectif des faux positifs et diminués de l'effectif des faux négatifs. Cela s'écrit (avec l'astérisque signifiant nouvel effectif) :

$$n_{BAAC\ seul}^* = n_{BAAC\ seul} - n_{Faux\ Neg} + n_{Faux\ Pos}$$

$$n_{Reg\ seul}^* = n_{reg\ seul} - n_{Faux\ Neg} + n_{Faux\ Pos}$$

$$n_{intersection}^* = n_{intersection} + n_{Faux\ Neg} - n_{Faux\ Pos}$$

Nous donnons ci-dessous le détail des calculs.

*Pour rappel, un faux négatif est une paire constituée d'un individu de « BAAC seulement » et d'un individu de « Registre seulement » non reliés à tort. Ainsi l'effectif de faux négatifs est égal à l'effectif d'individus de « BAAC seulement » non reliés à tort, et aussi égal à l'effectif d'individus de « Registre seulement » non reliés à tort. De même, l'effectif de faux positifs est un nombre de paires, mais aussi égal au nombre de blessés du Registre reliés à tort, et égal au nombre de blessés des BAAC reliés à tort.*

*Nous rappelons que nous travaillons sur des données agrégées, et que, à ce stade, nous n'avons besoin que de la variable d'appartenance à l'un des trois sous-ensembles et des variables qui seront utilisées dans le modèle de capture-recapture. Sur la période 2006-2012 : il s'agit des principaux facteurs de biais de sélection : type de force de l'ordre, l'année, type de réseau, gravité de l'accident (mortel/corporel), type d'usager, présence/absence d'un tiers, hospitalisation (oui/non), et gravité MAIS.*

Nous notons les sous-ensembles exclusifs de la façon suivante :

- k=1 « BAAC seulement »
- k=2 « Registre seulement »
- k=3 « intersection BAAC-Registre »

Nous notons  $j$  les strates définies par la combinaison des principaux facteurs de biais de sélection. Nous notons  $p_{kj}$  la répartition des strates  $j$  dans le sous-ensemble  $k$ , c'est-à-dire :

$$p_{kj} = n_{kj} / n_k$$

Dans chacun des trois sous-ensembles exclusifs, nous avons besoin de la nouvelle répartition des facteurs de biais de sélection. Pour un sous-ensemble donné  $K$ , celle-ci est supposée être la même qu'avant la correction par les effectifs de faux négatifs et de faux positifs ; en d'autres termes :

$$p_{kj}^* = p_{kj}$$

Cette hypothèse vient du fait que les faux négatifs et les faux positifs ne sont ainsi que par manque d'efficacité de la procédure de chaînage et pas pour une autre raison. Par exemple : les blessés qui sont réellement dans l'intersection BAAC-Registre mais n'avaient pas été identifiés comme tels par la procédure de chaînage (pour cause d'erreurs dans les variables de chaînage) ressemblent aux blessés de l'intersection (en termes de caractéristiques de l'accident et du blessé) et ne ressemblent pas aux blessés de « BAAC seulement » ou de « Registre seulement ».

Dans chaque sous-ensemble, il suffit d'appliquer, au nouvel effectif, la répartition (croisée de la gravité du blessé, de l'accident, du type de force de l'ordre, du type d'utilisateur, tiers, réseau et année) observée avant la réaffectation de ceux qui appartiennent à ce groupe mais n'avaient pas été identifiés comme tels, c'est-à-dire :

$$n_{kj}^* = p_{kj} \times n_k^*$$

Le tableau ci-après donne un exemple des effectifs corrigés des faux négatifs et faux positifs du chaînage.

**Tableau 26 : effectifs moyens annuels de blessés selon la source, et corrigés des nombres de faux positifs et faux négatifs du chaînage, 2006-2012**

Type d'utilisateur et tiers (oui/non)	BAAC	Registre	Intersection (reliés)	BAAC U Registre (ensemble)	BAAC U Registre (corrigé)
2RM avec Tiers	582	918	390	1110	1098
2RM sans Tiers	100	1050	65	1085	1031*
vélo Avec tiers	138	301	88	351	341
vélo Sans tiers	6	855	4	857	808*
VL Avec Tiers	1 072	2 328	619	2 780	2 665
VL Sans Tiers	273	1 012	177	1 108	1 066
Piéton Avec Tiers	449	705	277	877	853
Autre Avec Tiers	117	114	61	170	163
Autre Sans Tiers	48	134	24	158	150
Total	2 785	7 417	1 706	8 496	8 174

\* situation problématique : l'effectif corrigé est inférieur à l'effectif observé dans le Registre

L'effectif BAAC Union Registre corrigé du nombre de faux positifs et faux négatifs, est plus faible que l'effectif non corrigé. Cela correspond à un nombre de reliés corrigés plus grand que le nombre de

reliés non corrigés. Cela correspond bien à une stratégie plus restrictive de chainage. L'écart est de 3,7% en moyenne annuelle sur 2006-2012.

## 2.9 Étape 2B : Modélisation du sous-enregistrement par capture-recapture

Lorsque deux (ou plusieurs) sources d'enregistrement d'une même population existent, il est possible avec la méthode de capture-recapture, sous certaines conditions, d'estimer le nombre de personnes échappant aux deux sources et ainsi l'effectif total de la population étudiée. La méthode de capture-recapture et son application en épidémiologie sont exposées brièvement ici.

### 2.9.1 Synthèse bibliographique sur capture-recapture

Trois articles (Hook and Regal, 1995; IWGDMF, 1995a; Gallay et al., 2002) fournissent une revue bibliographique particulièrement intéressante de la méthode de capture-recapture, et avec un point de vue épidémiologique ; nous présentons ici une brève synthèse bibliographique de son utilisation. La méthode de capture-recapture a été mise en place en écologie animale, pour quantifier des populations de poissons, d'oiseaux... Par exemple, s'il s'agit de quantifier la population de poissons dans un étang, le principe est de faire une première capture (pêche), de marquer les individus capturés, de les relâcher dans l'étang, puis de faire une deuxième capture, et parmi ceux capturés, d'identifier s'il s'agit de leur première capture, ou de leur deuxième (déjà marqués). À partir des effectifs de chaque sous-groupe (capturés seulement la 1<sup>ère</sup> fois, seulement la 2<sup>ème</sup> fois, ou les deux fois), et de calculs de probabilités, sous certaines hypothèses à vérifier, on estime l'effectif de ceux jamais capturés et donc l'effectif total.

On peut ensuite transposer les captures à des listes d'enregistrement, et les présenter selon le schéma ci-dessous.

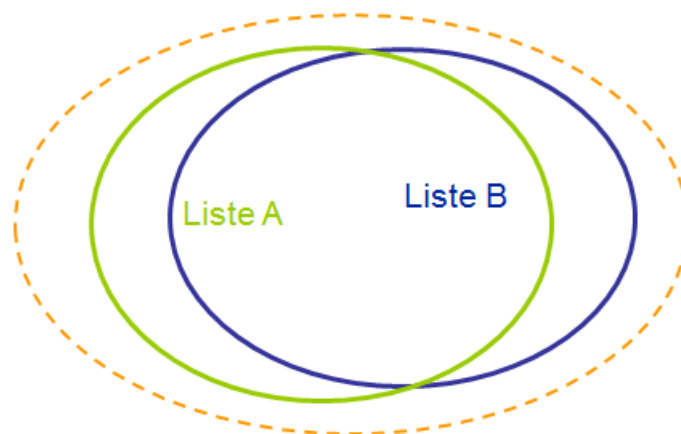


Figure 7 : schéma de deux recensements (ou « listes »)

Une autre façon de présenter les choses consiste à les écrire sous forme d'un tableau :

**Tableau 27: répartition des effectifs selon l'appartenance des sujets à deux sources d'enregistrement**

		Liste B		
		oui	non	
Liste A	oui	$n_{AB}$	$n_{A\bar{B}}$	$n_A$
	non	$n_{\bar{A}B}$	$n_{\bar{A}\bar{B}}$	
		$n_B$		$n$

Si l'on suppose l'indépendance d'enregistrement entre la liste A et la liste B, autrement dit, que la proportion d'appartenance à la liste A, parmi la population totale, est la même que la proportion (d'appartenance à la liste A) parmi ceux appartenant à la liste B, ce qui s'écrit  $n_A/n = n_{AB}/n_B$ , alors

on obtient l'estimateur intuitif de Petersen  $\hat{n} = \frac{n_A \times n_B}{n_{AB}}$  ; il est également l'estimateur du maximum

de vraisemblance.

L'approche a été utilisée dans d'autres domaines que l'écologie : en démographie, pour mesurer la couverture d'un recensement (Fienberg, 1992; Garton et al., 1996), en épidémiologie, surtout depuis les années 80 (Hook and Regal 1995; IWGDMF 1995; Gallay, Nardone et al. 2002). Elle sert à estimer des effectifs des populations, en particulier des populations « cachées » (Neugebauer and Wittes, 1994) : séropositifs (Frischer et al., 1991; Abeni et al., 1994), personnes atteintes du SIDA (Bernillon et al., 2000), toxicomanes (Hay et al., 1999; Maxwell, 2000; Gemmell et al., 2004; Brady et al., 2008; Leclerc et al., 2014; Espelt et al., 2015), prostituées (Bloor et al., 1991; Brewer et al., 2006), personnes sans domicile fixe (Fisher et al., 1994), enfants vivant dans la rue (Bezerra et al., 2011; Pullum et al., 2012) personnes subissant des violences domestiques (Oosterlee et al., 2009). Elle sert aussi à estimer le taux de couverture de registres ou encore à estimer des incidences corrigées de la sous-déclaration ou du sous-enregistrement. Les deux derniers objectifs sont essentiellement observés dans les domaines suivants : diabète (IWGDMF, 1995a, 1995b), malformations congénitales, cancer (Sherman, 1981; Robles et al., 1988; Hilsenbeck et al., 1992; Brenner et al., 1994; Schouten et al., 1994; Brenner et al., 1995; Crocetti et al., 2001), VIH et SIDA (Frischer et al., 1993; Abeni et al., 1994; Johri et al., 1998; Bernillon et al., 2000), traumatologie (Roberts and Scragg, 1994; LaPorte et al., 1995; Johnson et al., 1997; Rosenman et al., 2006; Siegler et al., 2008; Palusci et al., 2010).

La méthode de capture-recapture est utilisée en épidémiologie des traumatismes de la circulation routière. Elle a été appliquée sur les sous-populations suivantes : enfants (Roberts and Scragg, 1994; Jarvis et al., 2000), adolescents et jeunes adultes (Morrison and Stone, 2000), piétons et cyclistes (Dhillon et al., 2001; Tin et al., 2013), piétons en fauteuil roulant (Kraemer and Benton, 2015), conducteurs de poids lourds (Meuleners et al., 2006), mais aussi à des zones géographiques : villes (Razzak and Luby, 1998; Tercero and Andersson, 2004), région (Chokocho et al., 2013) ou une île (la Réunion) (Aptel et al., 1999). Une étude concernant les blessés d'accidents de trajets professionnels utilise également la méthode capture-recapture (Thomas et al., 2012). Elle est aussi utilisée pour évaluer la responsabilité de l'alcool dans les accidents de la circulation routière (Miller et al., 2012).

L'application de la méthode de capture-recapture en épidémiologie est sujette à des prises de position très tranchées. Certains chercheurs pensent qu'il est bien mieux d'appliquer cette méthode que de ne rien faire, et que tout Registre évaluant son taux de couverture doit l'utiliser. D'autres chercheurs sont franchement réfractaires et rejettent la méthode en bloc. Le sujet est polémique ; il faut dire que la méthode est victime de sa simplicité d'application : elle est parfois utilisée un peu trop vite, sans

vraiment tenir compte des conditions d'application. D'autres chercheurs tentent de calmer la polémique, en discutant point par point les conditions d'application. Une liste de 17 recommandations a notamment été écrite (Hook and Regal, 1999, 2000a) concernant l'application de la méthode, la vérification des conditions d'application, la présentation des étapes de l'analyse et des résultats. Dans le domaine de l'épidémiologie des traumatismes, une poignée de chercheurs se montre aussi très sceptique (Jarvis et al., 2000; Morrison and Stone, 2000) ; cependant, leur discussion des conditions d'applications n'apparaît pas complète.

Une étude récente concerne le sous-enregistrement de blessés répertoriés selon trois sources, les données policières, les données hospitalières, et « the injuries board » qui recense les réclamations de personnes souhaitant être indemnisées à cause de blessures dues à des accidents de la route (Short and Caulfield, 2016).

Concernant les méthodes d'estimation, il y a eu historiquement les estimateurs « simples » (nommés estimateur de Petersen, de Sekar et Deming...) limités aux situations à deux sources d'enregistrement (avec notamment l'appellation de « méthode à deux listes »), où il suffit d'une calculatrice pour estimer le nombre de sujets non-observés. S'est ensuite développée la modélisation explicite. Celle-ci permet de prendre facilement en compte des variables liées à la probabilité d'enregistrement, mais aussi, des éventuelles interactions entre les listes, lorsqu'il y en a trois ou plus. La modélisation la plus répandue est celle basée sur une loi de Poisson, avec un modèle log-linéaire (Cormack, 1989; IWGDMF, 1995a; Chao et al., 2001); elle est liée à l'analyse des tableaux de contingence. Une autre modélisation se base sur la loi multinomiale et le modèle logistique (Alho, 1990; Tilling and Sterne, 1999). Il existe aussi des développements non-paramétriques, bayésiens (Madigan and York, 1997; Hook and Regal, 2000b) et autres qui n'ont pas été explorés ici.

Nous appliquons ici l'approche capture-recapture grâce à une modélisation explicite, par un modèle multinomial logistique. La formalisation est donnée en annexe. Cela nous permet de prendre en compte l'ensemble des facteurs majeurs de biais de sélection, et aussi d'estimer directement les coefficients de correction du sous-enregistrement des données des forces de l'ordre.

### 2.9.2 Conditions d'application

L'approche capture-recapture se base sur quatre conditions clés d'application (Hook and Regal, 1995; IWGDMF, 1995a, 1995b; Gallay et al., 2002):

- 1) population fermée,
- 2) indépendance d'enregistrement entre les deux sources,
- 3) homogénéité de la probabilité de capture par une source donnée,
- 4) identification parfaite des sujets communs aux deux sources.

Deux autres conditions existent (Gallay et al., 2002) mais sont rarement explicitées, car considérées comme évidentes. Il n'est cependant pas inutile de les donner :

- 5) même période de temps et même zone géographique,
- 6) pas d'erreur de classement des cas.

**La condition 1** de population fermée, signifie qu'il n'y a pas d'entrée ou sortie entre les deux sources d'enregistrement. Sur l'exemple des poissons, cela signifie qu'il n'y a pas d'apport ou de perte de poissons entre les deux pêches, ni naissances ni décès.

Sur les blessés de la route, cela signifie qu'il ne doit pas y avoir d'entrée ou de sortie entre les enregistrements des forces de l'ordre et du Registre, c'est-à-dire entre l'établissement d'un procès-verbal par les forces de l'ordre d'un côté, et la consultation dans un service hospitalier. Cette condition est vérifiée en grande partie. Peu de temps s'écoule entre la venue des forces de l'ordre sur le lieu de

l'accident et la consultation à l'hôpital (au plus quelques jours pour les plus légèrement blessés). La seule possibilité de « sortie » concerne les personnes blessées dans le Rhône se rendant dans un hôpital non couvert par le Registre, c'est-à-dire hors du Rhône et de ses hôpitaux limitrophes. Ce seraient typiquement des blessés légers (MAIS 1) n'habitant pas le Rhône.

Combien sont-ils ? Une indication de la proportion de blessés impliqués dans un accident dans le département du Rhône mais habitant hors du Rhône est donnée dans les données des forces de l'ordre : 24 % de voitures impliquées dans un accident corporel sont immatriculées dans un département autre que le Rhône.

Parallèlement, parmi les blessés occupants de voiture recensés dans le Registre, 16 % habitent un autre département que le Rhône. S'il n'y avait aucun biais de sélection, ni du Registre ni des données des forces de l'ordre, les deux proportions devraient être égales, et éventuellement égales à la proportion dans les BAAC soit 24%. Nous pourrions supposer alors qu'il nous manque 8 % (24 %-16 %) des automobilistes blessés. Il faut faire de même pour chaque type d'utilisateur car la mobilité interdépartementale est bien sûr liée au mode de transport. La proportion de personnes habitant hors du Rhône est très faible parmi les blessés piétons et cyclistes. Elle est forte parmi les blessés usagers de véhicules utilitaires et poids lourds. **Cela conduit à ce que les estimations finales d'effectifs de blessés, surtout légers, soient des valeurs planchers.**

**La condition 2** d'indépendance d'enregistrement entre les deux sources signifie que les enregistrements sont faits de façon indépendante. Sur l'exemple des poissons, c'est le cas par défaut. Pour les blessés de la route, ce n'est pas le cas. Il y a une dépendance positive pour les blessés graves : en effet, en présence d'un blessé supposé grave, les forces de l'ordre appellent en général les pompiers ou le SAMU. La réciproque existe, mais dans une moindre mesure. Cette corrélation porte sur l'alerte et la présence sur le lieu de l'accident des équipes des forces de l'ordre et des équipes médicales. Cette corrélation « sur le terrain » entraîne une corrélation positive sur les probabilités d'enregistrement dans les données des forces de l'ordre et dans le Registre. Il a été démontré (IWGDMF, 1995a) qu'une corrélation positive se traduit par une sous-estimation du nombre total de blessés. **Nous avons ici une dépendance positive et donc les effectifs et incidences estimés de blessés graves seront vraisemblablement des valeurs planchers.**

**La condition 3** d'homogénéité de capture signifie que tous les individus d'intérêt ont la même probabilité d'être enregistrés par une source donnée. Sur l'exemple des poissons : que tous les poissons, quelle que soit leur taille, leur âge, ont la même probabilité d'être pêchés. En épidémiologie, il est en fait fréquent que l'homogénéité de capture ne soit vérifiée qu'à l'intérieur de sous-groupes, définis par exemple par les niveaux de gravité d'une maladie. Cela peut et doit être pris en compte dans la mise en application de la méthode, soit par stratification sur la variable définissant les sous-groupes soit par inclusion de la variable dans une modélisation explicite, ce que nous faisons dans notre approche (cf. Tableau 29). Pour les blessés de la route, ces variables correspondent à ce que nous appelons biais de sélection. Du côté de l'enregistrement dans les BAAC, les facteurs de biais de sélection les plus importants sont la gravité lésionnelle, le type d'utilisateur, la présence/absence de tiers, le type de réseau, et le type de force de l'ordre (Amoros et al., 2006) ; la gravité de l'accident (accident mortel ou corporel) s'est aussi avérée jouer sur la probabilité d'enregistrement des blessés impliqués dans ces accidents. En ce qui concerne les biais de sélection du Registre (Amoros, 2007), il s'agit de la gravité lésionnelle, de la distance entre le lieu d'accident et l'hôpital le plus proche, et éventuellement du type d'utilisateur (sous-notification des cyclistes), ces biais sont cependant de bien moindre ampleur que ceux des BAAC

**La condition 4** d'identification parfaite des sujets communs aux deux sources signifie que le chaînage entre les 2 sources d'enregistrement est parfait : on ne « loupe » aucun rapprochement entre 2 enregistrements de la même victime (i.e. il n'y a aucun « faux négatif » par rapport au chaînage), et pour chaque paire reliée entre les 2 sources, il s'agit bien de la même victime (i.e. il n'y a aucun « faux positif »).

En pratique, cette condition est difficile à satisfaire complètement, et à évaluer. Nous avons cependant amélioré le chaînage, dans l'étape 2, en estimant le nombre de faux positifs et de faux négatifs, et en corrigeant le résultat du chaînage à l'aide de ces effectifs.

**La condition 5** « même période de temps et même zone géographique » paraît une évidence. Dans le cas des poissons, cela signifie que les deux pêches doivent être très rapprochées dans le temps, et sur la même zone de l'étang.

La période de temps est ici 2006-2012. Le chaînage des victimes entre les données des forces de l'ordre et le Registre se fait année par année. La période de temps est donc bien la même entre les deux sources d'enregistrement.

La zone géographique est le département du Rhône, comme lieu d'accident. Du côté des données des forces de l'ordre, le département où s'est produit l'accident est toujours renseigné, et c'est cette variable qui sert dans les données nationales à identifier et à sélectionner les accidents ayant eu lieu dans le département du Rhône. Du côté du Registre, le critère d'inclusion géographique est aussi le département du Rhône, en tant que lieu d'accident. Cependant, dans le Registre, le lieu de l'accident (commune) est non renseigné pour un quart des victimes. On peut penser que par défaut les services hospitaliers dans le département du Rhône incluent soit toutes les personnes qui se présentent suite à un accident de la route, soit font une sélection sur « accident dans le Rhône » mais sans demander plus de précision ; en effet, la pertinence de la variable « lieu de l'accident » n'est pas évidente pour les soignants et il est donc compréhensible que cette variable soit peu renseignée. Le personnel du Registre envoie un courrier aux victimes du Registre lorsque cette information est manquante mais elle le reste chez 25,6 % des blessés (nom de la commune non renseigné).

Lors de l'enquête auprès des blessés identifiés seulement dans le Registre (Amoros, 2007), il est apparu que certains accidents, parmi ceux où le lieu n'était pas renseigné, avaient en fait eu lieu hors du Rhône. Cela correspond à 2 % des répondants, parmi les blessés par accident en 2001 (période enquêtée), identifiés seulement dans le Registre.

La condition n°5 de l'approche capture-recapture implique d'exclure ces victimes blessées dans un accident hors Rhône. Elles sont cependant difficilement identifiables.

En pratique nous avons procédé de la manière suivante : nous généralisons le pourcentage de tels blessés (parmi les répondants) à l'ensemble des blessés identifiés dans le Registre seulement, et à l'ensemble des années de la période étudiée. Nous appliquons ainsi ces 2 % aux blessés identifiés seulement dans le Registre, année par année ; cela nous donne un effectif de blessés à exclure. Nous tirons au sort ces blessés, parmi les blessés identifiés seulement dans le Registre ET dont le lieu de l'accident est manquant.

**La condition 6** « pas d'erreur de classement des cas » signifie que les cas qui nous intéressent (poissons de telle espèce, ou blessés de la route) sont parfaitement identifiables et identifiés, ou autrement dit, qu'ils sont des vrais cas (qu'il n'y ait ni faux négatifs ni faux positifs, par rapport à la définition d'un cas). Ici, il s'agit de s'assurer que les personnes vérifient les deux critères : « blessées » et « lors d'un accident de la circulation routière ».

Du côté des BAAC, l'étude (Amoros 2007) des blessés identifiés dans les BAAC seulement portant sur l'année 2011 a mis en évidence que certaines personnes classées « blessées » ne l'étaient pas : typiquement il s'agit de personnes accidentées envoyées dans un service hospitalier pour observation (femmes enceintes, nourrissons par exemple), ou de personnes en état de choc psychologique, mais non blessées physiquement (la classification AIS ne retient pas le choc psychologique comme une blessure), ou encore de personnes déclarées « blessées » par précaution vis-à-vis des assurances ou d'éventuelle procédure judiciaire mais qui se révèlent indemnes.

Afin de vérifier la condition n°6 de l'approche capture-recapture, il nous faut exclure ces personnes des données des forces de l'ordre. Nous procédons de la même façon que pour satisfaire à la condition n°5 : nous généralisons la proportion observée (4,1 %) sur les blessés de cette étude (ceux avec PV exploitable), à l'ensemble des blessés identifiés dans les BAAC seulement, et à l'ensemble de la période étudiée. Nous tirons au sort les personnes à exclure (année par année), parmi les blessés identifiés



seulement dans les BAAC et parmi ceux classés blessés légers (au sens BAAC). Nous supposons en effet que ceux classés blessés graves sont réellement blessés et que l'erreur ne porte que sur la frontière blessé léger/indemne.

En ce qui concerne le critère « accident de la circulation routière », il paraît peu probable que d'autres types d'accidents (professionnel hors route ou de loisirs ou de la vie courante, exclusivement) soient classés comme accidents de la circulation routière par les forces de l'ordre. Il arrive que des accidents de la route soient des suicides ; il est difficile de les identifier.

Du côté du Registre, les personnes incluses sont assurément blessées puisqu'un des critères d'inclusion du Registre est « une personne présentant au moins une lésion au sens de la classification AIS (gravité AIS 1 au minimum ou AIS 9 indéterminé, correspondant à une ou plusieurs lésions mal définies) ».

En ce qui concerne le critère « accident de la circulation routière » l'enquête sur les blessés identifiés dans le Registre seulement (Amoros, 2007) a montré que quelques personnes s'avéraient ne pas correspondre à cette définition (huit parmi 547, soit 1,2 % en tenant compte du plan de sondage). Il s'agit d'accidents professionnels ou de la vie courante ou de loisirs, exclusivement, ou encore d'accidents avec un moyen de transport non routier.

Afin de vérifier la condition n°6 de l'approche capture-recapture, nous généralisons le pourcentage de « hors accident de la route » à l'ensemble des blessés identifiés dans le Registre seulement, et à la période 2006-2012. Nous tirons au sort les personnes à exclure (année par année) parmi les blessés identifiés dans le Registre seulement.

Les critères de définition des données des forces de l'ordre, qui sont ceux utilisés par le ministère chargé des transports pour communiquer les bilans officiels en termes de sécurité routière, sont plus restrictifs que ceux du Registre, et ont été appliqués aux victimes du Registre. Nous rappelons que, pour cette raison, les usagers de rollers, skateboard, trottinette se blessant seuls ou contre un piéton, inclus dans le Registre, sont exclus de l'extrapolation. Pour la même raison, les victimes du Registre où les accidents sont clairement renseignés comme hors réseau routier sont exclues de l'extrapolation.

### **2.9.3 Construction du modèle (logit multinomial)**

Pour modéliser l'approche de capture-recapture, nous utilisons un modèle logistique multinomial ; il permet d'inclure facilement les variables liées aux probabilités d'enregistrement. Nous nous sommes fortement appuyés sur un article de Tilling et Sterne (Tilling and Sterne, 1999). La présentation est détaillée en annexe ; elle reprend une partie de leur article.

### **2.9.4 Modèle d'enregistrement des blessés sur 2006-2012**

Il s'agit ici de choisir les variables à inclure dans le modèle.

En premier lieu, les probabilités d'enregistrement dans les BAAC et dans le Registre ne sont pas indépendantes ; elles sont positivement corrélées pour les blessés graves : sur le terrain, pour les accidents graves, les forces de l'ordre alertent les secours médicaux s'ils ne sont pas déjà présents ; la réciproque existe sans doute dans une moindre mesure. En d'autres termes, les BAAC et le Registre sont dépendants à travers la caractéristique de gravité lésionnelle. Cette variable doit donc être incluse dans le modèle.

