Les blessés par accidents de la route : estimation de leur nombre et de leur gravité lésionnelle, France, 1996-2004

Emmanuelle Amoros

▶ **To cite this version:**

Emmanuelle Amoros. Les blessés par accidents de la route : estimation de leur nombre et de leur gravité lésionnelle, France, 1996-2004. Sciences du Vivant [q-bio]. Université Claude Bernard - Lyon I, 2007. Français. NNT : . tel-00511718

HAL Id: tel-00511718

https://theses.hal.science/tel-00511718v1

Submitted on 25 Aug 2010

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.

N° d'ordre 272-2007 Année 2007

THÈSE

présentée

devant l'UNIVERSITÉ CLAUDE BERNARD - LYON 1

pour l'obtention

du DIPLÔME DE DOCTORAT

(Arrêté du 7 août 2006)

présentée et soutenue publiquement le

20 décembre 2007

par

Emmanuelle AMOROS

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)

Directeur de thèse : Bernard LAUMON

JURY :	Juliette BLOCH	membre du jury
	Jean BOUYER	rapporteur
	Jean CHAPELON	membre du jury
	Gérard DURU	membre du jury
	Bernard LAUMON	directeur de thèse
	Jean-Louis MARTIN	co-encadrant
	Jean-Christophe THALABARD	rapporteur
	Philippe VANHEMS	membre du jury

UNIVERSITE CLAUDE BERNARD - LYON I

Président de l'Université	**M. le Professeur L. COLLET**
Vice-Président du Conseil Scientifique	M. le Professeur J.F. MORNEX
Vice-Président du Conseil d'Administration	M. le Professeur J. LIETO
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. le Professeur D. SIMON
Secrétaire Général	M. G. GAY

SECTEUR SANTE

Composantes

UFR de Médecine Lyon R.T.H. Laënnec	Directeur : M. le Professeur P. COCHAT
UFR de Médecine Lyon Grange-Blanche	Directeur : M. le Professeur X. MARTIN
UFR de Médecine Lyon-Nord	Directeur : M. le Professeur J. ETIENNE
UFR de Médecine Lyon-Sud	Directeur : M. le Professeur F.N. GILLY
UFR d'Odontologie	Directeur : M. O. ROBIN
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : M. le Professeur F. LOCHER
Institut Techniques de Réadaptation	Directeur : M. le Professeur MATILLON
Département de Formation et Centre de Recherche en Biologie Humaine	Directeur : M. le Professeur P. FARGE

SECTEUR SCIENCES

Composantes

UFR de Physique	Directeur : Mme. le Professeur S. FLECK
UFR de Biologie	Directeur : M. le Professeur H. PINON
UFR de Mécanique	Directeur : M. le Professeur H. BEN HADID
UFR de Génie Electrique et des Procédés	Directeur : M. le Professeur G. CLERC
UFR Sciences de la Terre	Directeur : M. le Professeur P. HANTZPERGUE
UFR de Mathématiques	Directeur : M. le Professeur M. CHAMARIE
UFR d'Informatique	Directeur : M. le Professeur S. AKKOUCHE
UFR de Chimie Biochimie	Directeur : Mme. le Professeur H. PARROT
UFR STAPS	Directeur : M. C. COLLIGNON
Observatoire de Lyon	Directeur : M. le Professeur R. BACON
Institut des Sciences et des Techniques de l'Ingénieur de Lyon	Directeur : M. le Professeur J. LIETO
IUT A	Directeur : M. le Professeur M. C. COULET
IUT B	Directeur : M. le Professeur R. LAMARTINE
Institut de Science Financière et d'Assurances	Directeur : M. le Professeur J.C. AUGROS

Remerciements

Mes remerciements s'adressent avant tout à Bernard Laumon, directeur de l'UMRESTTE et mon directeur de thèse. Je lui suis grandement reconnaissante de m'avoir initiée à l'épidémiologie et notamment à l'épidémiologie de l'insécurité routière, et de m'avoir fait partager son intérêt pour ce domaine. Je le remercie de m'avoir accueillie dans son unité de recherche à diverses étapes de mon parcours et je lui suis reconnaissante de m'avoir confié ce travail de thèse, et en particulier ce sujet. J'ai grandement apprécié son encadrement, ses commentaires, ses conseils, ses réflexions méthodologiques, ses encouragements, son enthousiasme et son humour. Enfin, je lui suis reconnaissante de m'avoir fourni les conditions matérielles et administratives nécessaires au bon déroulement de ce travail.