Par ailleurs, la modélisation permet de prendre en compte la restriction d'homogénéité de capture à des sous-groupes : cela consiste à inclure dans le modèle les variables définissant ces sous-groupes. Il s'agit donc ici des facteurs de sous-enregistrement. Du côté des BAAC, les facteurs liés à un fort biais de sélection sont la gravité lésionnelle, la gravité de l'accident (mortel/corporel), le type d'utilisateur, la présence/absence de tiers, le type de réseau, le type de force de l'ordre, et l'année pour un type de force de l'ordre (Amoros, 2007). Du côté du Registre, il s'agit de la gravité lésionnelle, éventuellement

du type d'utilisateur, et de la distance à l'hôpital le plus proche (disposant d'un service d'urgences) (Amoros, 2007). Cette variable n'est pas introduite car elle n'est pas directement disponible dans les BAAC (il est théoriquement possible de la déterminer à partir des lieux d'accident et des adresses d'hôpitaux disposant d'un service d'urgence mais il n'est pas envisageable en pratique de la déterminer pour la totalité des blessés des BAAC nationaux).

Nous avons considéré l'inclusion de la variable année en quantitatif : cela permet de prendre en compte une éventuelle évolution (linéaire) du phénomène de sous-enregistrement. Nous avons choisi de ne pas inclure la variable année en qualitatif dans le modèle car nous craignons un sur-ajustement aux éventuelles irrégularités des pratiques d'enregistrement dans le département du Rhône. Il s'avère que la variable année en quantitatif est seulement significative pour la zone gérée par la gendarmerie.

Nous avons également ajouté au modèle une autre variable en lien avec la gravité du blessé, à savoir si le blessé est hospitalisé ou non. En effet, cette variable est pour les BAAC, un facteur important de biais d'enregistrement (cf. Tableau 16).

Les variables finalement retenues sont :

- Le type de force de l'ordre (CRS, police, gendarmerie)
- L'année en quantitatif pour la zone gendarmerie
- Le type de réseau routier
- La gravité de l'accident dans lequel le blessé est impliqué (mortel/corporel)
- L'hospitalisation (oui/non) du blessé
- La gravité MAIS3+ (oui/non ; observée ou prédite) du blessé
- L'interaction entre la gravité MAIS et le type de force de l'ordre
- Le type d'utilisateur (piétons, cyclistes, usagers de deux-roues motorisé, automobilistes, autres)
- La présence/absence d'un tiers dans l'accident
- L'interaction entre type d'utilisateur, tiers (oui/non) et MAIS

**Valeurs manquantes (des variables incluses dans le modèle) :**

Étant donné que l'approche capture-recapture vise à estimer le nombre exhaustif des sujets d'intérêt, il n'est pas question d'exclure des observations ; les observations avec valeur manquante sont traitées par de l'imputation simple.

Les variables type de réseau et force de l'ordre présentent une proportion non négligeable de valeurs manquantes parmi les observations du Registre : 21,2 % et 22,5 % respectivement. Nous avons fait de l'imputation simple (Donders et al., 2006). Le type de réseau routier a été imputé en fonction du type d'utilisateur (piétons, cyclistes, usagers de deux-roues motorisé, automobilistes, autres) et de la force de l'ordre en charge de la zone géographique de l'accident (CRS, gendarmerie, police), car ils sont très corrélés. Plus précisément, nous partons de la répartition observée (sur les blessés où il n'y a pas de valeurs manquantes) de la variable réseau en fonction des deux autres (type d'utilisateur et force de l'ordre). Les proportions observées nous donnent les probabilités d'une loi multinomiale à 4 modalités (AR, RN-RD, VC, autres), selon le profil de blessés. Le tableau ci-après présente cela.

**Tableau 28 : Répartition fictive du réseau selon le type d'utilisateur et la force de l'ordre en charge de l'accident (valeurs fictives)**

Type d'utilisateur	Force de l'ordre	Réseau		
		RN-RD	VC	Autre
VL	Gendarmerie	60%	30%	10%
VL	Police	40%	40%	20%
2RM	Gendarmerie	40%	50%	10%
2RM	Police	30%	60%	10%
Vélo	Gendarmerie	20%	60%	20%
Vélo	Police	10%	60%	30%
Piéton	Gendarmerie	10%	60%	30%
Piéton	Police	5%	65%	30%
Autre	Gendarmerie	50%	40%	10%
Autre	Police	40%	50%	10%

Ainsi, pour chaque blessé où le réseau est manquant, nous regardons le type d'utilisateur (ex. cycliste) et le type de force de l'ordre (ex. gendarmerie) et nous utilisons alors la distribution multinomiale correspondante (ex. AR=0, RN-RD = 0,20, VC = 0,60, autre = 0,20) ; nous tirons au sort la modalité réseau selon cette loi et remplaçons la valeur manquante par cette modalité.

De la même manière, la force de l'ordre a été imputée en fonction du réseau et du type d'utilisateur.

Pour les observations où le réseau et la force de l'ordre sont tous les deux manquants : le réseau est imputé selon le type d'utilisateur, puis, le type de force de l'ordre est imputé selon le type d'utilisateur et le réseau imputé. Il faut en effet respecter la cohérence entre réseau et type de force de l'ordre.

Les valeurs manquantes pour âge, sexe et type d'utilisateur ont été imputées en fonction des deux autres variables ou d'une seule autre si une seule était renseignée (imputation simple dans les deux cas). Ainsi, les valeurs manquantes pour l'âge sont imputées selon une loi multinomiale définie en fonction du croisement des variables sexe et type d'utilisateur. Les valeurs manquantes pour le sexe sont imputées en fonction des variables âge et type d'utilisateur, et celles manquantes pour le type d'utilisateur en fonction de l'âge et du sexe.

Le tableau ci-après donne les paramètres du modèle retenu sur 2006-2012. Les odds-ratios sont ceux de la probabilité d'être enregistré dans « Registre seulement » (ou dans l'intersection BAAC-Registre) pour une caractéristique donnée versus une caractéristique de référence.

**Tableau 29 : modèle capture-recapture, logit multinomial et multivarié, sur 2006-2012 (n=57 219 blessés) ; OR=odds-ratio associés à la probabilité d'être enregistrés dans le Registre seulement, dans l'intersection BAAC et Registre, ou dans les BAAC seulement. Catégorie prise comme référence : BAAC seulement**

Variable	modalité	Source	beta	OR	IC à 95%		
Ordonnée à l'origine		Registre seul	-1,56				
		Intersection	1,98				
Année en quantitatif Gendarmerie		Registre seul	0,04	1,04	1,00	1,08	
		Intersection	0,00	1,00	0,96	1,04	
Type de réseau	Autoroute	Registre seul	-0,97	0,37	0,31	0,45	
	Autoroute	Intersection	0,21	1,24	1,04	1,49	
	RN-RD	Registre seul	-0,78	0,45	0,40	0,50	
	RN-RD	Intersection	0,18	1,20	1,07	1,34	
	Voie communale (réf)	Registre seul	0,00	1,00	.	.	
	Voie communale (réf)	Intersection	0,00	1,00	.	.	
	Autre	Registre seul	-0,18	0,83	0,65	1,07	
Gravité accident	Autre	Intersection	0,08	1,08	0,84	1,40	
	Corporel	Registre seul	2,13	8,42	4,83	14,68	
	Corporel	Intersection	-0,74	0,47	0,32	0,69	
	Mortel (réf)	Intersection	0,00	1,00	.	.	
Hospitalisé	Mortel (réf)	Registre seul	0,00	1,00	.	.	
	Oui (réf)	Registre seul	0,00	1,00	.	.	
	Oui (réf)	Intersection	0,00	1,00	.	.	
Non	Non	Registre seul	1,37	3,97	3,59	4,38	
	Non	Intersection	-0,47	0,62	0,56	0,68	
	Force ordre * MAIS	CRS MAIS 1-2	Registre seul	-1,66	0,18	0,15	0,22
	CRS MAIS 1-2	Intersection	-0,18	0,83	0,69	0,99	
CRS MAIS 3+	Registre seul	-0,16	0,84	0,52	1,37		
	Intersection	-0,19	0,82	0,55	1,23		
Police MAIS 1-2	Registre seul	-1,49	0,22	0,19	0,26		
	Intersection	0,03	1,04	0,88	1,22		
Police MAIS 3+	Registre seul	-0,65	0,51	0,39	0,67		
	Intersection	0,12	1,12	0,88	1,43		
Gendarmerie MAIS 1-2	Registre seul	21,47	-	.	.		
	Intersection	1,18	-	.	.		
Gendarmerie MAIS 3+ (réf)	Intersection	0,00	1,00	.	.		
	Registre seul	0,00	1,00	.	.		
Type d'usager * tiers * MAIS	2RM avec tiers MAIS 1-2	Registre seul	0,99	2,70	1,94	3,75	
	2RM avec tiers MAIS 1-2	Intersection	0,30	1,35	1,02	1,79	
2RM avec tiers MAIS 3+	Registre seul	-0,20	0,81	0,56	1,17		
	Intersection	0,26	1,30	0,97	1,73		
2RM sans tiers MAIS 1-2	Registre seul	3,55	35,0	24,1	50,7		
	Intersection	0,17	1,18	0,84	1,66		
2RM sans tiers MAIS 3+	Registre seul	2,05	7,79	5,04	12,0		
	Intersection	0,33	1,39	0,92	2,08		
Vélo avec tiers MAIS 1-2	Registre seul	1,44	4,23	2,95	6,04		
	Intersection	0,25	1,29	0,94	1,76		
Vélo avec tiers MAIS 3+	Registre seul	0,59	1,81	1,05	3,10		
	Intersection	0,11	1,12	0,69	1,80		
Vélo sans tiers MAIS 1-2	Registre seul	6,18	484	204	>999		
	Intersection	0,28	1,33	0,51	3,45		
Vélo sans tiers MAIS 3+	Registre seul	4,99	148	35,7	613		
	Intersection	0,52	1,69	0,37	7,66		
VL avec tiers MAIS 1-2	Registre seul	1,31	3,74	2,70	5,18		
	Intersection	0,05	1,05	0,80	1,39		
VL avec tiers MAIS 3+ (réf)	Registre seul	0,00	1,00	.	.		
	Intersection	0,00	1,00	.	.		
VL sans tiers MAIS 1-2	Registre seul	2,29	9,95	7,11	13,9		
	Intersection	0,24	1,27	0,95	1,70		
VL sans tiers MAIS 3+	Registre seul	1,16	3,21	2,06	5,00		
	Intersection	0,30	1,35	0,91	2,00		
Piéton avec tiers MAIS 1-2	Registre seul	1,02	2,78	1,99	3,88		
	Intersection	0,09	1,10	0,83	1,46		
Piéton avec tiers MAIS 3+	Registre seul	-0,22	0,79	0,54	1,16		
	Intersection	-0,00	0,99	0,73	1,35		
Autre avec tiers MAIS 1-2	Registre seul	-0,04	0,96	0,66	1,38		
	Intersection	-0,11	0,89	0,65	1,21		
Autre avec tiers MAIS 3+	Registre seul	0,04	1,04	0,36	3,01		
	Intersection	0,00	1,00	0,44	2,24		
Autre sans tiers MAIS 1-2	Registre seul	1,52	4,61	3,08	6,88		
	Intersection	-0,27	0,76	0,52	1,11		
Autre sans tiers MAIS 3+	Registre seul	0,65	1,92	0,90	4,12		
	Intersection	-0,41	0,66	0,32	1,33		

*Italique = variables en interaction : les odds-ratios ne peuvent pas être lus directement dans ce tableau.*

L'interprétation de l'odds-ratio est ici complexe. Ici, la cote est la probabilité d'être enregistré dans le Registre seulement contre la probabilité d'être enregistré dans BAAC seulement. Le ratio entre deux cotes est ici le ratio entre la cote d'un groupe de blessés (ex : « hospitalisé ») et la cote du groupe de blessés de référence (pour la variable hospitalisé, la référence est « non hospitalisé »).

Pour les variables du modèle qui ne sont pas en interaction avec d'autres, les odds-ratios peuvent être lus directement dans le tableau : réseau, gravité de l'accident, hospitalisation (oui/non).

Ainsi, pour le réseau, les blessés sur autoroute sont moins bien enregistrés dans Registre seulement que dans BAAC seulement, par rapport aux blessés sur voies communales, ce qui signifie à contrario, que les blessés sur autoroute sont mieux enregistrés dans les BAAC seulement que les blessés sur voies communales (catégorie de référence).

Concernant la gravité de l'accident, les blessés impliqués dans un accident corporel (et non mortel) sont bien mieux enregistrés dans le Registre seulement que dans les BAAC seulement, par rapport aux blessés impliqués dans un accident mortel. Cela indique donc le contraire pour les BAAC.

Nous avons choisi ici de prendre « BAAC seulement » comme catégorie de référence pour le modèle afin d'obtenir des odds-ratios supérieurs à 1, plus facile à appréhender que des odds-ratios compris entre 0 et 1 que nous obtiendrions si nous avions pris « Registre seulement » comme référence.

En ce qui concerne les blessés déclarés hospitalisés ou non hospitalisés par les forces de l'ordre : les non hospitalisés sont mieux enregistrés dans le Registre seulement que dans les BAAC seulement, par rapport aux blessés hospitalisés. A contrario cela indique que les blessés légers (non hospitalisés) ne sont pas très bien enregistrés dans les BAAC.

Pour les variables en interaction avec d'autres, le tableau suivant donne les coefficients bêta du modèle et les odds-ratios associés.

**Tableau 30 : Odds-ratios entre différents profils de blessés, associés à la probabilité qu'ils soient enregistrés seulement dans le Registre**

Profils de blessés	Comparaison	bêta	OR
Blessés MAIS3+, accident avec tiers	Usagers de 2RM versus automobilistes	-0,21	0,81
Blessés MAIS3+, accident avec tiers	Cyclistes versus automobilistes	0,59	1,80
Blessés MAIS3+, accident avec tiers	Piétons versus automobilistes	-0,23	0,80
Blessés MAIS3+, accident avec tiers	Autres versus automobilistes	0,04	1,04
Blessés MAIS1-2, accident avec tiers	Usagers de 2RM versus automobilistes	-0,33	0,72
Blessés MAIS1-2, accident avec tiers	Cyclistes versus automobilistes	0,12	1,13
Blessés MAIS1-2, accident avec tiers	Piétons versus automobilistes	-0,30	0,74
Blessés MAIS1-2, accident avec tiers	Autres versus automobilistes	-1,35	0,26
Blessés MAIS3+, accident sans tiers	Usagers de 2RM versus automobilistes	0,89	2,42
Blessés MAIS3+, accident sans tiers	Cyclistes versus automobilistes	3,83	46,06
Blessés MAIS3+, accident sans tiers	Autres versus automobilistes	-0,51	0,60
Blessés MAIS1-2, accident sans tiers	Usagers de 2RM versus automobilistes	1,26	3,53
Blessés MAIS1-2, accident sans tiers	Cyclistes versus automobilistes	3,88	48,60
Blessés MAIS1-2, accident sans tiers	Autres versus automobilistes	-0,77	0,46
Automobilistes, blessés MAIS1-2	Accident sans tiers versus avec tiers	0,98	2,66
Automobilistes, blessés MAIS3+	Accident sans tiers versus avec tiers	1,17	3,20
Usagers de 2RM, blessés MAIS1-2	Accident sans tiers versus avec tiers	2,36	10,62
Usagers de 2RM, blessés MAIS3+	Accident sans tiers versus avec tiers	2,26	9,57
Cyclistes blessés, MAIS1-2	Accident sans tiers versus avec tiers	4,74	114,00
Cyclistes blessés, MAIS3+	Accident sans tiers versus avec tiers	4,40	81,77
Autres blessés, MAIS1-2	Accident sans tiers versus avec tiers	1,57	4,80
Autres blessés, MAIS3+	Accident sans tiers versus avec tiers	0,61	1,85
Blessés MAIS1-2	Zone Police versus zone gendarmerie	-22,97	NP
Blessés MAIS1-2	Zone CRS versus zone gendarmerie	-23,13	NP
Blessés MAIS3+	Zone Police versus zone gendarmerie	-0,66	0,52
Blessés MAIS3+	Zone CRS versus zone gendarmerie	-0,16	0,85
Piétons blessés, accident avec tiers, zone gendarmerie	MAIS1-2 versus MAIS3+	22,72	NP
Piétons blessés, accident avec tiers, zone Police	MAIS1-2 versus MAIS3+	0,41	1,51
Automobilistes blessés, accident avec tiers, zone gendarmerie	MAIS1-2 versus MAIS3+	22,79	NP
Automobilistes blessés, accident avec tiers, zone Police	MAIS1-2 versus MAIS3+	0,48	1,61
Automobilistes blessés, accident avec tiers, zone CRS	MAIS1-2 versus MAIS3+	-1,18	0,83
Automobilistes blessés, accident sans tiers, zone gendarmerie	MAIS1-2 versus MAIS3+	22,59	NP
Automobilistes blessés, accident sans tiers, zone Police	MAIS1-2 versus MAIS3+	0,29	1,33
Automobilistes blessés, accident sans tiers, zone CRS	MAIS1-2 versus MAIS3+	-0,36	0,70
Motocyclistes blessés, accident avec tiers, zone gendarmerie	MAIS1-2 versus MAIS3+	22,66	NP
Motocyclistes blessés, accident avec tiers, zone Police	MAIS1-2 versus MAIS3+	0,36	1,43
Motocyclistes blessés, accident avec tiers, zone CRS	MAIS1-2 versus MAIS3+	-0,30	0,74
Motocyclistes blessés, accident sans tiers, zone gendarmerie	MAIS1-2 versus MAIS3+	22,97	NP
Motocyclistes blessés, accident sans tiers, zone Police	MAIS1-2 versus MAIS3+	0,66	1,94
Motocyclistes blessés, accident sans tiers, zone CRS	MAIS1-2 versus MAIS3+	0,50	1,64
Cyclistes blessés, accident avec tiers, zone gendarmerie	MAIS1-2 versus MAIS3+	22,31	NP
Cyclistes blessés, accident avec tiers, zone Police	MAIS1-2 versus MAIS3+	0,007	1,01
Cyclistes blessés, accident avec tiers, zone CRS	MAIS1-2 versus MAIS3+	-0,65	0,52
Cyclistes blessés, accident sans tiers, zone gendarmerie	MAIS1-2 versus MAIS3+	22,65	NP
Cyclistes blessés, accident sans tiers, zone Police	MAIS1-2 versus MAIS3+	0,34	1,41
Cyclistes blessés, accident sans tiers, zone CRS	MAIS1-2 versus MAIS3+	-0,32	0,73

En termes de type d'utilisateur, il ressort nettement, pour les accidents sans tiers, que les blessés cyclistes sont bien mieux enregistrés dans le Registre seulement que dans les BAAC seulement, par rapport aux automobilistes, et rappelons le, toutes choses égales par ailleurs (notamment le MAIS prédit, le type de réseau, ...).

Cela reste vrai mais de bien moindre ampleur pour les blessés à deux-roues motorisés.

Pour les accidents avec tiers, il n'y a que les cyclistes qui sont mieux enregistrés dans le Registre seulement que dans BAAC seulement, par rapport aux automobilistes.

A contrario cela indique que les usagers de deux-roues motorisés et les piétons sont assez bien enregistrés dans BAAC seulement, par rapport aux automobilistes. Si l'on regarde cette caractéristique : accident sans tiers ou avec tiers, pour un type d'utilisateur donné et une gravité MAIS donnée, tous les blessés dans un accident sans tiers (versus avec tiers) sont mieux enregistrés dans Registre seulement que dans BAAC seulement. Cela montre un sous-enregistrement des blessés sans tiers coté BAAC.

En ce qui concerne la force de l'ordre, on note un très grand décalage pour les blessés MAIS1-2, entre la gendarmerie et les deux autres forces de l'ordre : les blessés légers en zone gendarmerie sont infiniment mieux enregistrés dans Registre seulement que dans BAAC seulement (les valeurs des bêta sont tellement grandes que l'odds-ratio prend des valeurs extrêmes et devient non-pertinent). Cela indique qu'il y a un sous-enregistrement très marqué des blessés légers dans les BAAC en zone gendarmerie. De plus, ce sous-enregistrement se détériore sur la période étudiée (2006-2012)

Sont présentés ci-après les coefficients de correction du sous-enregistrement des BAAC selon le type d'utilisateur, la présence ou l'absence de tiers, l'hospitalisation selon les BAAC, la gravité MAIS du blessé (scindée en deux catégories : MAIS1,2 vs MAIS3+), en 2012. Sans surprise, pour un même type d'accident, les blessés non hospitalisés ont des coefficients de correction du sous-enregistrement plus élevés que les blessés hospitalisés. Également, les blessés MAIS1,2 ont aussi des coefficients de correction du sous-enregistrement plus élevés que les MAIS3+.

**Tableau 31 : Coefficients de correction du sous-enregistrement des BAAC selon le type d'usager, la présence ou l'absence de tiers, l'hospitalisation selon les BAAC, la gravité MAIS, pour les hommes entre 30 et 59 ans, pour un accident uniquement corporel sur route nationale ou route départementale, renseigné par la police, 2012**

Type d'usager x tiers	Hospitalisation selon BAAC	Gravité MAIS	Coefficients de correction
2RM avec tiers	blessé hospitalisé	MAIS1,2	1,1
2RM avec tiers	blessé hospitalisé	MAIS3+	1,1
2RM avec tiers	Blessé non hospitalisé	MAIS1,2	1,5
2RM avec tiers	Blessé non hospitalisé	MAIS3+	1,4
2RM sans tiers	Blessé hospitalisé	MAIS1,2	2,2
2RM sans tiers	Blessé hospitalisé	MAIS3+	1,5
2RM sans tiers	Blessé non hospitalisé	MAIS1,2	8,8
2RM sans tiers	Blessé non hospitalisé	MAIS3+	4,2
Vélo avec tiers	Blessé hospitalisé	MAIS1,2	1,1
Vélo avec tiers	Blessé hospitalisé	MAIS3+	1,1
Vélo avec tiers	Blessé non hospitalisé	MAIS1,2	1,9
Vélo avec tiers	Blessé non hospitalisé	MAIS3+	1,9
Vélo sans tiers	Blessé hospitalisé	MAIS1,2	16,1
Vélo sans tiers	Blessé hospitalisé	MAIS3+	8,7
Vélo sans tiers	Blessé non hospitalisé	MAIS1,2	97,3
Vélo sans tiers	Blessé non hospitalisé	MAIS3+	50,5
VL avec tiers	Blessé hospitalisé	MAIS1,2	1,1
VL avec tiers	Blessé hospitalisé	MAIS3+	1,1
VL avec tiers	Blessé non hospitalisé	MAIS1,2	1,9
VL avec tiers	Blessé non hospitalisé	MAIS3+	1,6
VL sans tiers	Blessé hospitalisé	MAIS1,2	1,3
VL sans tiers	Blessé hospitalisé	MAIS3+	1,2
VL sans tiers	Blessé non hospitalisé	MAIS1,2	3,1
VL sans tiers	Blessé non hospitalisé	MAIS3+	2,3
Piéton avec tiers	Blessé hospitalisé	MAIS1,2	1,1
Piéton avec tiers	Blessé hospitalisé	MAIS3+	1,1
Piéton avec tiers	Blessé non hospitalisé	MAIS1,2	1,7
Piéton avec tiers	Blessé non hospitalisé	MAIS3+	1,5
Autre avec tiers	Blessé hospitalisé	MAIS1,2	1,0
Autre avec tiers	Blessé hospitalisé	MAIS3+	1,1
Autre avec tiers	Blessé non hospitalisé	MAIS1,2	1,3
Autre avec tiers	Blessé non hospitalisé	MAIS3+	1,6
Autre sans tiers	Blessé hospitalisé	MAIS1,2	1,3
Autre sans tiers	Blessé hospitalisé	MAIS3+	1,3
Autre sans tiers	Blessé non hospitalisé	MAIS1,2	2,6
Autre sans tiers	Blessé non hospitalisé	MAIS3+	2,7



## 2.10 Étape 2C : Projection du département du Rhône à la France entière

Ce qui nous intéresse, c'est l'extrapolation d'indicateurs de santé estimés au niveau départemental ou régional à une zone géographique plus large, régionale ou nationale. Les études publiées sur le sujet sont présentées en annexe 3. Une étude intéressante pour nous, car similaire, est l'estimation du nombre de toxicomanes consommateurs de cocaïne à partir du nombre observé d'héroïnomanes : la relation entre les deux est observée puis modélisée sur Londres, où les deux populations sont connues, puis appliquée à l'ensemble du Royaume-Uni (Gossop et al., 1994). Dans le domaine de la traumatologie routière, une étude de (James, 1991) est très proche de notre situation : le nombre de blessés de la route est donné par une source médicale : 6 hôpitaux impliqués. Ils sont comparés aux données policières, et semble-t-il en stratifiant sur le type d'utilisateur et la gravité : des taux de correction moyens sont alors estimés et ensuite appliqués aux données policières de l'ensemble du Royaume-Uni.

Dans la demande actuelle de la Commission Européenne à ses États membres de fournir le nombre de blessés de la route MAIS3+, certains pays européens estiment ces chiffres en appliquant aussi des facteurs correctifs à leurs données policières nationales.

Ce que l'on extrapole ici (par projection), du département du Rhône au territoire France métropolitaine, ce n'est pas l'accidentalité (observée sur le Rhône), mais les pratiques d'enregistrement des blessés des forces de l'ordre (du Rhône vers la France).

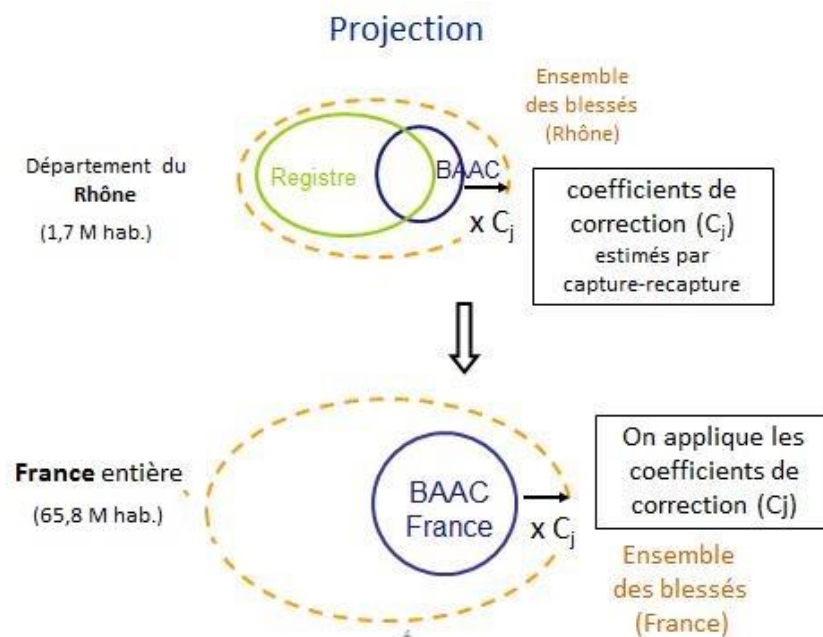


Figure 8: projection, du Rhône à la France entière

La méthode mise en œuvre est analogue à une standardisation indirecte (terminologie épidémiologique) d'un taux d'incidence, de cancer par exemple (Estève et al., 1994), effectuée classiquement sur âge et sexe. Ici il s'agit de la standardisation indirecte, de coefficients de correction de sous-enregistrement (au lieu d'incidences), sur les facteurs de sous-enregistrement (gravité de l'utilisateur, gravité de l'accident, type d'utilisateur, tiers, réseau et force de l'ordre) au lieu des facteurs de confusion (âge et sexe habituellement).