Je remercie vivement Jean-Louis Martin, qui a d'abord été mon encadrant de stage de l'ISUP, puis mon co-encadrant de thèse. Ses remarques, questions, suggestions, encouragements, m'ont été très utiles. J'ai apprécié ses digressions, son humour et sa disponibilité. Les réunions d'animation scientifique de l'UMRESTTE, qu'il organise, m'ont aussi beaucoup apporté, tant en qualité de participante que d'intervenante à diverses étapes de la thèse.

Je suis reconnaissante à Jean-Christophe Thalabard et Jean Bouyer d'avoir accepté d'être les rapporteurs de mon travail de thèse. Je les remercie vivement pour leurs remarques et commentaires, qui m'ont permis d'améliorer la qualité et la pertinence de cette thèse.
Je remercie Jean Bouyer pour ses cours d'épidémiologie, très clairs, alors que j'étais étudiante à l'ISUP ; ils m'ont accompagnée dans mon parcours.

Je remercie vivement Juliette Bloch, Jean Chapelon, Gérard Duru, et Philippe Vanhems, d'avoir accepté d'être les membres du jury de mon travail de thèse.

Ce travail n'aurait pas été possible sans l'implication de toutes les personnes au sein de l'Association pour le Registre des Victimes d'Accidents de la Circulation du Rhône (ARVAC ; Président Yves Noël Martin) et au sein de l'INRETS-UMRESTTE (Bernard Laumon, en tant que responsable scientifique du Registre et Amina Ndiaye, médecin responsable technique du Registre) qui ont participé au recueil ou à l'informatisation des données :
Ait Idir T, Ait Si Selmi T, Alloatti D, 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, Bec JF, Bejui J, Bel JC, Bérard E, Bérard J, Bernard JC, Berthet N, Bertrand JC, Besson L, Biot B, Biot V, Blanc C, Blanchard J, Bœuf C, Boisson D, Bonjean M, Bost J, Bouchedor C, Bouletreau P, Boyer V, Breda Y, Brilland R, Bussery S, Cabet N, Caillot L, Caillot JL, Cannamela A, Caregnato B, Carre M, Catala Y, Chagnon PY, Chambost M, Chantran C, Chardon P, Charnay P, Chatelain P, Chattard S, Chauvin F, Chavane H, Chazot G, Chettouane I, Chevreton N, Chevrillon E, Chevrillon S, Chiron M, Chotel P, Cochard P, Combe C, Contamin B, Coppard E, Cot T, Crettenet Z, Cristini A, Dal Gobbo B, De Angelis MP, Decourt L, Delfosse A, Demazière J, Deruty R, Desjardins G, Devaux J, Dohin B, Emonet A, Escarment J, Eyssette M, Fallavier L, Fanton L, Felten D, 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, Goubsky A, Granger M, Grattard P, Gueniaud PY, Guenot C, Guérin AC, Guignand M,Guillaumée F, Haddak M, Hamel D, Heckel T, Herzberg G, Jacquemard C, Javouhey E, Joffre T, Kohler R, Lablanche C, Lafont S, Lagier C, Lapierre B, Laplace MC, La Rosa C, Laurent R, Lebel M, Leblay G, Le-Xuan I, Lieutaud T, Lille R, Linné M, Lucas R, Machin B, Maello E, Malicier D, Mangola B, Marduel YN, Marie-Catherine M, Martin JL, Martinand G, Marty F, Mazouzi S, Messikh C, Meyer F, Meyrand S, Molard S, Monneuse O, Morel-Chevillet E, Mioulet E, Minjaud F, Mollet C, Monnet J, Moyen B, Neidhart JP, Ngandu E, Ny S, Ould T, Paget P, Paillot JC, Paris D, Patay B, Pauget P, Peillon D, Perrin G, Perrin-Blondeau D, Petit P, Piton JL, Plantier M, Pornon P, Pramayon C, Quelard B, Raquin L, Rezig M, 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, Sbraire N, Scappaticci N, Schiele P, Schneider M, Simonet C, Sindou M, Soldner R, Soudain M, Stagnara J, Stamm D, Suc B, Taesch MC, Tasseau F, Tell L, Thomas M, Tilhet-Coartet S, Tissot E, Toukou JC, Trifot M, Vallee B, Vallet G, Vancuyck A, Vergnes I, Verney MP, Voiglio EJ, Vourey G, Vuillard J, Westphal M, Willemen L.