L'estimation nationale du nombre réel de blessés est donnée par :

$$E_F = \sum_{j=1}^J C_j \times O_{F,j}, \text{ où :}$$

$j$  indice les strates définies par la combinaison des facteurs de biais d'enregistrement (gravité, type de force de l'ordre, type d'utilisateur, etc.), à la place des strates d'âge et de sexe (habituellement facteurs de confusion en épidémiologie),

$C_j$  sont les coefficients de correction du sous-enregistrement, de la zone de référence (Rhône) au lieu des taux d'incidences, dans la strate  $j$ ,

$O_{F,j}$  sont les effectifs observés des blessés selon les forces de l'ordre, de la population étudiée (F=France entière), au lieu des personnes-années de la population étudiée.

Par ailleurs, nous pouvons aussi utiliser la terminologie de la théorie des sondages (Ardilly, 1994) ; voir annexes.

Nous écrivons l'étape intermédiaire :

$$E_F = \sum_{j=1}^J E_{F,j} = \sum_{j=1}^J C_j \times O_{F,j},$$

Cela montre qu'il y a d'une part une post-stratification selon les strates  $j$  définies par les facteurs du sous-enregistrement :

$$E_F = \sum_{j=1}^J E_{F,j}.$$

Cela montre, qu'il y a d'autre part, dans chaque strate  $j$ , un redressement par le ratio :

$$E_{F,j} = C_j \times O_{F,j}.$$

Les dénombrements des forces de l'ordre ( $O_{F,j}$ ) sont redressés par un coefficient ( $C_j$ ) afin d'obtenir de meilleurs estimations, et ce coefficient est estimé sur un échantillon, définit ici par un département.

Cette formulation signifie, que pour une catégorie donnée, par exemple les cyclistes hospitalisés dans un accident corporel avec tiers sur une route départementale ou nationale en zone gendarmerie, nous considérons que le taux de sous-enregistrement modélisé sur le département du Rhône est une bonne estimation du taux de sous-enregistrement (dans la même catégorie) sur l'ensemble de la France. L'hypothèse sous-jacente est celle d'une homogénéité des pratiques des forces de l'ordre d'enregistrement des victimes sur l'ensemble de la France, **à l'intérieur des sous-groupes définis par la combinaison des facteurs de sous-enregistrement et notamment pour un type de force de l'ordre donné.**

En particulier, nous assimilons la police aux frontières et la préfecture de police de Paris à la force de l'ordre « police ».

Enfin, les coefficients de correction (du sous-enregistrement et de ses biais) sont estimés sur les données de 2006 à 2012. Ces coefficients incluent un coefficient dépendant de l'année, de façon quantitative et non qualitative (c'est-à-dire qu'il a la même valeur entre deux années consécutives quelles qu'elles soient et non par une valeur spécifique pour chaque année). Ce coefficient lié à l'année n'existe que pour la gendarmerie, où une évolution dans le temps du sous-enregistrement a été observée (une dégradation). Cela n'a pas été observé pour la police ou les CRS.

Ce coefficient correcteur selon l'année, et les autres coefficients correcteurs ont ensuite été appliqués aux données BAAC 2013-2016, au fur et à mesure de leur disponibilité.

### 3 RESULTATS

Tous les résultats concernent les estimations par capture-recapture sur BAAC-Registre (CRC) au niveau national.

#### 3.1 Nombre de blessés toutes catégories

##### 3.1.1 Nombre de blessés toutes gravités (MAIS1+)

Le tableau ci-après donne l'effectif estimé de blessés toutes gravités, corrigé du sous-enregistrement et des biais associés. Il présente aussi, pour comparaison, les effectifs des blessés toutes gravités selon les BAAC.

*Tableau 32 : effectifs annuels des blessés toutes gravités, selon capture-recapture ou selon les BAAC, France*

Année	Blessés MAIS1+ Estimations CRC	Blessés toutes gravités Effectifs BAAC*
2006	360 522	102 204
2007	382 189	103 268
2008	364 415	93 853
2009	361 479	91 002
2010	338 282	84 698
2011	327 114	81 399
2012	307 910	75 932
2013	274 263	70 607
2014	291 693	73 048
2015	284 254	70 818
2016	298 293	72 645

CRC = capture-recapture sur BAAC et Registre

\* données BAAC : de possibles indemnes ont été exclus parmi les blessés classés non graves

Comme attendu, les estimations sont bien plus élevées que les chiffres des forces de l'ordre, d'un facteur multiplicatif de 3,6-4 environ.

Les évolutions des estimations sont les mêmes que celles des BAAC, ce qui est la conséquence logique de la procédure utilisée. En effet, dans le département du Rhône où nous avons pu comparer les évolutions selon les BAAC et selon le Registre, nous avons constaté une stabilité dans le sous-enregistrement pour les CRS et la Police, et seulement une (petite) détérioration du sous-enregistrement pour la Gendarmerie.

Nous avons tenu compte de cela dans la modélisation en introduisant un effet année quantitatif pour la Gendarmerie (coefficient correcteur  $c = 1,04$  entre deux années consécutives, quelles qu'elles soient).

### 3.1.2 Nombre de blessés graves MAIS3+

Les tableaux ci-après donnent, pour chaque année, l'effectif de blessés graves (MAIS3+), estimé par capture-recapture et, pour rappel, l'effectif de blessés hospitalisés d'après les BAAC. Ces derniers sont donnés à titre de comparaison, notamment des évolutions, mais recouvrent des niveaux de gravité et de détermination différents. En effet, on observe (dans le Registre) qu'un blessé hospitalisé n'a qu'une probabilité de 50% d'être MAIS3+, et de surcroît un blessé MAIS3+ n'est pas toujours hospitalisé (5% d'exception) mais l'hospitalisation (oui/non) est le seul critère de gravité qui est présent dans les BAAC.

*Tableau 33 : effectifs annuels des blessés MAIS3 selon capture-recapture, et des blessés hospitalisés selon les BAAC, France*

Année	Blessés MAIS3+ estimations par CRC	Blessés hospitalisés BAAC*
2006	32 037	40 664
2007	32 036	38 619
2008	30 231	34 965
2009	29 509	33 325
2010	27 228	30 418
2011	26 754	29 677
2012	24 542	27 132
2013	23 291	25 966
2014	24 592	26 635
2015	24 273	26 595
2016	25 401	27 187

CRC= capture-recapture sur BAAC et Registre

\* données BAAC : de possibles indemnes ont été exclus parmi les blessés classés non graves

En termes d'ordre de grandeur, les deux effectifs sont proches (alors qu'ils ne devraient pas l'être). Étant donné que dans le Registre du Rhône on observe que, parmi les hospitalisés, 45% sont MAIS3+, cela donnerait en partant de l'estimation des MAIS3+, un nombre d'hospitalisés de 54 000 en 2015 (et non 27000 comme l'indiquent les BAAC).

En terme d'évolution, le nombre de blessés MAIS3+ estimé baisse moins vite que le nombre de blessés hospitalisés selon les BAAC.

### 3.1.3 Proportion de blessés MAIS3+ parmi les blessés MAIS1+

La gravité globale des blessés peut être mesurée par la proportion de blessés MAIS3+ parmi les blessés MAIS1+, Elle est stable, autour de 8,2%.

Tableau 34 : Effectifs de blessés MAIS1+, MAIS3+, selon capture-recapture, et leur ratio, France.

Année	Blessés MAIS3+	Blessés MAIS1+	Proportion de MAIS3+
2006	32 037	360 522	8,9%
2007	32 036	382 189	8,4%
2008	30 231	364 415	8,3%
2009	29 509	361 479	8,2%
2010	27 228	338 282	8,0%
2011	26 754	327 114	8,2%
2012	24 542	307 910	8,0%
2013	23 291	274 263	8,5%
2014	24 592	291 693	8,4%
2015	24 273	284 254	8,5%
2016	25 401	298 293	8,5%

Elle est légèrement plus élevée sur la France entière que sur le Rhône (7,3%), ce qui est attendu car la France est globalement plus rurale que le Rhône et les accidents en rural sont en moyenne plus graves qu'en urbain, en raison de vitesses pratiquées plus élevées.

### 3.1.4 Ratio Nombre de blessés MAIS3+/Nombre de tués

Le nombre de tués par accident de la circulation étant considéré comme très bien estimé, le ratio du nombre de blessés MAIS3+ par le nombre de tués est intéressant à observer.

Tableau 35 : effectifs de blessés MAIS3+ estimés selon capture-recapture, effectifs de tués (BAAC) et leur ratio, France

Année	MAIS3+	tués	Ratio MAIS3+/tués
2006	32 037	4 708	6,8
2007	32 036	4 620	6,9
2008	30 231	4 275	7,1
2009	29 509	4 271	6,9
2010	27 228	3 992	6,8
2011	26 754	3 963	6,8
2012	24 542	3 652	6,7
2013	23 291	3 268	7,1
2014	24 592	3 384	7,3
2015	24 273	3 461	7,0
2016	25 401	3 477	7,3

Ainsi au cours des onze années d'observation, ce ratio MAIS3+/tués reste constant, il est d'environ de 7 blessés graves pour 1 tué.

## 3.2 Blessés selon le type d'utilisateur

### 3.2.1 Nombre de blessés toutes gravités (MAIS1+)

Les tableaux ci-après donnent les effectifs corrigés par la méthode capture-recapture des blessés de la route MAIS1+ selon le type d'utilisateur ainsi que, les blessés toutes gravités selon les BAAC.

*Tableau 36 : effectifs annuels des blessés toutes gravités, estimés par capture-recapture selon le type d'utilisateur, France*

Année	2RM	Vélo	VL	Piéton	Autre*
2006	115 518	61 536	144 983	27 853	10 632
2007	124 623	73 673	145 662	28 288	9 942
2008	120 460	77 327	129 829	27 212	9 588
2009	119 377	75 510	130 320	26 788	9 483
2010	107 198	72 878	122 839	26 125	9 242
2011	103 203	75 576	115 229	25 406	7 700
2012	95 564	68 301	110 686	24 563	8 795
2013	84 711	54 735	103 214	23 740	7 863
2014	87 511	61 840	110 581	24 393	7 368
2015	81 423	61 710	109 956	23 236	7 929
2016	80 895	62 793	123 210	23 220	8 175

\*Pour rappel ; Autre : VU (véhicule utilitaire), TC/Train (TC : Transport en commun), PL (Poids lourd)

*Tableau 37 : effectifs annuels des BAAC\* des blessés toutes gravités, selon le type d'utilisateur, France*

Année	2RM	Vélo	VL	Piéton	Autre
2006	31 600	4 517	47 799	13 359	4 929
2007	33 651	4 638	46 953	13 408	4 618
2008	31 518	4 406	40 835	12 764	4 330
2009	30 081	4 380	40 006	12 503	4 032
2010	26 968	3 984	37 774	12 111	3 861
2011	26 112	4 312	35 539	11 916	3 520
2012	23 403	3 913	33 734	11 248	3 634
2013	20 857	3 784	31 391	11 005	3 570
2014	21 530	4 098	32 927	11 221	3 272
2015	20 370	4 029	32 406	10 753	3 260
2016	19 908	3 993	34 845	10 617	3 282

\* données BAAC : de possibles indemnes ont été exclus parmi les blessés classés non graves

Estimés par capture-recapture sur la période de stabilité de 2013-2016, les blessés toutes gravités à deux-roues motorisés sont quatre fois plus nombreux que les chiffres des BAAC, 15 fois plus nombreux pour les blessés cyclistes, 3,4 fois plus pour les blessés automobilistes, 2,2 pour les piétons et 2,3 pour les usagers autres.

Les estimations par type d'utilisateur donnent un classement différent selon les sources ou méthodes. Selon les BAAC, les blessés toutes gravités concernent d'abord les automobilistes, puis les usagers de deux-roues motorisés, puis les piétons, et enfin les cyclistes, alors que l'estimation par capture-

recapture indique aussi que le nombre de blessés automobilistes est plus élevé que celui des deux-roues motorisés, mais que le nombre de blessés cyclistes toutes gravités est bien au-dessus de celui des piétons (presque trois fois plus élevé).

Les effectifs des blessés MAIS1+ extrapolés et les effectifs des BAAC sont présentés ci-dessous graphiquement, pour chaque type d'usager.

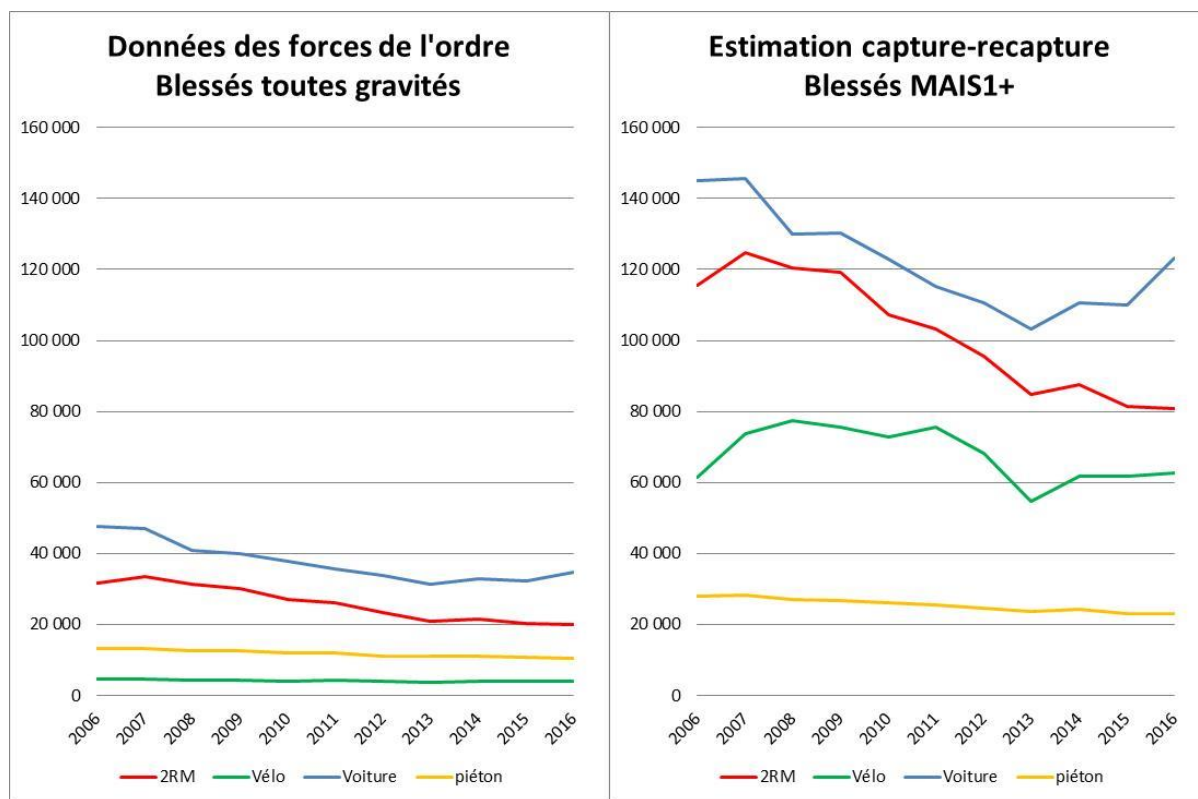


Figure 9: Nombre de blessés toutes gravités selon le type d'usager, 2006-2016

Graphiquement, on observe bien sûr encore plus l'ampleur du sous-enregistrement et les différences selon le type d'usager. Les évolutions sont similaires. Comme dit précédemment, le modèle de capture-recapture retenu colle aux évolutions des BAAC, sauf pour la zone Gendarmerie où un facteur multiplicatif supplémentaire de 1,04 est utilisé pour modéliser l'aggravation dans le temps du sous-enregistrement.

### 3.2.2 Nombre de blessés graves (MAIS3+)

Les tableaux suivants donnent les effectifs corrigés par la méthode capture-recapture des blessés de la route MAIS3+ selon le type d'utilisateur ainsi que les effectifs selon le type d'utilisateur des blessés hospitalisés d'après les BAAC.

*Tableau 38 : effectifs annuels des blessés MAIS3+ estimés par capture-recapture, selon le type d'utilisateur, France*

Année	2RM	Vélo	VL	Piéton	Autre
2006	13 371	3 503	9 615	4 267	1 281
2007	14 165	3 647	8 972	4 205	1 046
2008	13 606	3 727	7 855	4 012	1 030
2009	13 295	3 679	7 686	3 876	973
2010	11 778	3 613	7 166	3 736	935
2011	11 602	3 636	7 004	3 722	789
2012	10 515	3 162	6 487	3 483	894
2013	9 739	3 039	6 160	3 462	890
2014	9 966	3 802	6 459	3 556	810
2015	9 845	3 534	6 533	3 551	810
2016	9 738	4 216	7 127	3 520	800

*Tableau 39 : effectifs annuels des BAAC\* des blessés hospitalisés, selon le type d'utilisateur, France*

année	2RM	Vélo	VL	Piéton	Autre
2006	12 914	1 746	18 374	5 523	2 107
2007	13 250	1 614	16 780	5 312	1 663
2008	12 400	1 559	14 362	5 039	1 605
2009	11 731	1 554	13 799	4 789	1 452
2010	10 360	1 367	12 701	4 593	1 397
2011	10 125	1 424	12 340	4 593	1 195
2012	8 854	1 351	11 420	4 212	1 295
2013	8 174	1 344	10 908	4 199	1 341
2014	8 299	1 446	11 348	4 323	1 219
2015	8 310	1 516	11 244	4 331	1 194
2016	8 230	1 455	12 081	4 289	1 132

\* données BAAC : de possibles indemnes ont été exclus parmi les blessés classés non graves

Les effectifs des blessés MAIS3+ extrapolés et les effectifs des hospitalisés selon les BAAC sont également présentés graphiquement, pour chaque type d'utilisateur.



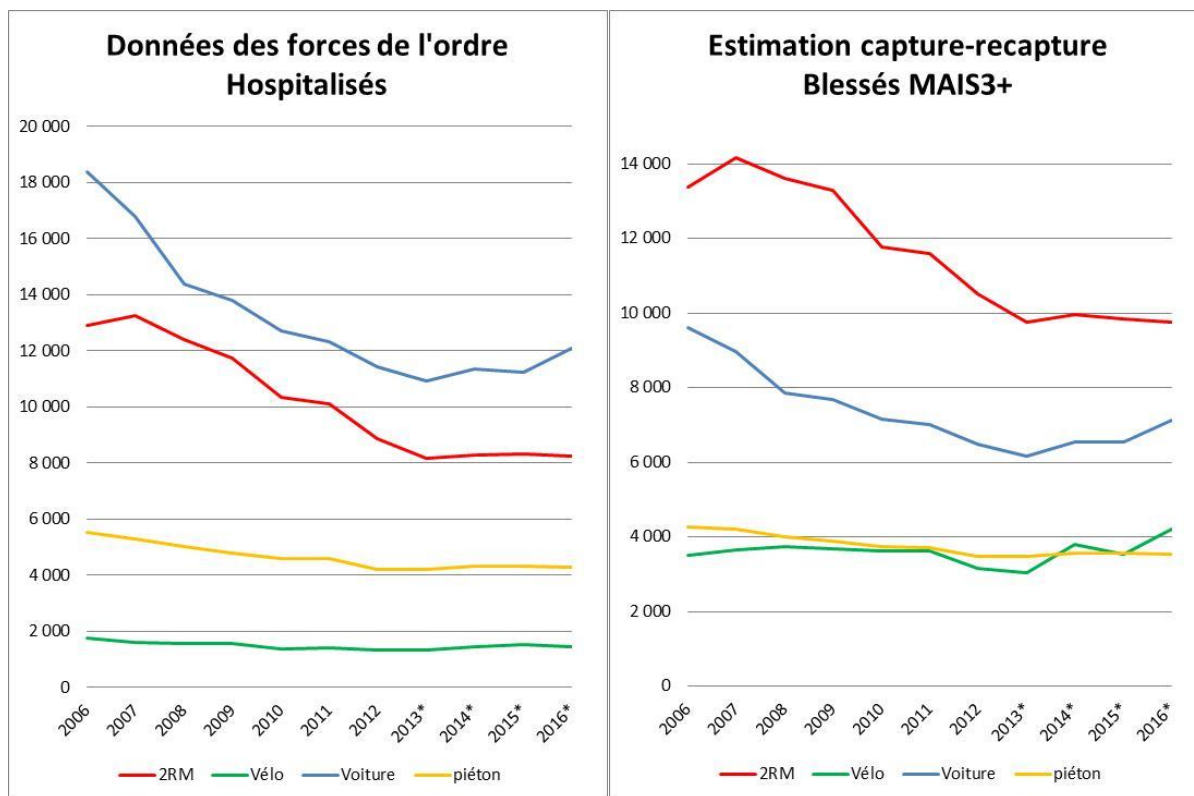


Figure 10: Nombre de blessés hospitalisés BAAC et de blessés graves MAIS3+ selon le type d'usager, 2006-2016

Le critère MAIS3+ corrigé du sous-enregistrement ne donne pas le même classement en termes d'effectifs entre les types d'usagers que le critère hospitalisation des BAAC. En effet, les chiffres d'hospitalisés des BAAC donnent plus de blessés hospitalisés automobilistes que de blessés hospitalisés usagers de deux roues motorisés alors qu'à l'inverse, selon le critère MAIS3+ corrigé, les usagers de deux roues motorisés sont plus nombreux à être blessé MAIS3+ que les automobilistes. Cela est vrai depuis 2003 (Amoros, 2007).

Pour les piétons et les cyclistes, les BAAC enregistrent de 4 000 à 5 000 blessés hospitalisés chez les piétons et 1 500 à 1 700 chez les cyclistes. Le critère MAIS3+ et surtout la correction du sous-enregistrement et de ses biais indiquent en termes d'effectifs qu'il y a quasiment autant de blessés graves piétons que de blessés graves cyclistes (environ 3 500 à 4 000/an).

En termes de kilomètres parcourus, les usagers de deux-roues motorisés représentent 2% du trafic alors que les automobilistes en représentent 70%. Rappelons que le nombre de tués est de 800/an pour les usagers de deux-roues motorisés et de 1 800/an pour les automobilistes.

En termes de nombre de trajets, les piétons sont à 23% de part modale et les cyclistes à 2-3%. Rappelons que les piétons tués sont au nombre de 500/an et les cyclistes au nombre de 150/an (sur 2013-2015).

Pour les cyclistes, les effectifs observés dans les BAAC étant très faibles (cf. Tableau 17 pour le Rhône), les estimations des coefficients de correction sont très peu précises (grande variance). En conséquence les chiffres des cyclistes sont à interpréter avec prudence (dans tout le rapport).

Les tableaux ci-après donnent les effectifs corrigés par la méthode capture-recapture des blessés de la route MAIS3+ selon le type d'usager combiné à la présence ou l'absence de tiers, ainsi que les effectifs des blessés hospitalisés d'après les BAAC.

**Tableau 40 : effectifs annuels des blessés MAIS3+ estimés capture-recapture, selon le type d'usager et l'absence/présence de tiers, France**

Année	2RM avec tiers	2RM sans tiers	Vélo avec tiers	Vélo sans tiers	VL avec tiers	VL sans tiers	Piéton (avec tiers)	Autre avec tiers	Autre sans tiers
2006	7 211	6 160	1 250	2 253	4 846	4 770	4 287	525	756
2007	7 390	6 775	1 184	2 463	4 573	4 399	4 205	416	631
2008	6 885	6 721	1 141	2 587	3 947	3 908	4 012	396	634
2009	6 498	6 767	1 165	2 514	3 743	3 943	3 876	372	601
2010	5 847	5 931	1 031	2 582	3 526	3 640	3 736	348	586
2011	5 653	5 949	1 104	2 532	3 354	3 650	3 722	312	477
2012	5 015	5 500	1 058	2 104	3 231	3 257	3 483	322	572
2013	4 456	5 183	1 056	1 983	3 067	3 094	3 462	342	548
2014	4 710	5 256	1 110	2 692	3 262	3 197	3 556	340	469
2015	4 627	5 218	1 188	2 345	3 193	3 340	3 551	324	486
2016	4 261	5 117	1 159	3 057	3 574	3 552	3 520	308	492

Pour les cyclistes, on note que les accidents sans tiers sont nettement plus nombreux que les accidents avec tiers.

**Tableau 41 : effectifs annuels des blessés hospitalisés BAAC\*, selon le type d'usager et l'absence/présence de tiers, France**

Année	2RM avec tiers	2RM sans tiers	Vélo avec tiers	Vélo sans tiers	VL avec tiers	VL sans tiers	Piéton (avec tiers)	Autre avec tiers	Autre sans tiers
2006	10 276	2 638	1 615	131	11 667	6 707	5 523	1 289	818
2007	10 330	2 920	1 491	123	10 714	6 066	5 312	988	675
2008	9 586	2 814	1 424	135	9 154	5 208	5 039	933	672
2009	8 953	2 778	1 431	123	8 581	5 218	4 789	857	595
2010	8 008	2 352	1 239	128	7 963	4 738	4 593	815	582
2011	7 732	2 393	1 296	128	7 608	4 732	4 593	722	473
2012	6 757	2 097	1 256	95	7 221	4 199	4 212	747	548
2013	6 149	2 025	1 251	93	6 930	3 978	4 199	799	542
2014	6 278	2 021	1 305	141	7 298	4 050	4 323	778	441
2015	6 199	2 111	1 396	120	7 158	4 086	4 331	744	450
2016	6 239	1 991	1 309	146	7 864	4 217	4 289	692	440

\* données BAAC : de possibles indemnes ont été exclus parmi les blessés classés non graves

On remarque que la répartition des blessés peut être très différente, suivant la présence d'un tiers ou non, et suivant le type d'usager. Pour rappel, dans les BAAC, les accidents sans tiers sont moins bien enregistrés que les accidents avec tiers et l'ampleur du sous-enregistrement varie selon le type d'usager. Les facteurs de correction tiennent compte de ces biais d'enregistrement.

La comparaison des deux tableaux confirme que le sous-enregistrement des blessés en absence de tiers impliqué est important y compris pour les blessés graves. Les facteurs de correction utilisés tiennent compte de ces biais d'enregistrement.

La comparaison des rapports des effectifs de victimes d'accidents avec et sans tiers est également intéressante à observer par type d'usager en termes d'évolution entre 2006 et 2016. Ainsi le rapport MAIS3+ avec tiers/MAIS3+ sans tiers est stable pour les occupants de voiture, alors qu'il a tendance à baisser pour les usagers de deux-roues motorisés et dans une moindre mesure pour les cyclistes. Cela pourrait être la conséquence d'une baisse des vitesses pratiquées des automobilistes plus marquée

que celle des usagers de deux-roues motorisés, occasionnant moins d'accidents VL-deux-roues (la majorité des accidents avec tiers), alors que les nombres d'accidents sans tiers des deux-roues motorisés diminuent moins dans la période (ce sont surtout des pertes de contrôle).

### 3.2.3 Proportion de blessés MAIS3+ pour les blessés MAIS1+

Les estimations du nombre de blessés MAIS1+ et MAIS3+ corrigées des biais de sous-enregistrement permettent de calculer les proportions correctes, c'est-à-dire non distordues, de blessés graves (parmi les blessés toutes gravités) par mode de transport.

**Tableau 42 : Effectifs annuels moyens estimés et ratio MAIS3+/MAIS1+ selon le type d'usager, 2006-2012, France corrigé et Registre du Rhône**

Type d'usager	MAIS3+	MAIS1+	Ratio (%) MAIS3+/MAIS1+ (CRC France)	Ratio (%) MAIS3+/MAIS1+ (Registre du Rhône)
2RM	12 619	112 278	11,2	9,8
Vélo	3 567	72 115	4,9*	6,7
VL	7 826	128 507	6,1	3,3
Piéton	3 900	26 605	14,7	10,7
Autre	993	9 340	10,6	7,7

CRC = capture-recapture sur BAAC et Registre

\*Sous enregistrement des blessés graves sans doute insuffisamment corrigé chez les cyclistes.

On obtient sur les chiffres nationaux estimés, l'ordre suivant de proportion croissante de MAIS3+ (parmi les MAIS1+) : vélo, VL, 2RM, piéton. Il est surprenant que la proportion de blessés graves soit plus faible chez les cyclistes que chez les automobilistes. Cela peut provenir d'un redressement moins bon pour les blessés cyclistes que pour les autres modes de transport, à cause de leur sous-enregistrement très marqué dans les BAAC, notamment pour les accidents sans tiers (pour rappel, Tableau 17, le Registre compte 855 blessés cyclistes sans tiers/an et les BAAC du Rhône 6/an).

Le tableau suivant distingue les proportions de blessés MAIS3+ (parmi les blessés MAIS1+) selon qu'un tiers était impliqué ou non dans l'accident.

**Tableau 43 : Ratio MAIS3+/MAIS1+ selon le type d'usager et l'absence/présence de tiers, 2006-2012, France corrigé et Registre Rhône**

Type d'usager	Tiers	MAIS3+	MAIS1+	Ratio (%) MAIS3+/MAIS1+ (CRC France)	Ratio (%) MAIS3+/MAIS1+ (Registre du Rhône)
2RM	Oui	6 357	45 059	14,1	13,3
	Non	6 262	67 218	9,3	7,5
Vélo	Oui	1 133	10 716	10,6	8,0
	Non	2 434	61 398	4,0*	6,4
VL	Oui	3 888	85 344	4,6	2,8
	Non	3 938	43 163	9,1	5,1
Piéton	Oui	3 900	26 605	14,7	13,0
Autre	Oui	385	4 700	8,2	4,9
	Non	608	4 640	13,1	7,3

CRC = capture-recapture sur BAAC et Registre

\*Sous enregistrement des blessés graves sans doute insuffisamment corrigé chez les cyclistes.

Le ratio MAIS3+/MAIS1+ est plus élevé sur la France (estimation) que sur le Rhône (Registre) pour toutes les situations sauf une (cyclistes dans un accident sans tiers). Cela est cohérent avec le fait que la France est moins urbanisée que le département du Rhône.

Pour la situation des cyclistes blessés dans un accident sans tiers, seule situation où la proportion de blessés graves est plus faible sur la France que sur le Rhône, cela peut venir, comme évoqué ci-dessus, de leur très faible effectif dans les BAAC (6/an versus 855/an dans le Registre) et donc d'un redressement fragile ; quelle particularité ont ces 6 accidentés pour être enregistrés dans les BAAC comparés à tous ceux qui manquent ?.

Cela a conduit à un coefficient correcteur très élevé et qui doit être pris avec des pincettes. Cela a aussi conduit à ne pas pouvoir, dans le modèle de prédiction du MAIS3+ (parmi les blessés BAAC), dissocier chez les cyclistes les catégories sans tiers et avec tiers en interaction avec blessés hospitalisés/non-hospitalisés, alors qu'elles l'ont été pour les autres usagers (cf. Tableau 10). Cette dissociation (avec/sans tiers) en interaction avec hospitalisé (oui/non) permet de prendre en compte la supposition, que, pour que des blessés sans tiers soient enregistrés dans les BAAC, il faut qu'ils soient d'une certaine gravité, même parmi ceux classés non-hospitalisés. Cela a pu être pris en compte pour tous les usagers, notamment les deux-roues motorisé, mais cela n'a pas pu être fait pour les cyclistes, alors qu'ils ont un « pattern » d'enregistrement proches des deux-roues motorisés, et en plus marqué.

Une fois qu'on a dit cela, on peut tout de même comparer les modes de transports entre eux.

Pour les accidents sans tiers, les blessés à deux-roues motorisés et les blessés automobilistes connaissent la même proportion de blessés graves (autour de 9%). Elle est plus élevée chez les blessés autres (VU, PL, bus, ...).

Pour les accidents avec tiers (qui est dans la grande majorité des cas un usager motorisé), la proportion de blessés graves est la plus faible chez les automobilistes (4,6%) puis chez les usagers « autres », puis chez les cyclistes (10,6%), puis chez les motocyclistes et piétons (respectivement 14,1% et 14,7%).

### 3.3 Blessés selon le sexe

#### 3.3.1 Nombre de blessés toutes gravités (MAIS1+)

Les tableaux ci-après donnent les estimations par sexe. Les femmes sont environ deux fois moins souvent blessées dans un accident de la route que les hommes alors qu'ils et elles se déplacent autant (ENDT, 2007-2008). De plus, entre 2006 et 2015, la baisse du nombre de blessés toutes gravités est deux fois plus forte chez les femmes que chez les hommes.

*Tableau 44 : Effectifs annuels des blessés MAIS1+ estimés selon capture-recapture selon le sexe, 2006-2016, France*

Année	Masculin	Féminin
2006	233 632	126 890
2007	250 067	132 122
2008	229 422	134 993
2009	235 354	126 125
2010	218 221	120 061
2011	214 318	112 796
2012	199 764	108 147
2013	180 490	93 773
2014	190 059	101 633
2015	197 723	86 530
2016	202 158	96 135
sex-ratio 2016	2,1	

Le tableau ci-dessous donne la répartition hommes-femmes des blessés par mode de transport.

*Tableau 45 : Effectifs annuels des blessés MAIS1+ estimés selon capture-recapture selon le type d'usager et le sexe, 2006-2016, France*

Année	2RM		Vélo*		VL		Piéton		Autre	
	M	F	M	F	M	F	M	F	M	F
2006	92 995	22 522	43 991	17 545	75 990	68 993	12 931	14 923	7 725	2 907
2007	99 476	25 148	55 281	18 392	75 109	70 553	13 012	15 276	7 189	2 753
2008	95 106	25 354	48 209	29 118	66 780	63 049	12 434	14 778	6 894	2 694
2009	93 547	25 830	56 025	19 485	66 697	63 623	12 152	14 636	6 932	2 551
2010	84 236	22 962	52 686	20 192	62 939	59 900	11 902	14 223	6 457	2 785
2011	82 807	20 396	53 867	21 709	60 387	54 842	11 582	13 823	5 675	2 025
2012	76 570	18 995	48 398	19 904	57 240	53 446	11 204	13 358	6 352	2 444
2013	68 117	16 594	41 550	13 185	53 921	49 293	10 917	12 823	5 985	1 878
2014	70 844	16 667	44 688	17 152	57 803	52 777	11 216	13 176	5 507	1 861
2015	68 422	13 001	49 106	12 604	62 247	47 709	11 887	11 349	6 061	1 868
2016	67 376	13 519	47 736	15 058	68 817	54 394	12 014	11 205	6 215	1 959
Sex-ratio 2016	5,0		3,2		1,3		1,1		3,2	

Les sex-ratios sont très variables entre les différents usagers. Cela vient à la fois des choix différents de modes (moins de femmes que d'hommes à vélo, et encore bien moins à deux-roues motorisés ou véhicules autres (VU, PL ...)) et de prises de risque différentes.

### 3.3.2 Nombre de blessés graves (MAIS3+)

Les tableaux ci-dessous donnent les effectifs estimés de blessés MAIS3+.

L'écart entre hommes et femmes est plus marqué que sur les blessés toutes gravités. Cela est sans doute en partie la conséquence d'une conduite à vitesse plus élevée pour les hommes, avec des véhicules plus rapides, avec un usage différent en moyenne de leur deux-roues motorisé (usage plus "ludique").

**Tableau 46 : Effectifs annuels des blessés MAIS3+ estimés selon capture-recapture selon le sexe, 2006-2016, France**

Année	Masculin	Féminin
2006	23 791	8 246
2007	24 134	7 902
2008	22 590	7 641
2009	22 115	7 395
2010	20 417	6 811
2011	20 104	6 650
2012	18 401	6 141
2013	17 385	5 906
2014	18 464	6 128
2015	18 634	5 640
2016	19 469	5 932
Sex-ratio 2016	3,3	

**Tableau 47 : Effectifs annuels des blessés MAIS3+ estimés selon capture-recapture selon le type d'usager et le sexe, 2006-2016, France**

Année	2RM		Vélo*		VL		Piéton		Autre	
	M	F	M	F	M	F	M	F	M	F
2006	11 673	1 698	2 766	737	6 114	3 501	2 203	2 065	1 036	245
2007	12 335	1 830	3 086	561	5 680	3 292	2 182	2 022	850	196
2008	11 846	1 760	2 813	914	5 009	2 845	2 097	1 915	824	207
2009	11 542	1 753	2 948	732	4 827	2 859	2 010	1 866	789	184
2010	10 215	1 563	3 019	595	4 498	2 668	1 943	1 794	743	191
2011	10 141	1 462	2 902	734	4 497	2 507	1 909	1 814	656	134
2012	9 194	1 322	2 611	551	4 087	2 400	1 782	1 701	727	167
2013	8 530	1 209	2 483	557	3 836	2 324	1 804	1 659	803	158
2014	8 746	1 220	3 201	600	4 044	2 415	1 811	1 745	662	148
2015	8 792	1 053	2 905	629	4 253	2 281	2 007	1 545	677	133
2016	8 693	1 046	3 505	711	4 589	2 538	2 020	1 500	662	138
Sex-ratio 2016	8,3		4,9		1,8		1,3		4,8	

\*Sous enregistrement des blessés graves sans doute insuffisamment corrigé chez les cyclistes.

En détaillant par type d'usager, on observe que le sex-ratio est en défaveur des hommes pour toutes les catégories d'usagers, y compris pour les piétons, mais dans une proportion beaucoup plus faible que pour les deux-roues, un mode de transport surtout pratiqué par les hommes.

### 3.3.3 Proportion de blessés MAIS3+ parmi les blessés MAIS1+

Avec la proportion de blessés MAIS3+ parmi les MAIS1+ ci-dessous, on voit directement la gravité plus élevée chez les hommes que chez les femmes, quel que soit le mode de transport.

*Tableau 48 : Effectifs annuels moyens estimés par capture-recapture des blessés MAIS3+, des blessés MAIS1+ et ratio MAIS3+/MAIS1+ selon le type d'usager et le sexe, 2006-2012, France*

Type d'usager	Sexe	MAIS3+	MAIS1+	Ratio (%) MAIS3+/MAIS1+ (CRC France)	Ratio (%) MAIS3+/MAIS1+ (Registre du Rhône)
2RM	F	1 627	23 030	7,1	5,5
	M	10 992	89 248	12,3	10,7
Vélo	F	689	20 906	3,3*	5,8
	M	2 878	51 208	5,6*	6,9
VL	F	2 868	62 058	4,6	2,5
	M	4 959	66 449	7,5	4,3
Piéton	F	1 882	14 431	13,0	10,3
	M	2 018	12 174	16,6	11,0
Autre	F	189	2 594	7,3	5,3
	M	803	6 746	11,9	8,7

CRC = capture-recapture sur BAAC et Registre

\*Sous enregistrement des blessés graves sans doute insuffisamment corrigé chez les cyclistes.

### 3.4 Blessés selon l'âge

#### 3.4.1 Nombre de blessés toutes gravités (MAIS1+)

Les tableaux ci-après donnent les effectifs estimés des blessés MAIS1+ selon l'âge.

*Tableau 49 : Effectifs annuels des blessés MAIS1+ estimés selon capture-recapture, selon l'âge, 2006-2016, France*

Année	00-09 ans	10-14 ans	15-19 ans	20-29 ans	30-54 ans	55-64 ans	65-74 ans	75+ ans
2006	12 757	15 762	56 131	99 406	128 827	24 602	11 507	11 531
2007	11 475	18 164	62 050	104 035	136 705	26 114	13 285	10 362
2008	10 575	18 139	57 090	98 237	129 513	27 498	12 161	11 204
2009	11 611	18 546	55 024	96 193	132 700	25 387	11 283	10 734
2010	11 913	15 158	53 940	85 817	121 331	26 303	13 029	10 793
2011	9 443	15 904	44 677	87 427	120 180	27 510	11 675	10 299
2012	9 723	11 532	43 534	79 930	118 827	22 089	11 643	10 633
2013	8 248	10 507	35 896	68 134	106 085	24 815	11 671	8 908
2014	10 880	11 626	35 387	75 136	110 487	23 443	14 061	10 672
2015	8 045	12 938	41 213	71 108	104 192	23 696	13 125	9 936
2016	8 879	11 422	39 066	74 232	106 244	30 262	15 091	13 095

Les effectifs de blessés toutes gravités baissent dans toutes les tranches d'âge sauf chez les plus âgés ; cela augmente chez les 65-74 ans et les 75 ans et plus. Cela provient sans doute du vieillissement de la population (ces classes d'âge augmentent en effectif dans la population générale).

#### 3.4.2 Nombre de blessés graves (MAIS3+)

*Tableau 50 : Effectifs annuels des blessés MAIS3+ estimés selon capture-recapture selon l'âge, 2006-2016, France*

Année	00-09 ans	10-14 ans	15-19 ans	20-29 ans	30-54 ans	55-64 ans	65-74 ans	75+ ans
2006	854	1 358	5 399	7 652	10 663	2 565	1 446	2 100
2007	766	1 321	5 368	7 775	10 687	2 732	1 458	1 929
2008	744	1 280	4 876	7 245	10 068	2 732	1 352	1 933
2009	731	1 261	4 494	6 936	10 112	2 698	1 292	1 985
2010	651	1 083	4 043	6 235	9 311	2 630	1 343	1 932
2011	584	987	3 829	6 243	9 298	2 659	1 263	1 891
2012	572	866	3 280	5 685	8 492	2 528	1 352	1 766
2013	551	817	2 920	5 091	8 342	2 636	1 270	1 664
2014	617	950	2 914	5 273	8 615	2 768	1 566	1 889
2015	547	848	3 067	5 201	8 240	2 907	1 649	1 815
2016	531	835	2 970	5 382	8 591	3 200	1 803	2 089

De la même façon, l'effectif des blessés MAIS3+ diminue dans toutes les classes d'âge sauf chez les 55-64 ans et les 65-74 ans (où il augmente). Il est stable chez les 75 ans et plus.



### 3.4.3 Proportion de blessés MAIS3+ parmi les blessés MAIS1+

Le tableau ci-après donne le pourcentage de blessés graves selon l'âge et le sexe.

*Tableau 51 : effectifs annuels moyens estimés par capture-recapture des blessés MAIS3+, des blessés MAIS1+ et ratio MAIS3+/MAIS1+ selon l'âge et selon le sexe, 2006-2012, France*

	MAIS3+	MAIS1+	Ratio (%) MAIS3+/MAIS1+ (CRC France)	Ratio (%) MAIS3+/MAIS1+ (Registre du Rhône)
<b>Age (en années)</b>				
00-09	700	11 071	6,3	4,4
10-14	1 165	16 172	7,2	6,8
15-19	4 470	53 206	8,4	6,4
20-29	6 825	93 006	7,3	5,3
30-54	9 804	126 869	7,7	6,7
55-64	2 649	25 643	10,3	10,7
65-74	1 358	12 083	11,2	12,5
75+	1 934	10 794	17,9	17,9
<b>Sexe</b>				
Féminin	7 255	123 019	5,9	4,5
Masculin	21 650	225 825	9,6	8,2

CRC = capture-recapture sur BAAC et Registre

Concernant l'âge, on remarque que les profils d'accident où le ratio MAIS3+/MAIS1+ (France) est plus élevé sont les accidents de personnes âgées ; en particulier le ratio augmente dès 55 ans et fortement à partir de 75 ans. Par ailleurs les hommes ont un ratio bien plus élevé que les femmes.

Le tableau ci-après donne la proportion de blessés graves selon le type d'utilisateur et l'âge.

**Tableau 52 : Effectifs annuels moyens des blessés MAIS3+ estimés par capture-recapture, des blessés MAIS1+ et ratio MAIS3+/MAIS1+ selon le type d'utilisateur et l'âge, 2006-2012, France**

Type d'utilisateur	Age	MAIS3+	MAIS1+	Ratio (%) MAIS3+/MAIS1+ (CRC France)
2RM	00-09	26	392	6,6
	10-14	338	2 822	12,0
	15-19	2 880	26 990	10,7
	20-29	3 659	34 841	10,5
	30-54	4 899	42 197	11,6
	55-64	645	4 084	15,8
	65-74	116	719	16,1
	75+	56	232	24,0
Vélo	00-09	91	2 060	4,4
	10-14	339	7 411	4,6
	15-19	351	8 650	4,1
	20-29	428	13 993	3,1
	30-54	1 170	27 089	4,3
	55-64	630	7 844	8,0
	65-74	309	3 199	9,7
	75 +	249	1 867	13,4
VL	00-09	195	5 085	3,8
	10-14	130	2 787	4,6
	15-19	877	14 156	6,2
	20-29	2 160	38 561	5,6
	30-54	2 480	46 679	5,3
	55-64	790	10 295	7,7
	65-74	493	5 697	8,7
	75 +	701	5 248	13,4
Piéton	00-09	374	3 306	11,3
	10-14	331	2 771	11,9
	15-19	287	2 608	11,0
	20-29	350	3 308	10,6
	30-54	791	6 507	12,1
	55-64	469	2 664	17,6
	65-74	405	2 209	18,3
	75 +	893	3 233	27,6
Autre	00-09	14	229	6,0
	10-14	28	380	7,4
	15-19	75	802	9,3
	20-29	228	2 303	9,9
	30-54	464	4 396	10,6
	55-64	114	757	15,1
	65-74	36	259	13,9
	75+	34	213	16,0

CRC = capture-recapture sur BAAC et Registre

L'augmentation de la gravité selon l'âge s'observe chez tous les types d'utilisateurs.

Tableau 53 : Effectifs annuels moyens des blessés MAIS3+ estimés par capture-recapture, des blessés MAIS1+ et ratio MAIS3+/MAIS1+ selon le type d'usager, l'âge et le sexe, 2006-2012, France

Type d'usager	Sexe	Age	MAIS3+	MAIS1+	Ratio (%) MAIS3+/MAIS1+ (CRC France)
2RM	F	00-09	8	148	5,3
		10-14	73	909	8,0
		15-19	453	6 693	6,8
		20-29	462	7 318	6,3
		30-54	549	7 338	7,5
		55-64	68	515	13,1
		65-74	10	76	12,8
		75+	5	33	14,7
	H	00-09	18	244	7,3
		10-14	265	1 914	13,8
		15-19	2 427	20 297	12,0
		20-29	3 198	27 524	11,6
		30-54	4 350	34 859	12,5
		55-64	578	3 569	16,2
65-74		106	643	16,5	
75+		51	199	25,5	
Vélo	F	00-09	19	697	2,7
		10-14	49	1 305	3,8
		15-19	62	1 872	3,3
		20-29	107	4 853	2,2
		30-54	226	8 202	2,8
		55-64	139	2 689	5,2
		65-74	47	909	5,1
		75+	41	381	10,6
	H	00-09	73	1 363	5,3
		10-14	289	6 107	4,7
		15-19	289	6 778	4,3
		20-29	321	9 141	3,5
		30-54	945	18 887	5,0
		55-64	491	5 155	9,5
65-74		262	2 291	11,5	
75+		209	1 487	14,0	
VL	F	00-09	79	2 470	3,2
		10-14	57	1 540	3,7
		15-19	266	5 993	4,4
		20-29	628	17 094	3,7
		30-54	930	23 694	3,9
		55-64	362	5 620	6,4
		65-74	234	3 029	7,7
		75+	311	2 618	11,9
	H	00-09	116	2 615	4,4
		10-14	73	1 247	5,8
		15-19	611	8 163	7,5
		20-29	1 532	21 467	7,1
		30-54	1 551	22 984	6,7
		55-64	428	4 675	9,2
65-74		259	2 667	9,7	
75+		390	2 630	14,8	

Type d'utilisateur	Sexe	Age	MAIS3+	MAIS1+	Ratio (%) MAIS3+/MAIS1+ (CRC France)
Piéton	F	00-09	110	1 219	9,0
		10-14	130	1 286	10,1
		15-19	138	1 528	9,0
		20-29	153	1 862	8,2
		30-54	344	3 543	9,7
		55-64	244	1 607	15,2
		65-74	220	1 316	16,7
		75+	544	2 069	26,3
	H	00-09	265	2 087	12,7
		10-14	201	1 485	13,6
		15-19	149	1 079	13,8
		20-29	196	1 446	13,6
		30-54	447	2 964	15,1
		55-64	226	1 056	21,4
65-74		184	894	20,6	
75+		349	1 164	30,0	
Autre	F	00-09	5	105	4,7
		10-14	13	208	6,2
		15-19	21	357	6,0
		20-29	38	552	6,8
		30-54	66	924	7,2
		55-64	23	230	9,8
		65-74	10	107	9,6
		75+	13	111	11,7
	H	00-09	9	124	7,1
		10-14	15	172	8,8
		15-19	53	446	11,9
		20-29	190	1 751	10,9
		30-54	398	3 473	11,5
		55-64	92	527	17,4
65-74		26	152	16,9	
75+		21	102	20,6	

CRC = capture-recapture sur BAAC et Registre

## 4 DISCUSSION

La définition du blessé grave adoptée par l'Europe repose sur l'échelle de gravité validée de l'Abbreviated Injury Scale. Le Registre du Rhône permet d'accéder à un nombre très important de victimes pour lesquelles chaque lésion a été codée avec cette échelle. Ceci n'est pas vrai dans la plupart des pays européens, qui, dans leur majorité, basent leurs estimations sur leurs données hospitalières. Celles-ci utilisent un codage valable pour toutes les pathologies (la CIM -10 ou la CIM-9), mais moins précis pour la traumatologie, et surtout n'ayant pas de niveau de gravité associé. Ces pays ont ainsi recours à des algorithmes de passage permettant d'indiquer pour chaque victime le niveau de gravité (équivalent à AIS $\geq$ 3 en oui/non) ce qui est évidemment moins satisfaisant que d'avoir le "vrai" codage AIS, ce d'autant plus qu'un certain nombre de codifications CIM n'ont pas d'équivalence en AIS.

Cependant, la plupart de ces pays qui fournissent les estimations des nombres de blessés graves MAIS3+ disposent de ces bases de données hospitalières au plan national.

En revanche, le Registre du Rhône présente deux avantages : la codification directe en AIS, et la présence de toutes les victimes, y compris non hospitalisées. Le principal point faible de nos données est bien sûr qu'elles ne sont disponibles que dans le département du Rhône, qui ne constitue pas un échantillon représentatif de la France.

La méthode que nous avons développée pour en tirer une estimation nationale comporte plusieurs étapes:

- Identifier les victimes corporelles d'accidents de la circulation routière dans le département du Rhône communes au Registre et au recueil des forces de l'ordre (BAAC)
- Estimer la gravité MAIS3+ pour les victimes non répertoriées par le Registre (dans le Rhône)
- Identifier les facteurs qui sont associés au non enregistrement dans les BAAC
- En fonction de cela, estimer par capture-recapture la population "réelle" des victimes dans le département du Rhône
- En déduire les coefficients permettant de passer du BAAC Rhône à cette population
- Appliquer ces coefficients correcteurs à l'ensemble des BAAC France métropolitaine et obtenir ainsi l'estimation nationale des blessés, notamment graves

En résumé, la méthode consiste à extrapoler la gravité MAIS3+ à l'ensemble des BAAC du Rhône, puis à redresser l'ensemble des BAAC France au moyen de coefficients qui tiennent compte des pratiques d'enregistrement des blessés par les forces de l'ordre (selon toutes les informations pertinentes présentes dans les BAAC).

Cette méthode comporte évidemment des points faibles qui ont déjà été précisés dans la section méthode, et d'autres qui sont évoqués ci-dessous. La validité des estimations qu'elle fournit ne peut pas être mesurée de façon certaine, puisque les nombres réels de blessés et de blessés graves en France ne sont pas connus, mais nous pouvons donner plusieurs éléments de validation qui confortent les résultats fournis. Il s'agit de validation internes, en examinant la cohérence de certains indicateurs, ou de validation externes en se comparant à d'autres sources de données françaises ou d'autres résultats obtenus en Europe.

### 4.1 Éléments de validité des résultats

D'après Tilling et Sterne (Tilling and Sterne, 1999) les tests habituels d'adéquation d'un modèle aux données ne sont pas utilisables car ce sont des tests d'adéquation aux données observées, alors que l'approche capture-recapture repose sur l'existence de données non observées (les victimes non recensées).

En revanche, afin d'évaluer la validité de notre procédure, nous comparons nos estimations soit avec d'autres résultats provenant de données à notre disposition (validation interne), soit avec des estimations extérieures quand elles existent (validation externe).

#### 4.1.1 Validation interne : Gravité selon le Registre du Rhône

Le Registre étant quasi exhaustif et donc représentatif de l'ensemble des blessés de la route du Rhône, la gravité mesurée sur ses données (par la proportion de MAIS3+ parmi les MAIS1+) est fiable.

La France étant moins urbaine que le Rhône on s'attend à ce que la gravité soit plus élevée : les accidents sont plus graves en moyenne en zone rurale (car les vitesses plus élevées). C'est ce que l'on observe : 8,6% sur les chiffres estimés France entière et 7,3% sur le Registre du Rhône.

Nous comparons aussi la proportion de MAIS3+ parmi les MAIS1+ observée dans le Registre, avec celle calculée sur les données BAAC France extrapolées suivant différentes caractéristiques : âge, sexe, mode de transport.

Nous vérifions (tableau 51) que sur les estimations nationales les hommes restent plus gravement blessés que les femmes (respectivement 9,6 et 5,9 % de blessés graves) ; les âgés restent plus gravement blessés que les plus jeunes (17,9% de graves chez les 75 ans et plus versus 6,3% chez les 0-9 ans).