Cette étude a pu être réalisée, d'une part, grâce à la mise à disposition de l'INRETS des données policières d'accidentologie par l'ONISR (Observatoire National Interministériel de Sécurité Routière), d'autre part par un financement sur trois ans de la DCSR (Direction de la Sécurité et Circulation Routières), dans le cadre du PREDIT (Programme National de Recherche et d'Innovation dans les Transports Terrestres).

Je remercie les personnes qui, victimes d'un accident de la route, ont accepté de se remémorer les circonstances de leur accident et de répondre à mon questionnaire d'enquête.

J'ai particulièrement apprécié l'aide que m'a apportée Geneviève Boissier, pour l'extraction et la saisie d'informations de quelques 900 procès-verbaux d'accident, l'aide pour l'envoi de 3000 lettres et questionnaires, et pour la saisie des réponses à ces questionnaires.

Je remercie Amina Ndiaye et Mireille Chiron pour avoir bien voulu coder la gravité des lésions de quelques 500 blessés, à partir de certificats médicaux.

Je remercie la directrice technique du Registre Amina Ndiaye, ainsi que Marie-Pierre Verney, Denise Perrin-Blondeau, et Blandine Gadegbeku, pour m'avoir donné des compléments d'informations sur le Registre.

Je remercie Sylviane Lafont pour ses relectures critiques et constructives du mémoire de thèse et des articles, pour ses encouragements, ses conseils et commentaires avisés.

Je remercie Christine Buisson pour ses conseils sur l'écriture d'un mémoire de thèse. J'ai également apprécié l'aide que m'ont apporté Mireille Chiron, Idlir Licaj, Aurélie Moskal, Nicolas Dapzol, Marlène Bernard et Karine Supernant en acceptant de relire tout ou partie de ce mémoire.

Je remercie Patricia Chapuis et Incarnation Montusclat, pour leur aide précieuse sur les aspects matériels, financiers, et administratifs.

Je remercie Lindsay Cant, Pierrette Charnay, Emmanuel Fort, Heikki Jahi, Anne-Céline Guérin, Mouloud Haddak, Martine Hours, Erik Lenguerrand, Nathalie Nallet, Gilles Vallet, Irène Vergnes, qui de différentes manières m'ont apporté aide ou soutien. Je remercie l'ensemble de mes collègues de l'unité UMRESTTE, auprès desquels j'ai trouvé une très bonne ambiance de travail.

Je remercie Vincent, Claude et Marie-Hélène pour leur soutien.

Je ne voudrais pas clore ces remerciements sans évoquer que le sujet d'étude est très concret : il est un drame pour les victimes et pour leurs proches, et il nous concerne tous, que ce soit en tant que personne exposée, ou en tant qu'acteur dans nos déplacements.
J'ose espérer que ce travail, en permettant une meilleure connaissance de l'insécurité routière, et notamment de son ampleur, contribuera à une prévention plus énergique de ce problème de santé publique.

Table des matières

Remerciements 4
Sommaire des annexes 13
Lexique et liste des abréviations 14
INTRODUCTION 15
MATÉRIEL 19
Données des forces de l'ordre 19
Le Registre des victimes d'accidents de la circulation routière 19
Données rapprochées 20
Exclusions 21
Période d'étude 21

Chapitre 1 Biais de sélection des données d'accidentologie 22
1.1 Biais de sélection des données des forces de l'ordre (analyse des blessés recensés par le Registre) 22
Introduction 22
Matériel 23
Méthode 23
Sous-enregistrement 23
Biais de sélection 23
Résultats 25
Discussion 28
Résumé 31
1.2 Enquête auprès des blessés identifiés dans le Registre seulement 32
Introduction 32
Matériel 32
Population d'étude 32
Méthode 33
Objectifs détaillés 33
Construction de l'échantillon (plan de sondage) 33
Recueil des données par questionnaire 35
Test des questionnaires 35
Analyse statistique 36
Résultats 36
Répartition des retours de questionnaires 36
Comparaison des répondants et des non-répondants 37
Blessés s'avérant hors critères BAAC 37
Alerte et intervention des forces de l'ordre 38
Taux de discordance sur les variables de chaînage 40
Discussion 41
Résumé 43
1.3 Étude des blessés identifiés dans les BAAC seulement 44
Introduction 44
Matériel 44
Population d'étude 44
Accès aux procès-verbaux 44
Méthode 44
Caractéristiques étudiées 45
Analyse statistique 46