Par type d'utilisateur, nous vérifions (tableau 43) que le « pattern » est conservé : les usagers de deux-roues motorisés, avec tiers ont la même gravité que les piétons (avec tiers par définition) et tous deux ont la gravité la plus élevée. Les blessés cyclistes, avec tiers, viennent en troisième position en termes de gravité. Pour les usagers de deux-roues motorisés et les cyclistes, la gravité des blessés avec tiers, reste plus élevée que celle des blessés sans tiers. Pour les automobilistes et les blessés « autres » (VU, PL, bus, ...), les blessés dans un accident avec tiers restent moins graves que les blessés sans tiers. Enfin, les blessés automobilistes restent les moins gravement blessés.

Ainsi la répartition de la gravité dans nos estimations finales est parfaitement cohérente avec ce qu'on observe à partir du registre.

#### 4.1.2 Validation interne : Gravité parmi les blessés non observés

Il est utile dans la méthode capture-recapture de présenter les données selon le tableau 2x2 ci-dessous, avec les quatre situations :

- blessés enregistrés dans les BAAC et dans le Registre,
- blessés enregistrés dans les BAAC mais pas dans le Registre,
- blessés enregistrés dans le Registre mais pas dans les BAAC,
- blessés enregistrés ni dans le Registre ni dans les BAAC.

On peut alors observer ou estimer la gravité (proportion de MAIS3+ pour les MAIS1+) dans les quatre sous-ensembles. On s'attend à trouver une gravité plus faible chez les blessés non observés (ni dans les données BAAC ni dans le Registre) en comparaison avec les trois autres situations. Le tableau ci-après donne la gravité dans les quatre situations obtenues.

*Tableau 54 : Proportions de MAIS3+/MAIS1+ (entre parenthèses : effectifs moyens annuels MAIS3+/MAIS1+) d'après les BAAC et le Registre dans le Rhône, 2006-2012*

		Dans le Registre ?	
		Oui	Non
Dans les BAAC ?	Oui	16% (344/2100) <i>(MAIS3+ observé)</i>	12% (84/709) <i>(MAIS3+ prédit sur BAAC)</i>
	Non	4% (196/5321) <i>(MAIS3+ observé)</i>	6% (92/1483) <i>(Non observés)</i>

Chez les blessés non-observés, mais estimés, on obtient une proportion de 6% de MAIS3+, et donc bien inférieure à la proportion (16%) des blessés MAIS3+ qui sont répertoriés à la fois dans le Registre et dans les BAAC. Cette proportion de 6% est aussi inférieure au 12% de MAIS3+ prédits chez les blessés enregistrés dans les BAAC mais pas dans le Registre. Elle est cependant légèrement au-dessus des 4% observés chez les blessés enregistrés dans le Registre mais pas dans les BAAC. Cela n'est pas parfait, mais c'est très proche de ce que l'on attend, alors que d'autres modélisations explorées (cf annexe 6) avaient conduit à des valeurs de 18 % ou 24% de proportion de MAIS3+ chez les non-observés, ce qui n'est pas plausible.

#### 4.1.3 Validation externe : l'Enquête Nationale Transport et Déplacements (ENTD)

Une validation externe est fournie par l'ENTD (Enquête Nationale Transports Déplacements) qui s'est déroulée en 2007-2008. Cette enquête se base sur 20 200 personnes (âgées de 6 ans et plus). Il leur a été demandé si elles ont eu un /des accidents corporels au cours des 5 dernières années, de la façon suivante : « (prénom) a-t-il/elle été victime d'un accident de la circulation entraînant des dommages corporels au cours des 5 dernières années ? (accidents ayant provoqué une blessure donnant lieu à un acte médical : médecin, hôpital...) ». Les réponses à cette question ont permis d'estimer (Haddak M (sous la coordination de) et al., 2013) le nombre de blessés de la route, toutes gravités, par année, France entière (tableau ci-après).

*Tableau 55 : estimation du nombre de blessés selon l'ENTD (2007-2008), personnes âgées de 6 ans et plus*

année	Effectif estimé ENTD
2003	369 303
2004	323 796
2005	399 581
2006	469 789
2007	416 544

On note une certaine fluctuation d'une année à l'autre (avec un biais de mémoire pour les années les plus éloignées dans le temps par rapport à la date de l'enquête, 2007-2008), mais on a globalement un ordre de grandeur de 400 000.

Les estimations obtenues par capture recapture sur la France entière sur 2006-2012, du nombre de blessés toutes gravités sont de 360 522 blessés toutes gravités en 2006 et 382 189 en 2007, auxquels on retranche les blessés de moins de 6 ans, ce qui nous donne 352 017 blessés en 2006 et 374 539 en 2007. C'est donc un peu en dessous des estimations de l'ENTD, mais dans le même ordre de grandeur, autour de 400 000.

#### 4.1.4 Validation externe : Effectifs de blessés selon le SDIS= Service Départemental d'Incendie et de Secours

Les Services Départementaux d'Incendie et de Secours peuvent fournir des chiffres de blessés de la route, pour les interventions qu'ils ont effectuées. Ainsi, sur la France entière (Bondaz et al., 2014), les chiffres sont pour 2012 de 272 000 blessés, ce qui est proche de notre estimation de 307 000 blessés toutes gravités pour la même année.

Tableau 56 : nombre de blessés toutes gravités, chiffres SDIS (source= INFOSDIS)

année	Blessés SDIS
2009	288 700
2010	273 862
2011	278 028
2012	271 796
2013	257 735
2014	270 142
2015	269 539

#### 4.1.5 Validation externe : ratio blessés toutes gravités /tués selon les pays

Un autre élément de validation vient du ratio blessés toutes gravités/tués. Ce ratio ne devrait pas être très différent entre pays européens semblables, ou plutôt devrait être du même ordre de grandeur. Dans le graphique ci-après on donne les ratios de quelques pays de l'Union Européenne pour 2009. Le ratio blessés toutes gravités/tués en France, non corrigé, vaut 14, en dessous de la moyenne européenne à 34, et aussi en dessous de l'Espagne et de l'Italie. En corrigeant du sous-enregistrement des BAAC, le ratio français atteint alors 70, ce qui le situe au même niveau que l'Allemagne et le Royaume-Uni, sur leurs données non corrigées.

Ces trois éléments extérieurs viennent confirmer l'ordre de grandeur des estimations du nombre de blessés toutes gravités.

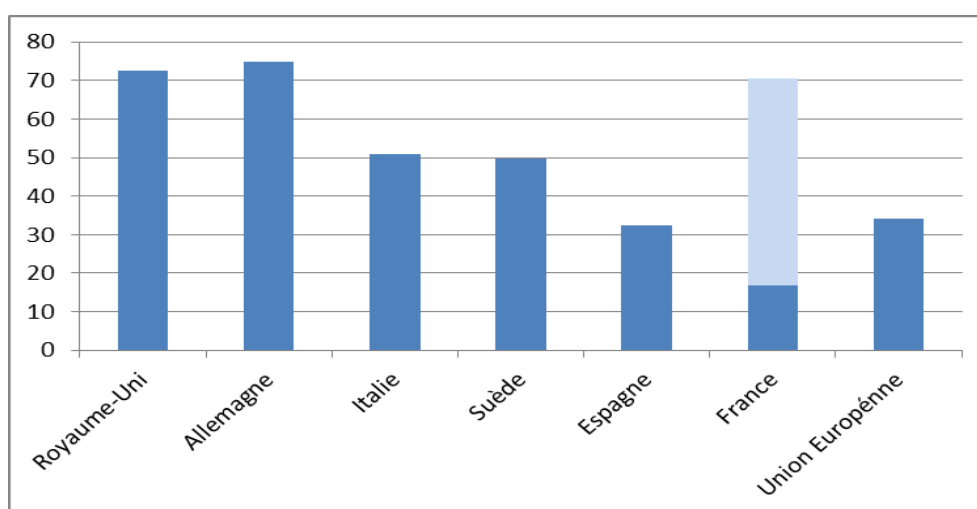


Figure 11 : ratios blessés toutes gravités / tués, 2009



#### 4.1.6 Validation externe : ratio blessés graves MAIS3+/tués selon les pays

Il est plus difficile de valider le nombre de blessés graves estimé car il n'est pas directement disponible dans d'autres sources. Un élément de validation du nombre de blessés graves vient du ratio blessés graves/tués. Comme pour le ratio blessés toutes gravités/tués, celui-ci devrait être du même ordre de grandeur entre les différents pays européens.

Concernant les méthodes d'estimation du nombre de blessés graves (i.e. MAIS3+), le Royaume-Uni se base sur une méthode de coefficients de correction sur les données policières, hormis l'Angleterre qui se base uniquement sur les données hospitalières. L'Espagne, l'Italie et la Suède se basent également uniquement sur les données hospitalières tandis que la Belgique se base sur des coefficients de correction sur les données policières. Les Pays-Bas se basent sur le chaînage entre les données policières et hospitalières, suivi d'une approche capture recapture (Perez et al., 2016). Pour finir, l'Allemagne se base sur une méthode utilisant la base de données GIDAS (German In-Depth Accident Study) et l'Official German Road Accident Statistics (OGRAS). Une estimation de la probabilité pour un blessé de la route d'être MAIS3+ selon le profil d'accident est effectuée à l'aide de GIDAS. Cela est possible car l' AIS y est disponible. Ces probabilités sont ensuite appliquées aux effectifs des profils des hospitalisés accidentés de la route dans la base OGRAS, permettant ainsi l'estimation du nombre de MAIS3+ au niveau national (Niesen et al., 2016).

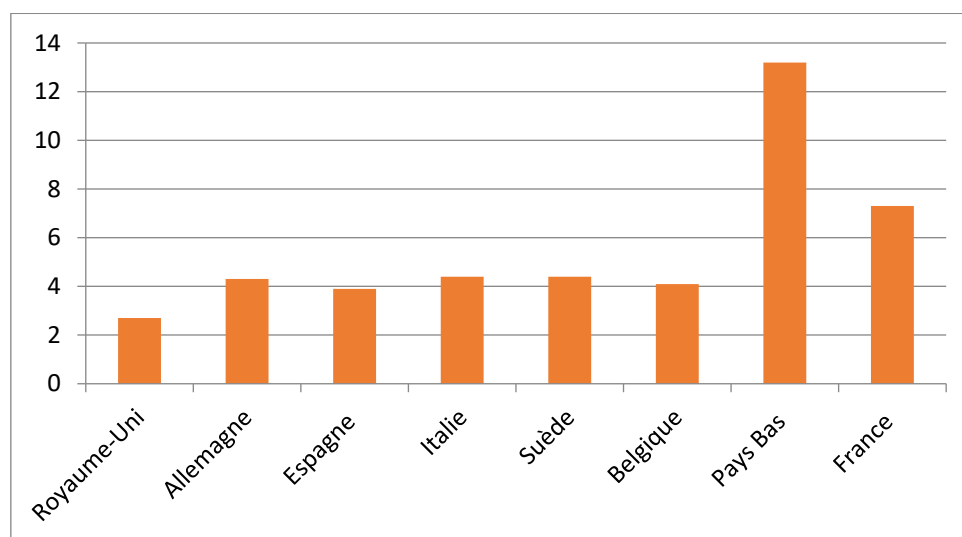


Figure 12 : ratios blessés graves MAIS3+ / tués, 2014

En 2014, le ratio de la France, qui vaut 7,3, est plus élevé que la majorité des ratios de l'Union Européenne qui sont de quatre blessés graves pour un tué. Le ratio français blessés MAIS3+/tués est cependant moins élevé que celui des Pays-Bas, qui vaut 12 et qui est le pays qui utilise la méthode la plus similaire à la nôtre.

Outre la méthode utilisée, ces différences peuvent s'expliquer par les variations de ce ratio selon le mode de transport, et la variation de répartition des modes de transport entre les pays: d'après nos données, en 2014, ce ratio est de 24 blessés graves pour un tué chez les cyclistes, 7 chez les piétons, 13 chez les deux-roues motorisés et 4 chez les automobilistes. Ce ratio étant très élevé chez les cyclistes et les Pays-Bas ayant une très forte proportion de cyclistes, cela peut expliquer que leur ratio global soit élevé (ils ont les mêmes ratios par type d'usager que les nôtres).

Il est à noter que les pays se basant uniquement sur leur source hospitalière sous-estiment sans doute l'effectif de blessés (puisque, dès qu'il y a une deuxième source, policière, et qu'on la croise avec la source hospitalière, on constate que chaque source est incomplète).

## 4.2 Limites

### 4.2.1 Manque de fiabilité du nombre estimé d'hospitalisés

Un nombre de blessés hospitalisés corrigé du sous-enregistrement peut être estimé (51 000 en 2012). Cependant, cette estimation n'est pas très fiable car basée sur un mélange entre la variable hospitalisé du Registre, fiable et la variable hospitalisé des BAAC, pas très fiable (cf. annexe 5).

Ce manque de fiabilité apparaît dans le fait que les différentes approches explorées (selon par exemple si capture-recapture incluait la variable MAIS ou pas, cf. annexe 6) ont abouti à des estimations du nombre d'hospitalisés qui varient du simple au double : 49 000 à 102 000.

En revanche, le nombre de blessés MAIS1+ était plutôt stable entre les différentes approches.

### 4.2.2 Nombre de blessés hospitalisés à partir du PMSI

Une approche différente de celle présentée dans ce rapport peut être établie pour l'estimation du nombre de blessés hospitalisés, basée sur le PMSI (Programme de Médicalisation des Système d'Information).

Le PMSI contient depuis les années 2000, au niveau national, l'ensemble des séjours hospitaliers (séjour où il y a eu une hospitalisation).

L'objectif premier de cette base de données est une gestion budgétaire. Cependant, elle peut être utilisée en épidémiologie.

Les personnes hospitalisées suite à une lésion traumatique (une blessure) peuvent être identifiées par la Classification Internationale des Maladies (CIM), en sélectionnant les codes S et T. Elles sont au nombre de 780 000 en 2012. Parmi ces hospitalisés pour traumatisme physique, il est en théorie possible d'identifier ceux dont la cause est un accident de la route, parmi l'ensemble des causes possibles (accidents de la vie courante, accidents du travail, ... ou blessures intentionnelles). Cette information est codée comme « cause externe » dans la CIM. Elle est cependant manquante à 80%. Sur les 20% où la cause externe est renseignée, soit 16 000 personnes hospitalisées, les accidents de la route représentent 17% des causes. Si l'échantillon où la cause externe est renseignée était représentatif de l'ensemble, le nombre d'hospitalisés suite à un accident de la route serait 17% de 780 000, soit 133 000. Des travaux sont en cours pour redresser cet échantillon.

### 4.2.3 Effets du Registre sur l'enregistrement par les forces de l'ordre du Rhône ?

Les services hospitaliers participant au Registre indiquent recevoir un certain nombre de visites des forces de l'ordre (qui viennent chercher des informations notamment sur l'hospitalisation (oui/non) des blessés de la route dont ils ont eu connaissance) et il semblerait que ce nombre ait augmenté avec les années d'existence du Registre. On peut ainsi se demander s'il y a un effet du Registre du Rhône sur les pratiques d'enregistrement des forces de l'ordre (du Rhône).

Le projet VOIESUR a donné l'opportunité de quantifier certains éléments à ce sujet.

Ce projet repose sur une base de données d'accidents constituée à partir de l'examen minutieux des PV des forces de l'ordre, ceci pour l'ensemble des accidents mortels survenus en France métropolitaine en 2011, sur un échantillon d'1/20<sup>ème</sup> des accidents corporels et pour l'ensemble des accidents corporels survenus dans le département du Rhône, pour cette même année. Dans les analyses faites au cours de ce projet, il est apparu que les PV du Rhône étaient un peu mieux renseignés que les PV France entière. Ainsi, la proportion de PV comprenant un bilan lésionnel est de 67% dans les PV 1/20<sup>ème</sup> France entière versus 83% dans le Rhône. De même, la vitesse à laquelle le véhicule roulait a pu être estimée (et pour cela il faut un certain nombre d'informations) dans 36% des PV 1/20<sup>ème</sup> France entière vs 49% des PV du Rhône, et pour la vitesse au choc : 11.5% versus 18.8%. Il semble donc qu'il y ait un effet du Registre dans le sens d'un meilleur contenu des PV. Est-ce que cela pourrait se traduire par un meilleur taux d'enregistrement des blessés dans le Rhône qu'ailleurs ?

Si tel était le cas cela signifierait que les coefficients de correction du sous-enregistrement, estimés sur le Rhône sous-estiment les coefficients correcteurs qu'il faudrait appliquer dans le reste de la France. Cela signifie que les effectifs totaux de blessés estimés sous-estimeraient la réalité.

Il est à noter que cela va dans le même sens, c'est-à-dire une sous-estimation des effectifs de blessés, que la dépendance positive entre l'enregistrement des blessés dans les deux sources, BAAC et Registre, pour le département du Rhône.

Dit autrement, cela signifie que les résultats obtenus sont sans doute des valeurs planchers.

## 5 REMERCIEMENTS

Nous tenons à remercier toutes les personnes qui ont participé au recueil ou à l'informatisation des données, pour l'Association pour le Registre des Victimes d'Accidents de la Circulation du Rhône (ARVAC : président Etienne Javouhey) et l'IFSTTAR-UMRESTTE (Bernard Laumon, responsable scientifique du Registre, Amina Ndiaye, coordinatrice médicale du Registre, et Blandine Gadegbeku, responsable de la gestion et de l'analyse des données du Registre) : Ait Idir T, Ait Si Selmi T, Alloatti D, Amoros E, Andrillat M, Artru F, Asencio Y, Assossou I, Auzaneau F, Bagès-Limoges F, Bagou G, Balogh C, Banssillon G, Banssillon V, Barnier N, Barth X, Basset M, Beaudry E, Bec JF, Bejui J, Bel JC, Belhadj A, Bérard E, Bérard J, Bergeron I, Bernard JC, Berthet N, Bertrand JC, Besson L, Biot B, Biot V, Blache D, Blanc C, Blanchard J, Blum O, Bœuf C, Boisson D, Bonjean M, Bost J, Bouchedor C, Boughtatene B, Bouletreau P, Boyer M., Boyer V, Breda Y, Brilland R, Busserly S, Cabet N, Caillot L, Caillot JL, Cannamela A, Caregnato B, Carre M, Catala Y, Chagnon PY, Chambost M, Chantran C, Chapuy P, Chardon P, Charnay P, Chatelain P, Chattard S, Chauvin F, Chavane H, Chazot G, Chenani M, Chettouane I, Chevreton N, Chevrillon E, Chevrillon S, Chiron M, Chotel P, Cochard P, Comarmond M, Combe C, Comte G, Contamin B, Coppard E, Cot T, Coutagnieux A, Crettenet Z, Cristini A, Cunin V, Danella-Marguiron C, Dailler F, Dal Gobbo B, David JS, De Angelis MP, Debrisay MC, Decourt L, Degrange P, Dejour H.†, Delfosse A, Demangel N†, Demazière J, Deruty R, Desjardins G, Devaux J, Dewhurst N, Dohin B, Drouet A, Du Besset M, Ducreux B, Ducrozet P, Dufour A, Dumortier C, El Khoury C, Emonet A, Escarment J, Evrard AS, Eyssette M, Fallavier L, Fanton L, Felten D, Feugere S, Feuglet P†, Fifis N, Figura J, Fisher G, Fischer LP, Floccard B, Floret D, Fournier G, Fraisse P, Fredenucci JF, Freidel M, Fuster P, Gadegbeku B, Galin L, Gaillard P, Gallon M, Garnier N, Garzanti A, Gaussorgues P, Gautheron V, Genevrier M, Gibaud F, Gillet Y, Gilly F, Girard A, Goubisky A, Gougne M. Granger M, Grattard P, Gueugniaud PY, Guenot C, Guérin AC, Guerin T, Guignand M, Guillaudon C, Guillaumée F, Haddak M, Hamel D, Haouas T, Heckel T, Herzberg G, Ho-Van-Truc P, Huguen-Schmitt D, Humbert C, Jacquemard C, Jault V, Jouanneau E, Joffre T, Kheniche F, Kohler R, Labed H, Lablanche C, Lafont S, Lagier C, Lapierre B, Laplace MC, La Rosa C, Lassaigne M, Laurent R, Lebel M, Leblay G, Le-Xuan I, Lieutaud T, Lille R, Linné M, Lorge S, Luauté J, Lucas R, Machin B, Maiello E, Malicier D, Mangola B, Marcotte G, Marduel YN, Marie-Catherine M, Martin JL, Martin YN, Martinand G, Marty F, Matricon P, Mazouzi S†, Melaine R, Menard B, Messikh C, Meyer F, Meyrand S, Mintsya-Eya C, Molard S, Monchanin C, Monneuse O, Morel-Chevillet E, Mioulet E, Minjaud F, Mokdadi A, Mollet C, Monnet J, Moyen B, Neidhart JP, Ngandu E, Ny S, Ollagnier F, Ould T, Paget P, Paillot JC, Paris D, Patay B, Pauget P, Peillon D, Perrin G, Perrin-Blondeau D, Petit P, Phelip H, Picaud S, Pinelli E, Paparel P, Piriou V, Piton JL, Plantier M, Pornon P, Potinet-Pagliarioli V, Pramayon C, Quelard B, Rakaa A, Raquin L, Remy C, Rerbal D, Rezig M, Ricard A, Richard A, Rigal F, Robert D, Rode G, Romanet JP, Rongieras F, Roset C, Rousson A, Roussouli P, Roux H, Ruhl C, Salamand J, Salord F, Sametzky P, Sayegh K, Sayous P, Sbraire N, Scappaticci N, Schiele P, Schneider M, Simonet C, Sigal A, Sindou M, Soldner R, Soudain M, Stagnara J, Stamm D, Suc B, Supernant K, Tardy H., Tasseau F, Taty M, Taveau G, Tazarourte K, Tell L, Theurey O, Thievon R, Thomas M, Tilhet-Coartet S, Tissot E, Toukou JC, Trifot M, Tronc F, Vallee B, Vallet G, Vancuyck A, Vergnes I, Verney MP, Viste A, Voiglio EJ, Vourey G, Vuillard J, Westphal M, Willemen Lt.

Nous remercions aussi MP Verney pour l'assistance secrétariale.

## 6 REFERENCES

- Abeni, D.D., Brancato, G., Perucci, C.A., 1994. Capture-recapture to estimate the size of the population with human immunodeficiency virus type 1 infection. *Epidemiology* 5, 410–4.
- Alho, J., 1990. Logistic regression in capture-recapture models. *Biometrics* 46, 623–635.
- Amoros, E., 2007. Les blessés par accidents de la route : estimation de leur nombre et de leur gravité lésionnelle, France, 1996-2004 ; modélisation à partir d'un registre médical (Rhône) et des données policières (France) (PhD thesis). Université Lyon 1.
- Amoros, E., Martin, J.L., Laumon, B., 2006. Under-reporting of road crash casualties in France. *Accident Analysis and Prevention* 38, 627–635.
- Aptel, I., Salmi, L.R., Masson, F., Bourdé, A., Henrion, G., Erny, P., 1999. Road accident statistics: discrepancies between police and hospital data in a French island. *Accident Analysis and Prevention* 31, 101–108.
- Ardilly, P., 1994. Les techniques de sondage. Editions Technip, Paris.
- Bernillon, P., Lievre, L., Pillonel, J., Laporte, A., Costagliola, D., 2000. Record-linkage between two anonymous databases for a capture-recapture estimation of underreporting of AIDS cases: France 1990-1993. *Int. J. Epidemiol.* 29, 168–174.
- Bezerra, K.F., Gurgel, R.Q., Ilozue, C., Castaneda, D.N., 2011. Estimating the number of street children and adolescents in two cities of Brazil using capture-recapture: Estimation using capture-recapture. *Journal of Paediatrics and Child Health* 47, 524–529. <https://doi.org/10.1111/j.1440-1754.2011.02015.x>
- Bloor, M., Leyland, A., Barnard, M., McKeganey, N., 1991. Estimating hidden populations: a new method of calculating the prevalence of drug-injecting and non-injecting female street prostitution. *British Journal of Addiction* 86, 1477–1483.
- Bondaz, M., Helbronner, C., Panhaleux, J., Paquette, O., Miramon, G., Sion, L., Aubert, D., Bodino, P., 2014. Evaluation de la politique de sécurité routière. Tome 1 (Rapport de diagnostic). Inspection générale de l'administration. Conseil général de l'environnement et du développement durable. Inspection générale de la police nationale. Inspection générale de la gendarmerie nationale.
- Brady, J.E., Friedman, S.R., Cooper, H.L.F., Flom, P.L., Tempalski, B., Gostnell, K., 2008. Estimating the Prevalence of Injection Drug Users in the U.S. and in Large U.S. Metropolitan Areas from 1992 to 2002. *Journal of Urban Health* 85, 323–351. <https://doi.org/10.1007/s11524-007-9248-5>
- Brenner, H., Stegmaier, C., Ziegler, H., 1995. Estimating completeness of cancer registration: an empirical evaluation of the two source capture-recapture approach in Germany. *Journal of Epidemiology and Community Health* 49, 426–30.
- Brenner, H., Stegmaier, C., Ziegler, H., 1994. Estimating completeness of cancer registration in Saarland/Germany with capture-recapture methods. *European Journal of Cancer* 30A, 1659–63.
- Brewer, D.D., Dudek, J.A., Potterat, J.J., Muth, S.Q., Roberts, J.M., Woodhouse, D.E., 2006. Extent, Trends, and Perpetrators of Prostitution-Related Homicide in the United States. *Journal of Forensic Sciences* 51, 1101–1108. <https://doi.org/10.1111/j.1556-4029.2006.00206.x>
- Chao, A., Tsay, P.K., Lin, S.H., Shau, W.Y., Chao, D.Y., 2001. The applications of capture-recapture models to epidemiological data. *Statistics in Medicine* 20, 3123–57.
- Chokotho, L.C., Matzopoulos, R., Myers, J.E., 2013. Assessing Quality of Existing Data Sources on Road Traffic Injuries (RTIs) and Their Utility in Informing Injury Prevention in the Western Cape Province, South Africa. *Traffic Injury Prevention* 14, 267–273. <https://doi.org/10.1080/15389588.2012.706760>
- Clark, D.E., 2004. Practical introduction to record linkage for injury research. *Injury Prevention* 10, 186–191.
- Cormack, R.M., 1989. Loglinear models for capture-recapture. *Biometrics* 45, 395–413.

- Crocetti, E., Miccinesi, G., Paci, E., Zappa, M., 2001. An application of the two-source capture-recapture method to estimate the completeness of the Tuscany Cancer Registry, Italy. *European Journal of Cancer Prevention* 10, 417–23.
- Deville, J., Sarndal, C., 1992. Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association* 87, 376–382.
- Dhillon, P.K., Lightstone, A.S., Peek-Asa, C., Kraus, J.F., 2001. Assessment of hospital and police ascertainment of automobile versus childhood pedestrian and bicyclist collisions. *Accident Analysis and Prevention* 33, 529–537.
- Donders, A., van der Heijden, G., Stijnen, T., Moons, K., 2006. Review: a gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology* 59, 1087–91.
- Espelt, A., Domingo-Salvany, A., Sánchez-Niubó, A., Marí-Dell’Olmo, M., Brugal, M.T., 2015. Estimating trends in the prevalence of problematic cocaine use (1999-2008). *Gaceta Sanitaria* 29, 353–357. <https://doi.org/10.1016/j.gaceta.2015.05.008>
- Estève, J., Benhamou, E., Raymond, L., 1994. *Descriptive epidemiology, Statistical methods in cancer research*. International agency for research on cancer, Lyon.
- Fienberg, S.E., 1992. Bibliography on capture-recapture modelling with applications to census undercount adjustments. *Survey methodology* 18, 143–154.
- Fisher, N., Turner, S.W., Pugh, R., Taylor, C., 1994. Estimating numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis. *British Medical Journal* 308, 27–30.
- Foster, P.J., Johnson, G.J., 2001. Glaucoma in China: how big is the problem? *British Journal of Ophthalmology* 85, 1277–82.
- Frey, C.M., Feuer, E.J., Timmel, M.J., 1994. Projection of incidence rates to a larger population using ecologic variables. *Statistics in Medicine* 13, 1755–70.
- Frischer, M., Bloor, M., Finlay, A., Goldberg, D., Green, S., Haw, S., McKeganey, N., Platt, S., 1991. A new method of estimating prevalence of injecting drug use in an urban population: results from a Scottish city. *International Journal of Epidemiology* 20, 997–1000.
- Frischer, M., Hickman, M., Kraus, L., Mariani, F., Wiessing, L., 2001. A comparison of different methods for estimating the prevalence of problematic drug misuse in Great Britain. *Addiction* 96, 1465–76.
- Frischer, M., Leyland, A., Cormack, R., Goldberg, D.J., Bloor, M., Green, S.T., Taylor, A., Covell, R., McKeganey, N., Platt, S., 1993. Estimating the population prevalence of injection drug use and infection with human immunodeficiency virus among injection drug users in Glasgow, Scotland. *American Journal of Epidemiology* 138, 170–81.
- Gallay, A., Nardone, A., Vaillant, V., Desenclos, J.C., 2002. La méthode capture-recapture appliquée à l'épidémiologie : principes, limites et applications. *Revue d'Epidémiologie et de Santé Publique* 50, 219–32.
- Garton, M.J., Abdalla, M.I., Reid, D.M., Russell, I.T., 1996. Estimating the point accuracy of population registers using capture-recapture methods in Scotland. *Journal of Epidemiology and Community Health* 50, 99–103.
- Gemmell, I., Millar, T., Hay, G., 2004. Capture-recapture estimates of problem drug use and the use of simulation based confidence intervals in a stratified analysis. *J Epidemiol Community Health* 58, 758–765.
- Gossop, M., Strang, J., Griffiths, P., Powis, B., Taylor, C., 1994. A ratio estimation method for determining the prevalence of cocaine use. *British Journal of Psychiatry* 164, 676–9.
- Haddak M (sous la coordination de), Bouaoun, L., Mintsya-Eya, C., Amoros, E., 2013. Rapport ISOMERR-Ménages n°3 : Mesure du risque d'accident de la route à l'aide d'indicateurs d'exposition au risque (Rapport de convention PREDIT4 (GO2)). Ifsttar-Umrestte, Bron (69).
- Hay, G., McKeganey, N., Hartnoll, R., Wiessing, L., 1999. Methodological guidelines to estimate the prevalence of problem drug use on the local level. *European Monitoring Centre for Drugs and Drug Addiction*, Lisbon.

- Hilsenbeck, S.G., Kurucz, C., Duncan, R.C., 1992. Estimation of completeness and adjustment of age-specific and age-standardized incidence rates. *Biometrics* 48, 1249–62.
- Hook, E.B., Regal, R.R., 2000a. On the need for a 16th and 17th recommendations for capture-recapture analysis. *Journal of Clinical Epidemiology* 53, 1275–7.
- Hook, E.B., Regal, R.R., 2000b. Accuracy of alternative approaches to capture-recapture estimates of disease frequency: internal validity analysis of data from five sources. *American Journal of Epidemiology* 152, 771–9.
- Hook, E.B., Regal, R.R., 1999. Recommendations for presentation and evaluation of capture-recapture estimates in epidemiology. *Journal of Clinical Epidemiology* 52, 917–26; discussion 929–33.
- Hook, E.B., Regal, R.R., 1995. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiologic Reviews* 17, 243–64.
- Howe, G.R., 1998. Use of computerized record linkage in cohort studies. *Epidemiologic Reviews* 20, 112–21.
- IWGDMF, 1995a. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation I: History and theoretical development. *American Journal of Epidemiology* 142, 1047–58.
- IWGDMF, 1995b. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation II: Applications in human diseases. *American Journal of Epidemiology* 142, 1059–68.
- James, H.F., 1991. Under-reporting of road traffic accidents. *Traffic Engineering and Control* 32, 573–583.
- Jarvis, S.N., Lowe, P.J., Avery, A., Levene, S., Cormack, R.M., 2000. Children are not goldfish - mark/recapture techniques and their application to injury data. *Injury Prevention* 6, 46–50.
- Jensen, O.M., Esteve, J., Moller, H., Renard, H., 1990. Cancer in the European Community and its member states. *European Journal of Cancer* 26, 1167–1256.
- Johnson, R.L., Gabella, B.A., Gerhart, K.A., McCray, J., Menconi, J.C., Whiteneck, G.G., 1997. Evaluating sources of traumatic spinal cord injury surveillance data in Colorado. *American Journal of Epidemiology* 146, 266–72.
- Johri, M., Kaplan, E., Levi, J., Novick, A., 1998. New approaches to HIV surveillance: means and ends. Summary report of conference held at Yale University, 21–22 May 1998, by the Law, Policy and Ethics Core, Center for Interdisciplinary Research on AIDS, Yale University. *AIDS Public Policy Journal* 14, 136–146.
- Jones, I.E., Cannan, R., Goulding, A., 2000. Distal forearm fractures in New Zealand children: annual rates in a geographically defined area. *New Zealand Medicine Journal* 113, 443–5.
- Kraemer, J.D., Benton, C.S., 2015. Disparities in road crash mortality among pedestrians using wheelchairs in the USA: results of a capture–recapture analysis. *BMJ open* 5, e008396.
- LaPorte, R.E., Dearwater, S.R., Chang, Y.F., Songer, T.J., Aaron, D.J., Anderson, R.L., Olsen, T., 1995. Efficiency and accuracy of disease monitoring systems: application of capture-recapture methods to injury monitoring. *American Journal of Epidemiology* 142, 1069–77.
- Leclerc, P., Vandal, A.C., Fall, A., Bruneau, J., Roy, É., Brisette, S., Archibald, C., Arruda, N., Morissette, C., 2014. Estimating the size of the population of persons who inject drugs in the island of Montréal, Canada, using a six-source capture–recapture model. *Drug and Alcohol Dependence* 142, 174–180. <https://doi.org/10.1016/j.drugalcdep.2014.06.022>
- Madigan, D., York, J., 1997. Bayesian methods for estimation of the size of a closed population. *Biometrika* 84, 19–31.
- Martin, J.L., Lenguerrand, E., 2008. A population based estimation of the driver protection provided by passenger cars: France 1996–2005. *Accident Analysis and Prevention* 40, 1811–1821. <https://doi.org/10.1016/j.aap.2008.07.001>
- Maxwell, J.C., 2000. Methods for estimating the number of “hard-core” drug users. *Substance Use and Misuse* 35, 399–420.
- Menegoz, F., Black, R.J., Arveux, P., Magne, V., Ferlay, J., Buemi, A., Carli, P.M., Chapelain, G., Faivre, J., Gignoux, M., Grosclaude, P., Mace-Lesec’h, J., Raverdy, N., Schaffer, P., 1997. Cancer

- incidence and mortality in France in 1975-95. *European Journal of Cancer Prevention* 6, 442–466.
- Meuleners, L.B., Lee, A.H., Cercarelli, L.R., Legge, M., 2006. Estimating crashes involving heavy vehicles in Western Australia, 1999-2000: A capture-recapture method. *Accident Analysis and Prevention* 38, 170–174.
- Miller, T.R., Gibson, R., Zaloshnja, E., Blincoe, L.J., Kindelberger, J., Strashny, A., Thomas, A., Ho, S., Bauer, M., Sperry, S., others, 2012. Underreporting of driver alcohol involvement in United States police and hospital records: capture-recapture estimates. *Annals* 87.
- Morrison, A., Stone, D.H., 2000. Capture-recapture: a useful methodological tool for counting traffic related injuries? *Injury Prevention* 6, 299–304.
- Nachbaur, C., Uhry, Z., Thélot, B., 2004. Estimation du taux d'incidence annuel. d'accidents de la vie courante, en France, en 2001. Presented at the Journées scientifiques de l'InVS.
- Neugebauer, R., Wittes, J., 1994. Voluntary and involuntary capture-recapture samples--problems in the estimation of hidden and elusive populations. *American Journal of Public Health* 84, 1068–9.
- Niesen, S., Auerbach, K., Lerner, M., Pastor, C., Schepers, A., Wisch, M., Lefering, R., Malczyk, A., Schmucker, U., 2016. Estimation of the Number of Seriously Injured Road Traffic Casualties in Germany Based on GIDAS, Using a Decision Tree Method and the Traumaregister DGU®, in: IRCOBI Conference Proceedings.
- Oosterlee, A., Vink, R.M., Smit, F., 2009. Prevalence of family violence in adults and children: estimates using the capture-recapture method. *The European Journal of Public Health* 19, 586–591. <https://doi.org/10.1093/eurpub/ckp065>
- Palusci, V.J., Wirtz, S.J., Covington, T.M., 2010. Using capture-recapture methods to better ascertain the incidence of fatal child maltreatment. *Child Abuse & Neglect* 34, 396–402. <https://doi.org/10.1016/j.chiabu.2009.11.002>
- Peabody, J.W., Schau, B., Lopez-Vidriero, M., Vestbo, J., Wade, S., Iqbal, A., 2005. COPD: A prevalence estimation model. *Respirology* 10, 594–602. <https://doi.org/10.1111/j.1440-1843.2005.00755.x>
- Perez, K., Weijermars, W., Amoros, E., Bauer, R., Bos, N., Dupont, E., Filtner, A.J., Houwing, S., Johannsen, H., Leskovsek, B., others, 2016. Practical guidelines for the registration and monitoring of serious traffic injuries, D7. 1 of the H2020 project SafetyCube.
- Pullum, T., Cappa, C., Orlando, J., Dank, M., Gunn, S., Mendenhall, M., Riordan, K., 2012. Systems and strategies for identifying and enumerating children outside of family care. *Child Abuse & Neglect* 36, 701–710. <https://doi.org/10.1016/j.chiabu.2012.09.003>
- Razzak, J.A., Luby, S.P., 1998. Estimating deaths and injuries due to road traffic accidents in Karachi, Pakistan, through the capture-recapture method. *International Journal of Epidemiology* 27, 866–870.
- Roberts, I., Scragg, R., 1994. Application of capture-recapture methodology to estimate the completeness of child injury surveillance. *Journal of Paediatrics and Child Health* 30, 513–4.
- Robles, S., Marrett, L., Clarke, E., Risch, H., 1988. An application of capture-recapture methods to the estimation of completeness of cancer registration. *Journal of Clinical Epidemiology* 41, 495–501.
- Rosenman, K.D., Kalush, A., Reilly, M.J., Gardiner, J.C., Reeves, M., Luo, Z., 2006. How much work-related injury and illness is missed by the current national surveillance system? *Journal of Occupational and Environmental Medicine* 48, 357–365.
- Sautory, O., 1993. La macro CALMAR, redressement d'un échantillon par calages sur marges, Documents de travail. INSEE, Paris.
- Schouten, L.J., Straatman, H., Kiemeny, L.A., Gimbrere, C.H., Verbeek, A.L., 1994. The capture-recapture method for estimation of cancer registry completeness: a useful tool? *International Journal of Epidemiology* 23, 1111–6.
- Sherman, G., 1981. Cancer incidence in Canada: completeness and ecological correlations. University of Toronto, Toronto.



- Short, J., Caulfield, B., 2016. Record linkage for road traffic injuries in Ireland using police hospital and injury claims data. *Journal of Safety Research* 58, 1–14. <https://doi.org/10.1016/j.jsr.2016.05.002>
- Siegler, A., Roberts, L., Balch, E., Bargues, E., Bhalla, A., Bills, C., Dzung, E., Epelboym, Y., Foster, T., Fulton, L., Gallagher, M., Gastolomendo, J.D., Giorgi, G., Habtehans, S., Kim, J., McGee, B., McMahan, A., Riese, S., Santamaria-Schwartz, R., Walsh, F., Wahlstrom, J., Wedeles, J., 2008. Media Coverage of Violent Deaths in Iraq: An Opportunistic Capture-Recapture Assessment. *Prehospital and Disaster Medicine* 23, 369–371. <https://doi.org/10.1017/S1049023X00006026>
- Stutts, J.C., Williamson, J.E., Sheldon, F.C., 1988. Bicycle accidents: an examination of hospital emergency room reports and comparison with police accident data. *Transportation Research Record* 1168, 60–71.
- Takala, J., 1999. Global estimates of fatal occupational accidents. *Epidemiology* 10, 640–6.
- Tercero, F., Andersson, R., 2004. Measuring transport injuries in a developing country: an application of the capture-recapture method. *Accident Analysis and Prevention* 36, 13–20.
- Thomas, A.M., Thygeson, S.M., Merrill, R.M., Cook, L.J., 2012. Identifying Work-Related Motor Vehicle Crashes in Multiple Databases. *Traffic Injury Prevention* 13, 348–354. <https://doi.org/10.1080/15389588.2012.658480>
- Tilling, K., Sterne, J.A., 1999. Capture-recapture models including covariate effects. *American Journal of Epidemiology* 149, 392–400.
- Tin, S.T., Woodward, A., Ameratunga, S., 2013. Completeness and accuracy of crash outcome data in a cohort of cyclists: a validation study. *BMC Public Health* 13, 420.

## 7 ANNEXES

## Annexe 1 : Estimation du nombre de faux positifs et faux négatifs lié au chaînage BAAC et Registre du Rhône

Lors de la procédure de rapprochement ou chaînage (record-linkage), on compare les victimes des BAAC (Rhône) à celles du Registre en formant des paires constituées d'un enregistrement d'une victime des BAAC et d'un enregistrement d'une victime du Registre. Ces paires peuvent être classées de la façon suivante, dans le tableau ci-après.

*Tableau 57 : description d'un résultat de chaînage entre deux fichiers d'enregistrement (unité=une paire d'enregistrements)*

décision :	situation réelle : enregistrements de :	
	même individu (Matched)	individus différents (Unmatched)
rapprochement	vrais positifs	faux positifs
non-rapprochement	faux négatifs	vrais négatifs

Positif/négatif se réfère au résultat de la décision : positif signifie « rapproché » ou « relié » ou « chaîné », et négatif signifie non-rapproché. Faux et vrai qualifient la décision par rapport à la réalité (mais non connue). Dans la terminologie du record-linkage, l'ensemble des paires d'enregistrements correspondant (réellement) au même individu est noté M pour Matched, et l'ensemble des paires d'enregistrements correspondant (réellement) à des individus distincts est noté U pour Unmatched. Matched et Unmatched constituent ensemble le gold standard. Enfin, « relié » se traduit par « linked ».

Les faux positifs sont les paires d'enregistrements reliées à tort, car ne correspondant pas à la même victime. Les faux négatifs sont les paires non-reliées, à tort, c'est-à-dire deux enregistrements non reliés alors qu'ils correspondent à la même victime (Clark, 2004).

Nous allons améliorer le résultat du chaînage par l'estimation du nombre de faux positifs et de faux négatifs.

### 7.1.1 Estimation du nombre de faux positifs

Un faux positif est une paire d'enregistrements de 2 victimes distinctes que l'on a reliées (à tort donc).

$P(\text{faux positifs})$

$=P(\text{rapprocher alors qu'il s'agit d'enregistrements de 2 victimes distinctes})$

$=P(\text{rapprocher} / \text{victimes distinctes})$

$=P(\text{concordance sur toutes ou une partie des variables de chaînage} / \text{victimes distinctes})$

La concordance sur toutes ou certaines variables est précisément définie. Cela correspond aux critères de décision de la procédure de rapprochement.

Les variables de chaînage sont les suivantes :

- 1) jour de l'accident
- 2) mois de l'accident
- (année de l'accident)

- 3) lieu de l'accident
- 4) mois de naissance de la victime
- 5) année de naissance de la victime
- 6) sexe de la victime
- 7) type d'usager de la victime (piéton, cycliste, usager deux-roues motorisés, automobiliste, autre)

Remarque 1 : l'année de l'accident n'est pas considérée comme une variable de chaînage car elle est en fait utilisée comme variable de blocage, c'est-à-dire que la procédure de chaînage est conduite par blocs, qui sont définis par les années. Cela suppose l'absence d'erreur sur cette variable. Une précaution est quand même prise : pour les accidents ayant lieu en fin d'année (d'après les BAAC), les enregistrements du Registre du début de l'année suivante sont explorés.

Remarque 2 : si lors de la comparaison des modalités BAAC et Registre d'une variable de chaînage, il y a une valeur manquante, nous considérons que le résultat de la comparaison est discordant.

Nous décrivons ce qui se passait sur la période 1996-2001 (car cela va nous servir pour la période 2006-2012) ; la décision de relier la paire se faisait si :

- les sept variables de chaînage concordent
- seulement six variables concordent
- seulement cinq variables concordent, mais les deux variables discordantes sont des variables considérées comme « mineures » : mois de naissance, sexe ou type d'usager de la victime.

Il y a eu un changement de stratégie de chaînage à partir de 2002. Depuis cette date, le chaînage entre les données des forces de l'ordre et le Registre se fait lors de la saisie des fiches de notification du Registre, avec l'idée, s'il y a chaînage pour la victime en cours de saisie, de récupérer une partie des informations des forces de l'ordre afin de les incorporer à l'enregistrement en cours côté Registre : notamment les informations sur les circonstances de l'accident, jugées plus fiables côté BAAC. Cela conduit à une décision de rapprochement qui est plus restrictive. Concrètement : seulement une discordance sur une seule variable est tolérée (et cela ne doit pas porter sur le mois de l'accident).

En résumé, depuis 2002, on décide de rapprocher deux enregistrements s'il y a :

- concordance sur les sept variables de chaînage, ou
- concordance sur six variables seulement (et la discordance ne concerne pas le mois de l'accident)

Ces deux situations étant disjointes, la probabilité de l'ensemble est la somme des probabilités des deux situations :

$P(\text{rapprocher} / \text{victimes distinctes})$

$= P(7 \text{ var concordent} / \text{victimes distinctes}) + P(\text{exactement 6 var concordent (sauf le mois)} / \text{victimes distinctes})$

Pour chacune des deux situations (écrites sur 2 lignes ci-dessous), en supposant l'indépendance des variables de chaînage, la probabilité de chacune s'écrit comme le produit des probabilités de concordance pour chaque variable de chaînage (entre les deux enregistrements), multipliée, dans la 2ème situation, par la probabilité de discordance sur une seule variable, autre que le mois de l'accident :

$$= \left[ \prod_{i=1}^{i=7} P(\text{var}_i \text{ concorde} / \text{victimes distinctes}) \right] + \left[ \sum_{j=1, j \neq 2}^7 \left( \prod_{i=1, i \neq j}^{i=7} P(\text{var}_i \text{ concorde} / \text{victimes distinctes}) \times P(\text{var}_j \text{ discorde} / \text{victimes distinctes}) \right) \right]$$

La concordance sur une variable alors qu'il s'agit de victimes distinctes correspond à une concordance par simple hasard (par exemple pour la variable type d'usager, ce seraient deux automobilistes). Pour les variables mois et jour (de naissance ou d'accident) pour lesquelles les modalités sont uniformément réparties (à condition que les effectifs soit suffisants), les probabilités de concordance par hasard seront de 1/12 (i.e. 0.083) pour le mois et environ 1/30 ( i.e. 0.033) pour le jour (Brenner and Schmidtman 1996).

Pour les autres variables, si l'on dispose du gold-standard (sous groupes Matched et Unmatched), alors les probabilités de concordance par hasard sont estimées par les proportions de concordances observées sur l'ensemble (Unmatched) des paires d'enregistrements de victimes (réellement) distinctes.

Comme nous ne le connaissons pas, nous choisissons d'approximer par l'ensemble des paires non-relées, voire par l'ensemble de toutes les paires possibles, car la différence entre ces deux sous-ensembles est négligeable. En effet, les tailles respectives sont (avec « # » signifiant effectif, et «  $\cap$  » signifiant intersection des 2 ensembles, i.e. blessés communs aux deux sources):

# paires non-relées = #BAAC \* #Registre - #(BAAC  $\cap$  Registre) et  
 # toutes les paires possibles = #BAAC \* #Registre,  
 avec #(BAAC  $\cap$  Registre) négligeable devant #BAAC \* #Registre.

Nous choisissons d'approximer par l'ensemble des paires non-relées, qui comprend quelques dizaines de millions de paires (cf. tableau ci-après).

**Tableau 58 : proportions de concordance sur les variables de chaînage, observées sur l'ensemble des paires non reliées, Registre et BAAC, Rhône, 1996-2004 et 2006-2012**

variables chaînage	1996 (n=45,4M)	1997 (n=43,6M)	1998 (n=50,2M)	1999 (n=51,9M)	2000 (n=44,1)	2001 (n=43,9M)	2002 (n=33,6M)	2003 (n=25,4M)	2004 (n=23,8M)
jour accident	3,3%	3,2%	3,3%	3,3%	3,3%	3,2%	3,3%	3,3%	3,3%
mois accident	8,4%	8,5%	8,4%	8,4%	8,4%	8,4%	8,4%	8,5%	8,5%
année accident	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
lieu de l'accident (commune)	9,0%	8,1%	10,6%	10,3%	8,1%	8,1%	6,8%	6,4%	5,6%
mois de naissance	8,3%	8,3%	8,3%	8,3%	8,3%	8,3%	8,2%	8,2%	8,3%
année de naissance	1,9%	1,9%	1,9%	1,9%	2,0%	2,0%	1,9%	1,9%	1,8%
sexe	52,1%	52,5%	52,0%	52,3%	52,5%	52,4%	52,5%	53,3%	52,6%
type d'usager	38,4%	38,5%	40,2%	37,8%	38,3%	39,2%	38,0%	33,3%	33,0%

*proportions de concordance sur les variables de chaînage, observées sur l'ensemble des paires non reliées, Registre et BAAC, Rhône, 1996-2004 et 2006-2012*

variables chaînage	2006 (n=22,5M)	2007 (n=21,8M)	2008 (n=18,6M)	2009 (n=21,5M)	2010 (n=21,6)	2011 (n=18,5M)	2012 (n=20,2M)
jour accident	3,30%	3,30%	3,20%	3,30%	3,20%	3,30%	3,20%
mois accident	8,50%	8,50%	8,50%	8,50%	8,60%	8,50%	8,40%
année accident	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
lieu de l'accident commune	8,70%	10,50%	9,80%	9,30%	9,10%	9,80%	9,10%
mois de naissance	8,30%	8,40%	8,40%	8,40%	8,40%	8,40%	8,30%
année de naissance	1,90%	2,10%	1,90%	2,00%	1,90%	1,80%	1,80%
sexe	51,70%	52,90%	52,30%	52,80%	52,10%	52,30%	52,30%
type d'utilisateur	32,30%	32,20%	30,10%	30,30%	30,10%	29,60%	30,70%

Nous vérifions que la proportion observée de concordance sur le jour de l'accident (3,2%) est effectivement proche de 1/30, et celles observées sur le mois d'accident (8,4%), et le mois de naissance (8,3%) sont effectivement proches de 1/12.

Les proportions de concordance (par hasard) sont liées à la répartition des modalités. La baisse de la proportion de concordance sur le type d'utilisateur peut s'expliquer par une baisse de la proportion d'automobilistes, au profit d'autres modes ; la variable type d'utilisateur devient alors plus discriminante.

Concernant le lieu de l'accident, nous minorons la proportion obtenue (9%) car celle-ci est la concordance par hasard sur la commune, alors que c'est la concordance sur commune ET route ou rue qui est utilisée dans le chaînage ; celle-ci est bien plus faible. Un taux de concordance par hasard est arbitrairement fixé à 1%.

#### Calcul du nombre de faux positifs :

Nous notons :

- FP = faux positifs
- FN = faux négatifs
- VN = vrais négatifs
- VP = vrais positifs
- # = effectif

alors :

$$\begin{aligned} \#FP &= P(FP) \times \#VN \\ \#FP &= P(FP) \times (\#BAAC \times \#Registre - \#VP) \\ \#FP &= P(FP) \times \#BAAC \times \#Registre - P(FP) \times \#VP \end{aligned}$$

le dernier terme est négligeable (de l'ordre de 10.E-3 en regard d'un effectif)  
d'où :  $\#FP = P(FP) \times \#BAAC \times \#Registre$

Les résultats sur 1996-2004 et 2006-2012 sont donnés dans le tableau ci-après.

**Tableau 59 : proportions sur U (=Unmatched), de faux positifs et effectifs de faux positifs ; résultats du chaînage Registre et BAAC, Rhône, par année, sur 1996-2004 et 2006-2012**

critère	1996	1997	1998	1999	2000	2001	2002	2003	2004
P(FP)	2,0 E-6	2,1 E-6	2,1 E-6	2,0 E-6	2,1 E-6	2,1 E-6	1,7 E-6	1,5 E-6	1,4 E-6
# FP	92	90	107	105	94	92	56	38	34

critère	2006	2007	2008	2009	2010	2011	2012
P(FP)	1,5 E-6	1,5 E-6	1,4 E-6	1,4 E-6	1,4 E-6	1,3 E-6	1,3 E-6
# FP	32,8	33,7	25,3	30,3	29,1	24,2	26,6

On constate que la probabilité de faux positifs est légèrement plus faible à partir de 2002, date à laquelle la stratégie de chaînage devient plus restrictive : on retrouve dans les valeurs estimées l'idée logique, que si le chaînage est plus restrictif, la proportion de faux positifs diminue. A l'inverse, la proportion de faux négatifs devrait augmenter.

### 7.1.2 Estimation du nombre de faux négatifs

Un faux négatif est une paire d'enregistrements de la même victime que l'on n'a pas rapprochée (à tort donc).

P(faux négatif)

- =P(ne pas rapprocher alors qu'il s'agit de 2 enregistrements d'une même victime)
- =P(ne pas rapprocher / même victime)
- =P(discordance sur toutes ou une partie des variables de chaînage / même victime)
- =1-P(concordance sur toutes ou une partie des variables de chaînage / même victime)

« Concordance sur toutes ou une partie des variables de chaînage » est précisément défini. Cela correspond aux critères de décision de la procédure de rapprochement. Nous les rappelons ci-après.