Chaînage supplémentaire 47
Résultats 47
Population d'étude 47
Comparaison des blessés « BAAC seulement » avec PV exploitable et ceux sans PV exploitable 48
Caractéristiques des blessés « BAAC seulement » 49
Taux d'erreur sur les variables servant au chaînage 52
Chaînage complémentaire sur les noms et prénoms 52
Discussion 53
Résumé 56

Chapitre 2 Erreurs de classement de la gravité lésionnelle dans les données des forces de l'ordre 57
Introduction 57
Matériel et Méthode 58
Résultats 60
Erreurs de classement de la gravité 60
Caractéristiques associées aux erreurs de classement de la gravité 61
Caractéristiques associées aux erreurs de sur-classement de la gravité 62
Caractéristiques associées aux erreurs de sous-classement de la gravité 63
Discussion 64
Interprétation des résultats 64
Forces et faiblesses de l'étude 65
Conclusion 66
Résumé 67

Chapitre 3 Bilan rhodanien : nombre de blessés et gravité lésionnelle 68
Introduction 68
Synthèse bibliographique sur capture-recapture 68
Matériel 69
Données policières d'accidentologie 70
Registre des victimes de la circulation routière dans le Rhône 70
Données rapprochées 70
Trois scénarios de chaînage 71
Méthode 74
Capture-recapture 74
Conditions d'application 74
Agrégation des données et variables de stratification 75
Traitement des valeurs manquantes 76
Statistiques présentées 76
Résultats 76
Discussion 79
Conclusion 82
Résumé 83

Chapitre 4 Bilan national : nombre de blessés et gravité lésionnelle 84
Introduction 84
Matériel 87
Méthode 87
États de santé étudiés 87

Étape 1 : Construction et application d'un modèle de prédiction du NISS 88
Étape 2 : Amélioration du résultat du chaînage 96
Étape 3 : Modélisation du sous-enregistrement par capture-recapture 98
Étape 4 : Application des coefficients de correction aux données policières nationales 110
Statistiques présentées : effectifs et incidences 111
Intervalles de confiance 111
Validité des modèles et de l'ensemble de la procédure d'estimation 111
Résultats 113
Blessés toutes gravités 113
Taux de létalité 118
Blessés hospitalisés 121
Blessés graves 125
Taux de gravité (chez les blessés non décédés) 131
Blessés avec séquelles majeures prévues 134
Discussion 140
Forces et faiblesses 140
Interprétation des résultats 144
Résumé 147

CONCLUSION 148
Bibliographie 151

Liste des figures

Figure 1 : principe de la projection 16
Figure 2 : ensemble de la procédure d'estimation du bilan national de l'insécurité routière 17
Figure 3 : schématisation des deux sources d'enregistrements des blessés dans le département du Rhône 23
Figure 4 : répartition du nombre de blessés (non-décédés) selon la source d'enregistrement, Rhône 1997-2001 25
Figure 5 : proportions de blessés classés « légers » et « graves » par les forces de l'ordre, en fonction du NISS (14 342 blessés non-décédés, intersection BAAC - Registre, Rhône, 1997-2001) 60
Figure 6 : répartition des blessés recensés par le Registre et par les BAAC 70
Figure 7: redéfinition de la taille de l'intersection BAAC-Registre 71
Figure 8: schéma de deux recensements (ou « listes ») 74
Figure 9 : principe de la projection 84
Figure 10 : ensemble de la procédure d'estimation du bilan de morbidité par accidents de la circulation routière, France entière 86
Figure 11 : redéfinition de la taille de l'intersection BAAC-Registre 97
Figure 12 : nombre de blessés toutes gravités, France, selon la source des données 114
Figure 13 : évolution du nombre des blessés toutes gravités, France, selon la source des données 114
Figure 14 : nombre de blessés toutes gravités selon le type d'usager, France, source = données des forces de l'ordre 115
Figure 15 : nombre de blessés toutes gravités selon le type d'usager, France, source = résultats de la projection 115
Figure 16 : incidence des blessés toutes gravités selon l'âge, France, source = résultats de la projection 116
Figure 17 : incidence des blessés toutes gravités selon le sexe, France, source = résultats de la projection 117
Figure 18 : taux de létalité, France, selon la source des données 118
Figure 19 : évolution en base 100 du taux de létalité, France, selon la source des données 118
Figure 20 : létalité, selon le type d'usager, France, source = données des forces de l'ordre 119
Figure 21 : létalité, selon le type d'usager, France, source = résultats de la projection 119
Figure 22 : nombre de blessés hospitalisés, France, selon la source des données 122
Figure 23 : évolution en base 100 du nombre de blessés hospitalisés, France, selon la source des données 122
Figure 24 : nombre de blessés hospitalisés prédits selon le type d'usager, France, source = données des forces de l'ordre 123
Figure 25 : nombre de blessés hospitalisés selon le type d'usager, France, source = résultats de la projection 123
Figure 26 : nombre de blessés graves, France, selon la source des données 126
Figure 27 : évolution du nombre de blessés graves, France, selon la source des données 126
Figure 28 : nombre de blessés graves (nécessitant plus de 6 jours d'hospitalisation), selon le type d'usager, France, source = données des forces de l'ordre 127
Figure 29 : nombre de blessés graves (NISS 9+), selon le type d'usager, France, source = résultats de la projection 127
Figure 30 : incidence de blessés graves (NISS 9+) selon l'âge, France, source = résultats de la projection 128
Figure 31 : incidence de blessés graves (NISS 9+) selon le sexe, France, source = résultats de la projection 129
Figure 32 : nombre de blessés (tous, blessés graves et blessés hospitalisés), France, selon la source des données 130
Figure 33 : taux de gravité, France, selon la source des données 132
Figure 34 : évolution en base 100 du taux de gravité, France, selon la source des données 132
Figure 35 : taux de gravité, selon le type d'usager, France, source = données des forces de l'ordre 133
Figure 36 : taux de gravité, selon le type d'usager, France, source = résultats de la projection 133

Figure 37 : effectif de tués à six jours, et de blessés avec séquelles majeures (résultats de la projection), ... 134
Figure 38 : effectifs de tués et de blessés avec séquelles majeures, selon le type d'usager, ... 136
Figure 39 : répartition des blessés avec séquelles majeures, selon la région corporelle atteinte, et le type d'usager, France, 1996-2004, source = résultats de la projection 136
Figure 40 : incidence de tués à six jours, selon l'âge, France, moyenne annuelle sur 1996-2004, ... 137
Figure 41 : incidence de blessés avec séquelles majeures, selon l'âge, France, moyenne annuelle sur 1996-2004, source = résultats de la projection 137
Figure 42 : répartition des blessés avec séquelles majeures, selon la région corporelle et l'âge, France, 1996-2004, source = résultats de la projection 138
Figure 43 : incidence de blessés avec séquelles majeures, selon le sexe, France, moyenne annuelle sur 1996-2004, source = résultats de la projection 139
Figure 44 : répartition des blessés avec séquelles majeures, selon la région corporelle, et le sexe, France, 1996-2004, source = résultats de la projection 139