Sur la période 1996-2001, la décision de relier la paire était prise si :

- les sept variables de chaînage concordent
- seulement six variables concordent
- seulement cinq variables concordent, mais les deux variables discordantes sont des variables considérées comme « mineures » : mois de naissance, sexe ou type d'utilisateur de la victime.

Depuis 2002, on décide un lien entre deux enregistrements s'il y a :

- concordance sur les 7 variables, ou
- concordance sur 6 variables seulement (et pas de discordance sur le mois de l'accident)

Ces deux situations étant disjointes, la probabilité de l'ensemble est la somme des probabilités des deux situations. En supposant l'indépendance des variables de chaînage, la probabilité de chacune des 2 situations s'écrit comme le produit des probabilités de concordance pour chaque variable de chaînage (entre les deux enregistrements), multipliée, dans la 2ème situation par la probabilité de discordance sur une seule variable, autre que le mois de l'accident :

P(concordance sur toutes ou une partie des variables de chaînage / même victime)

$$= P(7 \text{ variables concordent/ même victime}) + P(6 \text{ variables concordent/ même victime})$$

$$= \left[ \prod_{i=1}^{i=7} P(\text{var}_i \text{ concorde/ même victime}) \right]$$

$$+ \left[ \sum_{j=1, j \neq 2}^7 \left( \prod_{i=1, i \neq j}^{i=7} P(\text{var}_i \text{ concorde/ même victime}) \times P(\text{var}_j \text{ discorde/ même victime}) \right) \right]$$

D'après Brenner et Schmidtman, si on dispose du « gold standard », i.e. si on connaît le sous-ensemble M des (vraies) paires correspondant à une même victime, on remplace ces probabilités par les proportions observées sur l'ensemble M.

Comme nous ne disposons pas du gold standard, nous pouvons utiliser le sous-ensemble des paires rapprochées comme une approximation grossière du sous-ensemble des paires des mêmes victimes. Mais cela donnera plutôt une sur-estimation des probabilités de concordance. Nous décidons donc de diminuer ces proportions.

Un autre élément conduit aussi à diminuer ces proportions : chacune de ces probabilités est égale à 1-taux d'erreur (Howe, 1998), et cela ne paraît pas réaliste d'avoir des taux d'erreur aussi faibles que 2 ou 3%. Dans la littérature, selon une étude citée par Howe (Howe, 1998), le taux d'erreur sur l'année de naissance était de 13%.

**Tableau 60 : proportions de concordance sur quelques variables de chaînage, cité par Brenner, d'après Newcombe (Newcombe 1988),**

variables de chaînage	concordance parmi les paires de même individu
nom	96,5%
prénom	79,0%
initiale du 2 <sup>ème</sup> prénom	88,8%
jour de naissance	85,1%
mois de naissance	93,3%
année de naissance	77,3%
région de naissance	98,1%

Le tableau ci-après donne les proportions de concordance observées sur l'ensemble des paires reliées, par année.

**Tableau 61 : proportions de concordance sur les variables de chaînage, observées sur les paires reliées, Registre et BAAC, Rhône, 1996-2004**

variables de chaînage	1996 n=2919	1997 n=2616	1998 n=3169	1999 n=3040	2000 n=2761	2001 n=2809	2002 n=2013	2003 n= 1882	2004 n=1780
jour acc.	97,7%	97,9%	98,8%	98,3%	98,5%	97,3%	99,5%	99,7%	99,6%
mois acc	99,9%	99,8%	99,9%	99,8%	99,8%	99,9%	100,0%	100,0%	100,0%
année acc	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
lieu acc commune	85,4%	79,7%	82,4%	82,1%	80,6%	82,0%	98,4%	98,9%	98,8%
mois de naissance	88,5%	95,9%	94,5%	95,5%	95,8%	95,8%	97,2%	97,3%	98,4%
année de naissance	92,3%	97,0%	95,8%	96,2%	97,0%	96,5%	97,8%	98,4%	98,1%
sexe	96,9%	96,3%	96,9%	95,9%	96,0%	97,0%	98,6%	98,1%	98,7%
type d'utilisateur	95,9%	94,3%	95,6%	95,7%	95,1%	96,1%	98,0%	95,7%	98,7%



Les proportions sont élevées, et encore plus à partir de 2002 (surtout pour la variable commune), où la décision de relier est plus restrictive. Pour que les proportions de ces années-là ne soient pas dépendantes du changement dans la stratégie de chaînage, nous utilisons les proportions de l'ensemble de la période 1996-2001 (ce n'était pas visible sur les paires non-relées car le changement est dilué sur les quelques millions de paires non reliées). Ces proportions restent des proportions estimées sur les paires reliées, qui sur-estiment donc la probabilité de concordance sur les vrais positifs. Pour avoir des probabilités proches de celles de la littérature, nous minorons de 5% les proportions observées. Le tableau ci-après donne ces nouvelles valeurs.

**Tableau 62 : proportions de concordance sur les variables de chaînage, observées sur les paires reliées et minorées de 5%, Registre et BAAC, Rhône, 1996-2001**

variables de chaînage	1996	1997	1998	1999	2000	2001	moyenne
jour accident	93%	93%	94%	93%	94%	92%	<b>93%</b>
mois acc	95%	95%	95%	95%	95%	95%	<b>95%</b>
année acc	100%	100%	100%	100%	100%	100%	<b>100%</b>
lieu accident (commune)	81%	76%	78%	78%	77%	78%	<b>78%</b>
mois de naissance	84%	91%	90%	91%	91%	91%	<b>90%</b>
année de naissance	88%	92%	91%	91%	92%	92%	<b>91%</b>
sexe	92%	91%	92%	91%	91%	92%	<b>92%</b>
type usager	91%	90%	91%	91%	90%	91%	<b>91%</b>

#### Calcul du nombre de faux négatifs :

La probabilité de faux négatifs est définie sur l'ensemble des vrais positifs.

Avec les notations précédentes, on a :

$$\# \text{ FN} = P(\text{FN}) \times \# \text{ VP}$$

$$\# \text{ FN} = P(\text{FN}) \times (\# \text{ Pos} - \# \text{ FP} + \# \text{ FN})$$

$$(1 - P(\text{FN})) \times \# \text{ FN} = P(\text{FN}) \times (\# \text{ Pos} - \# \text{ FP})$$

$$\# \text{ FN} = P(\text{FN}) \times (\# \text{ Pos} - \# \text{ FP}) / (1 - P(\text{FN}))$$

Le tableau ci-après donne les résultats sur 1996-2004 et 2006-2012.

**Tableau 63 : proportions de concordance sur M (=Matched), de faux négatifs, d'effectifs de faux négatifs ; résultats du chaînage Registre et BAAC, Rhône, par année, 1996-2004 et 2006-2012**

critère	1996	1997	1998	1999	2000	2001	2002	2003	2004
P(concordance/ens. M)	84,9%	86,1%	87,0%	86,9%	86,7%	87,1%	82,4%	82,4%	82,4%
P(FN)	15,1%	13,9%	13,0%	13,1%	13,3%	12,9%	17,6%	17,6%	17,6%
# FN	505	408	457	445	410	403	417	393	369

critère	2006	2007	2008	2009	2010	2011	2012
P(concordance/ens. M)	82,40%	82,40%	82,40%	82,40%	82,40%	82,40%	82,40%
P(FN)	17,60%	17,60%	17,60%	17,60%	17,60%	17,60%	17,60%
# FN	375	361	332	346	379	354	358

On note que les probabilités de faux négatifs sont plus élevées à partir de 2002, ce qui est attendu pour une stratégie de chaînage plus restrictive.

## Annexe 2 : Modélisation multivariée de l'approche capture-recapture

Nous présentons d'abord la modélisation dans le cadre simple où il y aurait indépendance d'enregistrement entre les deux sources.

### 7.1.3 Sous l'hypothèse d'indépendance des deux sources

On formalise l'enregistrement par les deux sources BAAC et Registre avec le tableau suivant :

		Registre (R)		
		oui	non	
BAAC (B)	oui			$N_B$
	non			
		$N_R$		$N ?$

Avec

$N_B$ =Nombre de blessés dans les BAAC (B=BAAC)

$N_R$ =Nombre de blessés dans le Registre (R=Registre)

Le modèle s'écrit de la façon suivante :

$\Pi_B$  = probabilité d'être enregistré dans les BAAC

$\Pi_R$  = probabilité d'être enregistré dans le Registre

Dans le tableau croisé, cela correspond aux probabilités sur les marges :

		Registre (R)		
		oui	non	
BAAC (B)	oui			$\Pi_B$
	non		$\Pi_0$	
		$\Pi_R$		

On note  $l_B$  et  $l_R$  les logit de ces probabilités

$$l_B = \text{logit}(\Pi_B) = \ln \left( \frac{\pi_B}{1-\pi_B} \right)$$

$$l_R = \text{logit}(\Pi_R) = \ln \left( \frac{\pi_R}{1-\pi_R} \right)$$

Sous l'hypothèse d'indépendance des deux sources, la probabilité pour tout individu d'être manqué par les deux recensements vaut :

$$\Pi_0 = (1 - \Pi_B) (1 - \Pi_R) = \frac{1}{(1 + \exp(l_B)) \times (1 + \exp(l_R))}$$

Une variable indicatrice  $q_i$  est définie ; elle identifie si l'individu  $i$  est capturé par aucune source ( $q_i=0$ ), par la source BAAC seulement ( $q_i=1$ ), par la source Registre seulement ( $q_i=2$ ) ou par les deux sources ( $q_i=3$ ). En termes d'effectifs, cela s'écrit, dans le tableau croisé, de la façon suivante :

		Registre (R)		
		oui	non	
BAAC (B)	oui	$N_3$	$N_1$	$N_B$
	non	$N_2$	$\widehat{N}_0$	
		$N_R$		$\widehat{N}$

$N = N_1 + N_2 + N_3$  (en termes d'observé)

$\hat{N} = N + \hat{N}_0$  (le « ^ » dénote l'estimateur)

avec  $\hat{N}$  estimateur du nombre réel des blessés,  $\hat{N}_0$ , estimateur du nombre réel de blessés manqués

On conditionne sur les individus observés ( $q \neq 0$ ) afin d'estimer les paramètres du modèle, pour ensuite pouvoir estimer la taille de la population étudiée. On note  $k=1,2,3$  les sous-ensembles exclusifs « BAAC seulement », « Registre seulement » et « intersection BAAC-Registre ». On note  $p_{ik}$  avec  $k=1, 2, 3$  la probabilité que  $q_i = k$ , sachant que l'individu  $i$  est observé.

Par hypothèse d'homogénéité de capture, les  $p_{ik}$  sont identiques pour tous les individus  $i$  et donc  $p_{ik}$  est remplacé par  $p_k$ . On a les probabilités conditionnelles suivantes :

$p_1 = P(\text{être dans BAAC seulement}) / P(\text{être observé})$

$p_2 = P(\text{être dans Registre seulement}) / P(\text{être observé})$

$p_3 = P(\text{être dans l'intersection BAAC - Registre}) / P(\text{être observé})$

Cela correspond à l'écriture ci-dessous dans le tableau croisé :

		Registre (R)		
		oui	non	
BAAC (B)	oui	$p_3$	$p_1$	$\Pi_B$
	non	$p_2$	$\Pi_0$	
		$\Pi_R$		

On a donc, par hypothèse d'indépendance des deux sources :

$$p_1 = \frac{\Pi_B(1 - \Pi_R)}{\Pi_B(1 - \Pi_R) + (1 - \Pi_B)\Pi_R + \Pi_B\Pi_R} = \frac{\exp(l_B)}{\exp(l_B) + \exp(l_R) + \exp(l_B + l_R)}$$

De la même façon :

$$p_2 = \frac{\exp(l_R)}{\exp(l_B) + \exp(l_R) + \exp(l_B + l_R)}, \text{ et}$$

$$p_3 = \frac{\exp(l_B + l_R)}{\exp(l_B) + \exp(l_R) + \exp(l_B + l_R)}$$

On formule un modèle logit sur ces probabilités conditionnelles  $p_k$ . La variable à expliquer est une variable catégorielle nominale ; on utilise un modèle logit généralisé, en prenant  $k=1$  i.e. « BAAC seulement » comme référence. Le modèle s'écrit :

$$\text{logit}(p_k) = \log(p_k/p_1) = \beta_k, \quad k=2,3$$

et en posant  $\beta_1=0$  (correspondant à « BAAC seulement » pris comme référence).

En estimant le modèle à partir des données, on obtiendra une estimation des paramètres  $\beta_k$ . Or, les paramètres  $\beta_k$  sont reliés à  $l_1$  et  $l_2$  de la façon suivante :

$$\beta_2 = \log\left(\frac{p_2}{p_1}\right) = \log\left(\frac{\exp(l_R)/(\exp(l_B) + \exp(l_R) + \exp(l_B + l_R))}{\exp(l_B)/(\exp(l_B) + \exp(l_R) + \exp(l_B + l_R))}\right) = \log\left(\frac{\exp(l_R)}{\exp(l_B)}\right) = l_R - l_B$$

$$\beta_3 = \log\left(\frac{p_3}{p_1}\right) = \log\left(\frac{\exp(l_B + l_R)}{\exp(l_B)}\right) = l_B + l_R - l_B = l_R$$

et donc réciproquement :

$$I_R = \beta_3$$

$$I_B = \beta_3 - \beta_2$$

On estime le modèle sur les observations, et on obtient les estimations  $\hat{\beta}_2$  et  $\hat{\beta}_3$

En remplaçant les paramètres  $I_B$  et  $I_R$  par leurs estimations, fonctions des estimations des paramètres  $\beta_k$ , on obtient une estimation de  $\Pi_0$  :

$$\hat{\Pi}_0 = \frac{1}{(1 + \exp(\hat{I}_B)) \times (1 + \exp(\hat{I}_R))}$$

$$= \frac{1}{(1 + \exp(\hat{\beta}_3 - \hat{\beta}_2)) \times (1 + \exp(\hat{\beta}_3))}$$

$$= \frac{\exp(\hat{\beta}_2)}{(\exp(\hat{\beta}_2) + \exp(\hat{\beta}_3)) \times (1 + \exp(\hat{\beta}_3))}$$

On peut alors estimer le nombre total N. En effet :

$$\hat{\Pi}_0 = \frac{\hat{N}_0}{\hat{N}} \Leftrightarrow$$

$$\hat{\Pi}_0 = \frac{\hat{N} - (N_1 + N_2 + N_3)}{\hat{N}} \Leftrightarrow$$

$$1 - \hat{\Pi}_0 = \frac{(N_1 + N_2 + N_3)}{\hat{N}} \Leftrightarrow$$

$$\hat{N} = \frac{(N_1 + N_2 + N_3)}{1 - \hat{\Pi}_0}$$

La modélisation logit multinomiale nous donne aussi une estimation des probabilités  $\Pi_B$  et  $\Pi_R$  :

$$\hat{\Pi}_B = \frac{1}{1 + \exp(-\hat{I}_B)} = \frac{1}{1 + \exp(\hat{\beta}_2 - \hat{\beta}_3)} \text{ et}$$

$$\hat{\Pi}_R = \frac{1}{1 + \exp(-\hat{I}_R)} = \frac{1}{1 + \exp(-\hat{\beta}_3)}$$

L'inverse de  $\Pi_B$  correspond au coefficient de correction du sous-enregistrement des données des forces de l'ordre, facteur multiplicatif à appliquer aux effectifs des BAAC.

$$\hat{C}_B = \frac{1}{\hat{\Pi}_B} = 1 + \exp(\hat{\beta}_2 - \hat{\beta}_3), \text{ coefficient de correction du sous-enregistrement des données}$$

des forces de l'ordre.

On peut aussi estimer  $C_R$ , coefficient de correction du sous-enregistrement du Registre :

$$\hat{C}_R = \frac{1}{\hat{\Pi}_R} = 1 + \exp(-\hat{\beta}_3),$$

Enfin, on peut estimer  $C_3$ , coefficient de correction du sous-enregistrement de l'intersection BAAC Rhône et Registre :

$$\hat{C}_3 = \frac{1}{\hat{p}_3} = \left[ 1 + \exp(\hat{\beta}_2) + \exp(\hat{\beta}_3) \right] / \exp(\hat{\beta}_3)$$

Cela est utile pour toute analyse restreinte à ces données, qui ont l'avantage de combiner caractéristiques accidentologiques et caractéristiques médicales (par exemple : étudier les blessures des automobilistes sur telle région du corps, en fonction du lieu d'impact sur le véhicule)

#### 7.1.4 Sous l'hypothèse de dépendance des deux sources

Nous considérons maintenant la situation plus réaliste où les deux sources sont dépendantes, dans le sens où la probabilité d'enregistrement dans l'une et dans l'autre sont liées à une même caractéristique (notée  $X$ ) des individus à enregistrer. L'hypothèse d'indépendance des deux sources est alors posée, dans des catégories définies par  $X$ , c'est-à-dire conditionnellement à  $X$ .

Toujours en notant  $i$  l'individu, les probabilités d'enregistrement dans les sources BAAC (B) et Registre (R) sont notées  $\Pi_{Bi}$  et  $\Pi_{Ri}$ .

On modélise leur lien à la caractéristique  $X$  (quantitative) de la façon suivante :

$$\text{logit}(\Pi_{Bi}) = l_B + k_B X_i$$

$$\text{logit}(\Pi_{Ri}) = l_R + k_R X_i$$

Rien ne contraint  $k_B$  et  $k_R$  à être de valeurs égales.

On prend comme caractéristique une variable qualitative telle que la gravité lésionnelle en deux classes : BL, BH (gravité BAAC, BL= blessé léger, BH=blessé hospitalisé), que l'on note  $m$  ( $m=1, 2$ ). Pour une catégorie  $m$  donnée, le tableau croisé BAAC x Registre s'écrit :

catégorie $m$		Registre (R)		
		oui	non	
BAAC (B)	oui			$\Pi_{Bm}$
	non		$\Pi_{0m}$	
		$\Pi_{Rm}$		

Pour tout individu appartenant à la catégorie  $m$ , on pose :

$$\text{logit}(\Pi_{Bm}) = l_B + k_{Bm}$$

$$\text{logit}(\Pi_{Rm}) = l_R + k_{Rm}$$

La probabilité d'un individu  $i$ , de la catégorie  $m$ , d'être manqué (par les deux sources) est donnée, sous l'hypothèse d'indépendance des 2 sources à l'intérieur de la catégorie  $m$ , par :

$$\Pi_{0i} = (1 - \Pi_{Bi}) (1 - \Pi_{Ri})$$

Cela s'écrit, pour tout individu  $i$  dans la catégorie  $m$  :

$$\Pi_{0m} = (1 - \Pi_{Bm}) (1 - \Pi_{Rm})$$

$$\Pi_{0m} = \frac{1}{(1 + \exp(l_B + k_{Bm}))} \times \frac{1}{(1 + \exp(l_R + k_{Rm}))}$$

On définit la variable indicatrice  $q_i$  comme précédemment : elle identifie si l'individu  $i$  est capturé par aucune source ( $q_i=0$ ), par la source BAAC seulement ( $q_i=1$ ), par la source Registre seulement ( $q_i=2$ ) ou par les deux sources ( $q_i=3$ ). En termes d'effectifs, le tableau croisé s'écrit, pour une catégorie  $m$  :

catégorie $m$		Registre (R)		
		oui	non	
BAAC (B)	oui	$N_{3m}$	$N_{1m}$	$N_{Bm}$
	non	$N_{2m}$	$\widehat{N_{0m}}$	
		$N_{Rm}$		$\widehat{N_m}$

Avec

$$N_m = N_{1m} + N_{2m} + N_{3m} \text{ (en termes d'observé)}$$

$$\widehat{N_m} = N_m + \widehat{N_{0m}}$$

avec  $\widehat{N_m}$  estimateur du vrai effectif des blessés, et  $\widehat{N_{0m}}$ , estimateur du vrai nombre de blessés manqués.

On a : 
$$\widehat{N} = \sum_{m=1}^2 \widehat{N}_m$$

De façon identique, on conditionne sur les individus observés ( $q \neq 0$ ) afin d'estimer les paramètres du modèle, pour ensuite pouvoir estimer la taille de la population étudiée. On note toujours  $k=1,2,3$  les sous-ensembles exclusifs « BAAC seulement », « Registre seulement » et « intersection BAAC-Registre ». On note  $p_{ik}$  avec  $k=1, 2, 3$  la probabilité que  $q_i=k$ , sachant que l'individu  $i$  est observé.

Par hypothèse d'homogénéité de capture à l'intérieur de la catégorie  $m$ , les  $p_{ik}$  sont identiques pour tous les individus  $i$  et donc  $p_{ik}$  est remplacé par  $p_{km}$ . On a les probabilités conditionnelles suivantes :

$$p_{1m} = P(\text{être dans BAAC seulement}) / P(\text{être observé})$$

$$p_{2m} = P(\text{être dans Registre seulement}) / P(\text{être observé})$$

$$p_{3m} = P(\text{être dans l'intersection BAAC - Registre}) / P(\text{être observé})$$

Cela s'écrit, pour la catégorie  $m$  :

catégorie $m$		Registre (R)		
		oui	non	
BAAC (B)	oui	$p_{3m}$	$p_{1m}$	$\Pi_B$
	non	$p_{2m}$	$\Pi_{0m}$	
		$\Pi_R$		

Pour tout individu  $i$  appartenant à la catégorie  $m$ , on a, par hypothèse d'indépendance des 2 sources, à catégorie  $m$  fixée :

$$p_{1m} = \frac{\Pi_{Bm}(1 - \Pi_{Rm})}{\Pi_{Bm}(1 - \Pi_{Rm}) + (1 - \Pi_{Bm})\Pi_{Rm} + \Pi_{Bm}\Pi_{Rm}} = \frac{\exp(l_B + k_{Bm})}{\exp(l_B + k_{Bm}) + \exp(l_R + k_{Rm}) + \exp(l_B + k_{Bm} + l_R + k_{Rm})}$$

De la même façon :

$$p_{2m} = \frac{\exp(l_R + k_{Rm})}{\exp(l_B + k_{Bm}) + \exp(l_R + k_{Rm}) + \exp(l_B + k_{Bm} + l_R + k_{Rm})}$$

$$p_{3m} = \frac{\exp(l_B + k_{Bm} + l_R + k_{Rm})}{\exp(l_B + k_{Bm}) + \exp(l_R + k_{Rm}) + \exp(l_B + k_{Bm} + l_R + k_{Rm})}$$

Ces probabilités sont aussi formalisées par un modèle multinomial généralisé, qui s'étend sans problème à l'inclusion d'une variable X :

$$\text{logit}(p_{ki}) = \log\left(\frac{p_{ki}}{p_{1i}}\right) = \beta_k + \theta_k X_i, \text{ avec } k=2,3, \text{ et } k=1 \text{ la catégorie de référence } (\beta_1=0, \text{ et } \theta_1=0)$$

Ou, si l'on inclut une variable catégorielle, telle que la gravité BAAC en 2 catégories ( $m=1, 2$ ), on écrit, pour tout individu  $i$  appartenant à la catégorie  $m$  :

$$\text{logit}(p_{km}) = \log\left(\frac{p_{km}}{p_{1m}}\right) = \beta_k + \theta_{km}, \text{ avec } k=2,3, \text{ et } k=1 \text{ la catégorie de référence } (\beta_1=0, \text{ et } \theta_{1m}=0)$$

En ajustant le modèle, on va obtenir une estimation des paramètres  $\beta_k$  et  $\theta_{km}$ .

Or, ils sont reliés aux paramètres  $l$  et  $k$  de la façon suivante :

$$\begin{aligned} \beta_2 + \theta_{2m} &= \log\left(\frac{p_{2m}}{p_{1m}}\right) \\ &= \log\left(\frac{\exp(l_R + k_{Rm}) / (\exp(l_B + k_{Bm}) + \exp(l_R + k_{Rm}) + \exp(l_B + k_{Bm} + l_R + k_{Rm}))}{\exp(l_B + k_{Bm}) / (\exp(l_B + k_{Bm}) + \exp(l_R + k_{Rm}) + \exp(l_B + k_{Bm} + l_R + k_{Rm}))}\right) \\ &= \log\left(\frac{\exp(l_R + k_{Rm})}{\exp(l_B + k_{Bm})}\right) \\ &= (l_R + k_{Rm}) - (l_B + k_{Bm}) \end{aligned}$$

$$\begin{aligned} \beta_3 + \theta_{3m} &= \log\left(\frac{p_{3m}}{p_{1m}}\right) \\ &= \log\left(\frac{\exp(l_B + k_{Bm} + l_R + k_{Rm}) / (\exp(l_B + k_{Bm}) + \exp(l_R + k_{Rm}) + \exp(l_B + k_{Bm} + l_R + k_{Rm}))}{\exp(l_B + k_{Bm}) / (\exp(l_B + k_{Bm}) + \exp(l_R + k_{Rm}) + \exp(l_B + k_{Bm} + l_R + k_{Rm}))}\right) \\ &= \log\left(\frac{\exp(l_B + k_{Bm} + l_R + k_{Rm})}{\exp(l_B + k_{Bm})}\right) \\ &= l_R + k_{Rm} \end{aligned}$$

Pour la catégorie  $m$  de référence, on a  $\beta_1=0$ , et  $\theta_{1m}=0$ , et donc on en déduit :

$$\beta_2 = l_R - l_B$$

$$\beta_3 = l_R$$

et ensuite :

$$\theta_{2m} = k_{Rm} - k_{Bm}$$

$$\theta_{3m} = k_{Rm}$$

et donc, réciproquement :

$$l_B = \beta_3 - \beta_2$$

$$l_R = \beta_3$$

$$k_{Bm} = \theta_{3m} - \theta_{2m}$$

$$k_{Rm} = \theta_{3m}$$

On ajuste le modèle, on remplace alors les paramètres  $\beta$  et  $\theta$  par leurs estimations, et on en déduit une estimation des  $\Pi_{0m}$  :

$$\begin{aligned}\hat{\Pi}_{0m} &= \frac{1}{(1 + \exp(\hat{I}_B + \hat{k}_{Bm})) \times (1 + \exp(\hat{I}_R + \hat{k}_{Rm}))} \\ &= \frac{1}{(1 + \exp(\hat{\beta}_3 - \hat{\beta}_2 + \hat{\theta}_{3m} - \hat{\theta}_{2m})) \times (1 + \exp(\hat{\beta}_3 + \hat{\theta}_{3m}))}\end{aligned}$$

On peut alors estimer le nombre total par catégorie  $m$ ,  $\hat{N}_m$ .

En effet :

$$\begin{aligned}\hat{\Pi}_{0m} &= \frac{\hat{N}_{0m}}{\hat{N}_m} \Leftrightarrow \\ \hat{\Pi}_{0m} &= \frac{\hat{N}_m - (N_{1m} + N_{2m} + N_{3m})}{\hat{N}_m} \Leftrightarrow \\ 1 - \hat{\Pi}_{0m} &= \frac{(N_{1m} + N_{2m} + N_{3m})}{\hat{N}_m} \Leftrightarrow \\ \hat{N}_m &= \frac{(N_{1m} + N_{2m} + N_{3m})}{1 - \hat{\Pi}_{0m}}\end{aligned}$$

On peut en déduire le nombre total, en sommant sur les catégories  $m$  :

$$\hat{N} = \sum_{m=1}^2 \hat{N}_m$$

On estime aussi les probabilités non conditionnelles aux observations, d'être enregistré dans les BAAC ou dans le Registre :

$$\begin{aligned}\hat{\Pi}_{Bm} &= \frac{1}{1 + \exp(-\hat{I}_B - \hat{k}_{Bm})} = \frac{1}{1 + \exp(\hat{\beta}_2 - \hat{\beta}_3 + \hat{\theta}_{2m} - \hat{\theta}_{3m})} \text{ et} \\ \hat{\Pi}_{Rm} &= \frac{1}{1 + \exp(-\hat{I}_R - \hat{k}_{Rm})} = \frac{1}{1 + \exp(-\hat{\beta}_3 - \hat{\theta}_{3m})}\end{aligned}$$

En prenant l'inverse de la probabilité non conditionnelle d'être enregistré dans les BAAC, on obtient le coefficient de correction du sous-enregistrement des BAAC, pour la catégorie  $m$  :

$$\hat{C}_{Bm} = \frac{1}{\hat{\Pi}_{Bm}} = 1 + \exp(\hat{\beta}_2 - \hat{\beta}_3 + \hat{\theta}_{2m} - \hat{\theta}_{3m}) = 1 + \exp(\hat{\beta}_2 + \hat{\theta}_{2m} - \hat{\beta}_3 - \hat{\theta}_{3m}),$$

On peut aussi estimer  $C_{Rm}$ , coefficient de correction du sous-enregistrement du Registre.

$$\hat{C}_{Rm} = \frac{1}{\hat{\Pi}_{Rm}} = 1 + \exp(-\hat{\beta}_3 - \hat{\theta}_{3m}),$$

Enfin, on peut estimer les coefficients de correction du sous-enregistrement des données de l'intersection BAAC du Rhône et Registre :

$$\hat{C}_3 = \frac{1}{\hat{p}_3} = \left[ 1 + \exp(\hat{\beta}_2 + \hat{\theta}_{2m}) + \exp(\hat{\beta}_3 + \hat{\theta}_{3m}) \right] / \exp(\hat{\beta}_3 + \hat{\theta}_{3m})$$



On généralise sans problème à plusieurs caractéristiques liées aux probabilités d'enregistrement. C'est-à-dire, si la probabilité d'enregistrement est fonction de plusieurs variables (gravité, type d'utilisateur, absence ou présence de tiers, ...), on peut (et on doit) prendre cela en compte. Le modèle multinomial généralisé s'écrit alors, avec l'inclusion de variables  $X_s$  (qui peuvent être continues ou catégorielles).

$$\text{logit}(p_{ki}) = \log\left(\frac{p_{ki}}{p_{li}}\right) = \beta_k + \sum_s \theta_{ks} X_{is} \quad \text{pour } k=2,3, \text{ avec } k=1 \text{ la source de référence.}$$

NB : les coefficients de correction du sous-enregistrement sont ici indicés par  $m$  ;  $m$  identifie les catégories définies par la combinaison des variables incluses dans le modèle. Dans la suite du document, ils sont indifféremment indicés par  $m$  ou  $j$ , la signification ne changeant pas : cela indice les variables qui jouent sur la probabilité d'enregistrement des blessés dans les BAAC

Nous utilisons le logiciel SAS, version 9.4, procédure LOGISTIC et l'option glogit.

## Annexe 3 : Méthodes de redressement, dans le cadre de la théorie des sondages

Nous rappelons ici ce qu'est le redressement dans le cadre de la théorie des sondages (Ardilly 1994). En général, on cherche à redresser un échantillon lorsque l'échantillon obtenu par plan de sondage n'est pas assurément représentatif de la population cible. Le redressement se définit comme une amélioration de l'estimation (dans une étude avec un plan de sondage) en utilisant de l'information auxiliaire.

L'idée est d'utiliser une information auxiliaire X corrélée à la variable d'intérêt Y. On doit avoir suffisamment d'informations sur X pour pouvoir redresser l'échantillon (avec un système de poids) de telle sorte qu'il soit représentatif de la répartition de X dans la population cible. La corrélation de X avec Y conduit à ce que l'échantillon ainsi redressé donne une meilleure estimation de Y.

Trois méthodes de redressement sont définies :

- redressement par post-stratification
- redressement par le ratio
- redressement par régression

### 1. Redressement par post-stratification :

Il s'agit d'un redressement sur une ou plusieurs variables auxiliaires, catégorielles ou rendues catégorielles. Celles-ci forment des post-strates, par opposition aux variables de stratification utilisées dans le plan de sondage. L'information auxiliaire se résume à la connaissance de la répartition dans la population générale de ces post-strates (et non de la répartition dans l'échantillon).

Au lieu de l'estimateur brut, on prend pour estimateur une moyenne pondérée des estimateurs par strate, où le poids est l'effectif relatif de la strate dans la population.

Ex : estimer le revenu moyen de la population française. Le revenu est corrélé avec l'âge ; on redresse donc l'estimation en utilisant la répartition des tranches d'âge dans la population française, et non celle de l'échantillon, qui peut par hasard ne pas être représentative de la population.

$$\hat{Y}_R = \sum_{h=1}^H w_h \times \bar{y}_h = \sum_{h=1}^H \frac{N_h}{N} \times \bar{y}_h,$$

avec h=identifiant des post-strates,  $w_h$  le nouveau poids

Post-stratification en fonction de plusieurs variables auxiliaires :

Si on dispose des effectifs pour toutes les strates définies par le croisement de toutes les modalités des variables auxiliaires : même méthode que celle au-dessus.

Si on ne dispose des informations que sur les marges, c'est-à-dire pour une variable auxiliaire à la fois, on utilise une procédure de calage sur les marges (Deville and Sarndal, 1992; Sautory, 1993), qui ressemble à la post-stratification sur une variable, mais avec des itérations pour tenir compte de plusieurs variables auxiliaires.

### 2. Redressement par le ratio ou par le quotient

Pour améliorer l'estimation, on utilise une variable auxiliaire (quantitative) qui a une relation de proportionnalité avec la variable d'intérêt.

On suppose :  $Y_i = R \times X_i + U_i$ , avec X variable auxiliaire, et  $U_i$  résidus de somme nulle et donc l'estimateur redressé s'écrit :

$$\hat{Y} = \bar{y} \times \frac{\bar{X}}{\bar{x}}, \text{ avec } R \approx \frac{\bar{y}}{\bar{x}}, \text{ estimé sur l'échantillon}$$

pondération (des observations) :

$$w_i(s) = \frac{1}{n} \times \frac{\bar{X}}{\bar{x}}$$

Cela signifie que l'on redresse par un coefficient correcteur ; ce coefficient est égal au ratio entre la valeur moyenne de l'information auxiliaire dans la population cible et celle dans l'échantillon.

### 3. Redressement par régression

Pour améliorer l'estimation, on utilise une variable auxiliaire (quantitative) qui est en relation linéaire avec la variable d'intérêt.

On suppose :  $Y_i = a + b \times X_i + U_i$ , avec X variable auxiliaire, et  $U_i$  résidus de somme nulle

et donc l'estimateur redressé s'écrit :  $\hat{Y} = \bar{y} + \hat{b} \times (\bar{X} - \bar{x})$ .

On redresse les observations récoltées au niveau de l'échantillon par une régression.

Il est facile d'étendre le redressement par régression à plusieurs variables auxiliaires.

## Annexe 4 : Synthèse bibliographique sur la projection, du régional au national, en épidémiologie

Ce qui nous intéresse, c'est l'extrapolation d'indicateurs de santé estimés au niveau local ou régional à une zone géographique plus large, régionale ou nationale. Nous n'avons pas considéré les études avec un plan de sondage sur tout un territoire ; pour celles-ci, il s'agit essentiellement de faire de l'inférence, en tenant compte du plan de sondage.

Par indicateurs de santé, nous entendons en particulier l'incidence, c'est-à-dire le nombre de nouveaux cas sur une année en général ; le taux d'incidence est le nombre de nouveaux cas rapportés à la population exposée, qui peut être la population générale. Par exemple, le taux d'incidence de cancers, est le nombre de personnes nouvellement diagnostiquées comme atteintes d'un cancer, sur une année, rapportée à la population générale. Pour les accidents de la route, le taux d'incidence de blessés de la route est le nombre de blessés de la route, sur une année donnée, rapporté à la population générale.

Nous avons trouvé des publications en épidémiologie des addictions (drogues), du cancer, des traumatismes... et nous avons pu classer ces études, selon les méthodes implicites ou explicites utilisées. Nous présentons ci-dessus les méthodes que nous avons dégagées, avec les exemples correspondants.

### Extrapolation brute

Le taux d'incidence observé sur une région est utilisé comme taux d'incidence du pays. Précisément, l'incidence observée sur une région est appliquée à (l'effectif de) la population nationale, et on obtient ainsi le nombre de cas incidents au niveau national. Cela correspond à la règle de trois : on a trois effectifs, qui sont le nombre de cas incidents sur une région, la population de la région, la population nationale, et par « règle de trois » on obtient le nombre de cas incidents de la population nationale.

Exemple :

Estimation du nombre d'accidents professionnels mortels dans le monde, à partir d'incidences observées dans certains pays, puis appliquée aux autres pays, selon la similarité entre pays (Takala, 1999)

Hypothèse sous-jacente :

Le taux d'incidence observé dans une région, est supposé être le même que celui de la région d'application (ou du territoire tout entier).

### Extrapolation en ajustant sur des facteurs de confusion

- a) Extrapolation en ajustant sur catégories d'âge et de sexe, ou autres, définis au niveau individuel

Principe :

L'incidence est observée sur une région, par strates de sexe et d'âge, et appliquée, par strates de sexe et d'âge, à la population générale.

Exemples :

- estimation du nombre de personnes avec fractures distales de l'avant-bras sur l'ensemble de la Nouvelle-Zélande, à partir de l'incidence observée, par sexe et groupe d'âge, sur une ville majeure, Dunedin, 111000 habitants (Jones et al., 2000)

- estimation de la prévalence de glaucome en chine : application de la prévalence observée en Mongolie, à la Chine rurale, et application de la prévalence observée à Singapour, à la Chine urbaine, par sexe et groupe d'âge (Foster and Johnson, 2001).

Cela revient à pondérer les taux d'incidence par les effectifs par strate d'âge et de sexe, i.e. à de la post-stratification (cf. annexe) sur âge et sexe ; il s'agit aussi de standardisation indirecte, selon la terminologie en épidémiologie, où le nombre attendu de cas incidents de la population étudiée est estimé en appliquant le taux d'incidence, observé sur une population de référence, à la distribution d'âge et de sexe de la population étudiée.

Hypothèse sous-jacente :

Le taux d'incidence dans une région, par strate d'âge et de sexe, est le même, à l'intérieur de chaque strate, que celui de la région d'application (ou du territoire tout entier).

b) Extrapolation en ajustant sur des facteurs de confusion, définis à un niveau agrégé

Principe :

Une relation entre l'incidence d'une maladie et des caractéristiques régionales est observée et modélisée, et on applique ensuite le modèle obtenu à la population générale.

Exemples :

- estimation de l'incidence de cancer sur tous les États-Unis à partir de la relation entre l'incidence de cancer et des variables socio-économiques et ethniques ; relation établie pour certains États, puis appliquée aux autres États (Frey et al., 1994)

- estimation de la prévalence de broncho-pneumopathie chronique obstructive (COPD), en fonction de sa relation avec la structure démographique, la proportion de fumeurs, la proportion d'urbains, le développement économique, la pollution environnementale... : relation observée et modélisée dans certains pays, puis appliquée à d'autres pays (Peabody et al., 2005).

Détail de la méthode (avec variables définies à un niveau agrégé) :

Étape 1 : obtention, sur plusieurs régions, de l'incidence observée et de variables corrélées agrégées et observées ;

Étape 2 : modélisation de l'incidence (observée) en fonction des variables régionales observées ; la modélisation est de type modèle linéaire généralisé ;

Étape 3 : application du modèle obtenu aux autres régions, où les mêmes variables régionales sont observées, afin d'obtenir une estimation du nombre des incidents sur ces régions ; en sommant sur les régions, on obtient une estimation nationale.

Cela correspond à la méthode de redressement par régression (cf. annexe), où les États ou pays constituent les individus de l'échantillon.

Hypothèse sous-jacente :

La relation entre le taux d'incidence et certaines caractéristiques macroscopiques (démographiques, économiques, ...) est supposée être la même dans la région d'observation et dans la région d'application (ou dans le territoire tout entier).

## **Extrapolation à partir d'une morbidité corrélée ou de la mortalité associée**

### a) Extrapolation en fonction d'un indicateur similaire, et d'éventuels facteurs de confusion, définis à un niveau individuel

Principe :

L'incidence d'intérêt est modélisée par une relation proportionnelle à l'incidence d'une morbidité associée mieux connue, ou au taux de mortalité associé à la morbidité d'intérêt, sur une (des) région(s) où les deux sont observées. Extrapolation aux autres régions où seule l'incidence de la morbidité associée ou le taux de mortalité est observé. (Il y a éventuellement stratification sur âge et sexe).

Exemples :

- estimation de la prévalence de toxicomanes consommateurs de cocaïne à partir de la prévalence observée d'héroïnomanes : relation observée et modélisée sur Londres, puis appliquée à l'ensemble du Royaume-Uni (Gossop et al., 1994)

- estimation de l'incidence des cancers (par site anatomique) en fonction du taux de mortalité de ces cancers, par strate d'âge et de sexe ; pour la France, relation entre incidence et mortalité observée et modélisée dans les départements avec registres de cancer, puis appliquée à toute la France (Jensen et al., 1990).

En épidémiologie des traumatismes :

- estimation du nombre de blessés de la route traités dans les services d'urgences, proportionnellement au nombre de blessés de la route hospitalisés ; relation observée sur deux groupes d'hôpitaux de l'État de Caroline du Nord (au nombre de 10 et 15 hôpitaux) et appliquée à tout l'État (Stutts et al., 1988).

- estimation du nombre de blessés de la vie courante traités dans les services d'urgences, proportionnellement au nombre d'hospitalisés pour traumatisme ; relation observée sur trois hôpitaux du réseau EPAC, Enquête permanente sur les accidents de la vie courante (Bordeaux, Annecy, Béthune) et appliquée à la France entière (Nachbaur et al., 2004).

Détail de la méthode utilisant l'incidence d'une morbidité associée :

Étape 1 : obtention, sur une (ou des) région(s), de l'incidence d'intérêt et de l'incidence d'une morbidité similaire et corrélée ;

Étape 2 : modélisation de l'incidence d'intérêt (observée) en fonction de l'incidence (observée) de cette morbidité similaire ; la modélisation est de type proportionnelle ; et éventuellement par strate d'âge et sexe.

Étape 3 : application du modèle obtenu aux autres régions où l'incidence de cette morbidité associée est observée, afin d'obtenir une estimation du nombre des cas incidents pour la morbidité d'intérêt sur ces régions.

Cela correspond à un redressement par le ratio (cf. annexe), et éventuellement par post-stratification.

Hypothèse sous-jacente :

La relation observée et modélisée (entre l'incidence d'intérêt et une morbidité associée) sur la région d'observation est supposée être la même dans les régions d'application.

### b) Extrapolation en fonction d'un indicateur similaire et de caractéristiques macroscopiques

Principe :

L'incidence d'intérêt est modélisée par une relation à partir de l'incidence d'une morbidité associée mieux connue ou de la mortalité associée, et de caractéristiques générales sur une (des) région(s) où toutes sont observées. La relation modélisée est appliquée aux autres régions où seule l'incidence de la morbidité associée et les caractéristiques générales sont observées. La modélisation est de type modèle linéaire généralisé.

Exemples :

- estimation de l'incidence des cancers (par site anatomique) en fonction du taux de mortalité de ces cancers, de l'âge, et de l'année, en France (avec étape supplémentaire de lissage des données de mortalité) ; relation entre incidence et mortalité observée et modélisée dans les départements avec registres de cancer puis appliquée à toute la France (Menegoz et al., 1997).

- estimation du nombre de toxicomanes en fonction du nombre de décès liés à l'usage de drogue, du nombre de toxicomanes admis en traitement médical, du nombre de toxicomanes verbalisés par la police, de la quantité de drogue saisie, et du nombre de séropositifs par injection de drogue ; relation observée sur quatre régions (parmi une quinzaine du Royaume-Uni) et appliquée à l'ensemble du Royaume-Uni (Frischer et al., 2001).

Cela correspond au redressement par régression (cf. annexe), où les régions ou départements ou villes constituent les individus de l'« échantillon ».

Hypothèse sous-jacente :

La relation observée et modélisée sur la région d'observation est supposée être la même dans les régions d'application.

Enfin, nous avons rencontré un exemple un peu différent mais plus proche de notre situation : le nombre de blessés de la route est estimé par 6 hôpitaux en Grande-Bretagne et comparé aux données policières du même lieu, en stratifiant sur type d'usager et gravité. Cette comparaison conduit à estimer des coefficients de corrections des données policières (à appliquer pour atteindre les chiffres obtenus par les hôpitaux). Ces coefficients de correction sont ensuite appliqués aux données policières de l'ensemble de la Grande-Bretagne (James, 1991).

Cet exemple ressemble à notre situation. L'article indiquait que ces 6 hôpitaux n'étaient pas forcément représentatifs, et que le résultat donnait juste une indication de l'ampleur réelle de la morbidité routière. On peut analyser cette approche comme étant un redressement par post-stratification, et à l'intérieur de chaque strate, d'un redressement par le ratio.

## Annexe 5 - Le cas particulier de la variable hospitalisé (oui/non)

On a vu que dans les deux étapes principales (prédiction du nombre de MAIS3+ parmi les BAAC et correction du sous-enregistrement dans les BAAC) la variable hospitalisé (oui/non) des BAAC jouait un rôle important. En effet, dans le modèle de prédiction du nombre de MAIS3+, cette variable est, comme attendu, positivement corrélée au MAIS3+ (Tableau 7). Chez les blessés classés non-hospitalisés 3% sont MAIS3+ et, chez les blessés classés hospitalisés 40% sont MAIS3+.

Dans le modèle de correction du sous-enregistrement elle joue fortement sur la probabilité d'un blessé d'être enregistré dans les BAAC. Par exemple, en zone police, les blessés classés non-hospitalisés ont deux fois moins de chances d'être enregistrés dans les BAAC, que les blessés classés hospitalisés (Tableau 16).

Il s'agit dans ces deux cas, de la variable hospitalisé (oui/non) des BAAC (et non celle du Registre) puisqu'il s'agit d'appliquer le modèle de prédiction du MAIS3+ et de la probabilité d'enregistrement aux seules données BAAC. Cependant, cette variable n'est pas fiable à 100%.

En effet, lors du chaînage entre les BAAC et le Registre du Rhône, on peut remarquer que pour 18% des blessés, la variable hospitalisation des BAAC et la variable hospitalisation du Registre ne correspondent pas. Le Registre est la source la plus fiable pour l'hospitalisation, et nous considérons ainsi la variable hospitalisation (oui/non) du Registre comme le Gold Standard.

**Tableau 64 : Effectifs annuels moyens des blessés selon le critère d'hospitalisation (oui/non) d'après les BAAC et le Registre dans le Rhône sur les blessés communs aux deux sources, 2006-2012**

		Registre		
		Hospitalisé	Non hospitalisé	Total
BAAC	Hospitalisé	429	188	617
	Non hospitalisé	120	1 038	1 158
Total		549	1226	1 775

Ces effectifs permettent d'estimer la probabilité que les BAAC indiquent qu'un blessé est hospitalisé sachant qu'il est hospitalisé d'après le Registre (sensibilité) et la probabilité que les BAAC indiquent qu'un blessé n'est pas hospitalisé sachant qu'il n'a pas été hospitalisé d'après le Registre (spécificité). On peut également estimer la probabilité qu'un blessé est réellement hospitalisé sachant qu'il a été classé hospitalisé par les BAAC (valeur prédictive positive) et estimer la probabilité qu'un blessé est réellement non hospitalisé sachant qu'il a été classé non hospitalisé par les BAAC (valeur prédictive négative).

**Tableau 65 : Critère de validation de la comparaison des blessés BAAC et Registre selon le critère d'hospitalisation**

Pourcentage d'accord	Sensibilité	Spécificité	Valeur prédictive positive	Valeur prédictive négative
82,65	78,20	84,64	69,50	89,67

On observe une sensibilité et une valeur prédictive positive relativement éloignées du 100%.

Ces discordances entre les deux variables (hospitalisé oui/non) des BAAC et du Registre impliquent que l'estimation nationale du nombre de blessés hospitalisés, possible dans l'approche en cours, n'est cependant pas fiable.

En même temps, cela n'enlève en rien l'utilité et la pertinence de cette variable dans les deux étapes clés de notre extrapolation.



## Annexe 6 : Résumé des différentes approches testées et critères de sélection et de validation de l'approche.

Cette annexe présente les résultats d'estimations du nombre de blessés graves de la route selon les différentes approches que nous avons exploré, et comment nous avons choisi celle retenue.

Il s'agissait d'abord de choisir dans quel ordre effectuer les deux principales étapes :

- Capture-recapture pour corriger du sous-enregistrement et des biais associés
- Prédire la proportion de MAIS3+ chez les blessés des BAAC.

Dans la méthode capture-recapture, pour corriger le sous-enregistrement, il faut tenir compte (entre autres) de la gravité, car elle joue beaucoup sur la probabilité d'un blessé d'être enregistré dans les BAAC. Pour la gravité, deux critères sont envisageables : le critère MAIS et l'hospitalisation (Oui/Non). Soit nous prenons la variable hospitalisé (oui/non) qui existe dans les deux bases de données mais qui n'est pas très fiable, soit nous effectuons capture-recapture en fonction de la gravité MAIS et, dans ce cas il faut la prédire sur les données BAAC. Cela signifie aussi, que capture-recapture se base sur une variable en partie prédite, ce qui peut s'avérer plus fragile.

Une dernière possibilité est d'effectuer capture-recapture selon, à la fois la variable hospitalisé (oui/non) et la gravité MAIS ; c'est l'approche retenue. Nous craignons des soucis de colinéarité, ce qui n'a pas été le cas.

Le tableau ci-après donne les effectifs estimés, de blessés toutes gravités MAIS1+, de blessés MAIS3+, et d'hospitalisés. En jaune est surlignée l'approche retenue.

*Effectifs estimés de blessés MAIS1+, de blessés MAIS3+ et de blessés hospitalisés pour chacune des approches (2012)*

N°	ÉTAPE 1	ÉTAPE 2	MAIS1+	MAIS3+	Hospitalisés	Ratio MAIS3+/Hospitalisés
A	CRC selon l'hospitalisation (oui/non)	Prédiction de P(MAIS3+)	326 000	35 600	48 600	73,2
B	Prédiction de P(MAIS3+)	CRC selon le MAIS prédit	312 000	22 000	101 600	21,6
C	Prédiction de P(MAIS3+)	CRC selon l'hospitalisation et le MAIS prédit	307 910	24 500	51 200	47,8
D	Prédiction de P(MAIS3+)	CRC selon l'hospitalisation (oui/non)	307 000	35 700	48 600	73,4

CRC = capture-recapture

- Sensibilité des estimations

On note que d'une approche à l'autre, l'ordre de grandeur du nombre de MAIS1+ estimé reste le même autour de 300 000 - 320 000. Il reste aussi conforme à l'ordre grandeur des estimations de blessés toutes gravité estimés par l'Enquête Nationale Transports et déplacements, ou des données des services d'incendie et de Secours (SDIS)

En revanche, Le nombre de MAIS3+ estimé est soit autour de 22-24 000, soit autour de 35 000, ce qui est moins négligeable comme variation.

Enfin, le nombre d'hospitalisés estimés varie du simple ou double selon l'approche. Cela n'est pas très satisfaisant. Cependant, cela s'ajoute ou correspond au souci de fiabilité de cette variable. Pour rappel

lorsqu'on croise cette variable entre les BAAC et le Registre sur les blessés communs, on note environ 20% de discordance sur cette variable.

- Proportion de MAIS3+/MAIS1+ chez les blessés non observés.

Capture-recapture estime le nombre de blessés qui ne sont ni dans le Registre ni dans les BAAC (au niveau du Rhône). Dans ce sous-ensemble, on s'attend, à ce que la proportion de MAIS3+ parmi les MAIS1+ soit plus faible que chez les blessés répertoriés dans le Registre ou dans les BAAC (en termes ensemblistes : BAAC Union Registre). En effet on sait que ce sont surtout les blessés légers qui font défaut dans les BAAC.

Le tableau ci-dessous indique (colonne « Non observé »), cette proportion de MAIS3+ qu'on obtient dans ce sous-ensemble, selon les différentes approches. On compare cette proportion à celle de BAAC Union Registre (blessés enregistrés dans les BAAC et/ou le Registre. Ainsi, les approches A et D donnent respectivement des proportions de 17% et 24% de MAIS3+ parmi les MAIS1+, bien plus élevée que dans BAAC Union Registre autour de 7-8%. Nous éliminons donc ces deux approches.

*Proportion, en pourcentage, de MAIS3+ parmi les MAIS1+ (2006-2012)*

N°	ÉTAPE 1	ÉTAPE 2	France		Rhône	
			Extrapolé	Extrapolé	Registre U BAAC	Non observé
A	CRC selon l'hospitalisation (oui/non)	Prédiction de P(MAIS3+)	11,4	9,6	7,8	17,8
B	Prédiction de P(MAIS3+)	CRC selon le MAIS prédit	7,3	6,9	7,8	2,2
C	Prédiction de P(MAIS3+)	CRC selon l'hospitalisation et le MAIS prédit	8,3	7,5	7,7	6,2
D	Prédiction de P(MAIS3+)	CRC selon l'hospitalisation (oui/non)	11,9	10,3	7,7	24,5

Nous ne gardons pas l'approche B bien qu'elle donne une proportion estimée de 2,2% de MAIS3+, c'est-à-dire bien en dessous des proportions observées dans BAAC Union Registre, car cette approche se base, dans capture-recapture, seulement sur la variable MAIS (observé mais surtout en partie prédit) pour mesurer la gravité des blessés et nous craignons que cela soit fragile ; nous préférons utiliser capture-recapture avec, pour mesurer la gravité des blessés les 2 variables MAIS (observé ou prédit) et hospitalisé (oui/non). Cela nous semble plus robuste, et nous n'avons pas rencontré de souci de colinéarité (en ayant ces 2 variables dans le modèle). Nous avons ainsi retenu l'approche C.

- Proportion de MAIS3+/MAIS1+ France vs Rhône

Le tableau ci-dessus donne aussi la gravité sur Rhône extrapolé et sur France extrapolé.

Les accidents en milieu rural étant plus graves en moyenne qu'en milieu urbain, et le département du Rhône plus urbain que la France, on s'attend à ce que la proportion de MAIS3+ parmi les MAIS1+ soit plus forte en France que dans le Rhône. C'est le cas des quatre approches ; cependant la différence est faible.