Liste des tableaux

Tableau 1 : risques relatifs ajustés et probabilités ajustées (standardisées) d'enregistrement dans les données des forces de l'ordre; analyse multivariée (50 936 blessés du Registre, Rhône, 1997-2001) . 26
Tableau 2 : répartition des effectifs observés des blessés selon le type d'usager, l'absence/présence de tiers, et la gravité lésionnelle, Rhône, 2001, Registre seulement. 34
Tableau 3: répartition des effectifs de blessés selon un plan de sondage stratifié et proportionnel, pour l'enquête « Registre seulement »,..... 34
Tableau 4 : répartition des effectifs de blessés selon un plan de sondage stratifié et proportionnel, avec gonflement des effectifs tenant compte du taux de réponse attendu, pour l'enquête « Registre seulement »..... 35
Tableau 5 : répartition des retours de questionnaires, enquête « Registre seulement » 37
Tableau 6 : répartition des réponses à l'enquête « Registre seulement » en fonction des critères BAAC (parmi les 547 questionnaires complétés) 37
Tableau 7 : proportion d'alerte des forces de l'ordre (parmi les 495 questionnaires exploitables de l'enquête « Registre seulement ») 38
Tableau 8 : proportion d'alerte des forces de l'ordre, en fonction du type d'usager, de la présence/absence de tiers et de la gravité lésionnelle (parmi les 495 questionnaires exploitables de l'enquête « Registre seulement ») 38
Tableau 9 : taux de procès-verbaux apparemment établis (parmi les 270 répondants de l'enquête « Registre seulement » avec présence des forces de l'ordre) 39
Tableau 10 : taux de procès-verbaux apparemment établis, en fonction du type d'usager, de la présence/absence de tiers et de la gravité lésionnelle (parmi les 270 répondants de l'enquête « Registre seulement » avec présence des forces de l'ordre) 39
Tableau 11 : taux de procès-verbaux apparemment établis (parmi les 495 questionnaires exploitables de l'enquête « Registre seulement »)..... 40
Tableau 12 : distribution des raisons données pour NE PAS avoir alerté les forces de l'ordre (parmi les 230 répondants dans ce cas, enquête « Registre seulement ») 40
Tableau 13 : discordance entre les informations du Registre et les réponses au questionnaire, enquête « Registre seulement » (parmi les 495 questionnaires exploitables)..... 41
Tableau 14 : répartition des blessés avec ou sans PV exploitable en fonction du mois de l'accident, étude « BAAC seulement » 48
Tableau 15 : répartition des blessés avec ou sans PV exploitable en fonction de la force de l'ordre, étude « BAAC seulement » 48
Tableau 16 : répartition des blessés avec ou sans PV exploitable en fonction de la gravité (BAAC), étude « BAAC seulement » 49
Tableau 17 : répartition des blessés avec ou sans PV exploitable en fonction de la distance lieu d'accident - hôpital, étude « BAAC seulement » 49
Tableau 18 : répartition des 868 personnes avec PV, selon le statut de blessé/non-blessé et selon le recours aux soins (groupes disjoints), étude « BAAC seulement »..... 50
Tableau 19 : répartition des 868 personnes avec PV, selon le recours aux soins, étude « BAAC seulement »..... 50
Tableau 20 : comparaison de la distribution du NISS entre les blessés « BAAC seulement » et les blessés de l'intersection BAAC - Registre, Rhône, 2001..... 51
Tableau 21 : répartition des blessés de l'étude « BAAC seulement » selon le recours aux soins et selon la gravité NISS..... 51
Tableau 22 : comparaison de la distance du lieu d'accident à l'hôpital le plus proche, entre les blessés « BAAC seulement » et les blessés de l'intersection BAAC - Registre, Rhône 2001 52
Tableau 23 : répartition des blessés « BAAC seulement » selon la consultation hospitalière, et la distance du lieu d'accident à l'hôpital le plus proche..... 52
Tableau 24: résultat détaillé du chaînage supplémentaire avec noms et prénoms, Rhône, 2001 (nombre de paires reliées parmi nombre de paires sélectionnées)..... 53
Tableau 25 : distribution des blessés selon les classements policier et médical de la gravité, Rhône, 1997-2001..... 60

Tableau 26 : caractéristiques liées aux erreurs de classement de la gravité par les forces de l'ordre, analyse multivariée sur 14 342 blessés (intersection BAAC - Registre, Rhône, 1997-2001) 61
Tableau 27 : caractéristiques liées au sur-classement de la gravité par les forces de l'ordre, analyse multivariée sur 13 259 blessés NISS 1-15 classés « blessés graves » (intersection BAAC - Registre, Rhône, 1997-2001) 62
Tableau 28 : caractéristiques liées au sous-classement de la gravité par les forces de l'ordre, analyse multivariée sur 1083 blessés NISS 16-75 classés « blessés légers » (intersection BAAC - Registre, Rhône, 1997-2001) 63
Tableau 29 : description d'un résultat de chaînage entre deux fichiers d'enregistrement (unité=paire d'enregistrements) 70
Tableau 30: répartition des effectifs selon l'appartenance des sujets à deux sources d'enregistrement 74
Tableau 31 : résultat du chaînage de routine entre les données des forces de l'ordre et le Registre, Rhône, 2001 77
Tableau 32 : résultats de l'analyse capture-recapture stratifiée, agrégés sur type d'usager et tiers (oui/non), pour chaque scénario, Rhône, 2001 78
Tableau 33 : résultats détaillés de l'analyse capture-recapture stratifiée, scénario n° 3, Rhône, 2001 . 79
Tableau 34 : paramètres du modèle de prédiction P(NISS4+), construit sur 22 704 blessés de l'intersection BAAC-Registre, Rhône, 1996-2004 92
Tableau 35 : paramètres du modèle de prédiction P(NISS 9+), construit sur 22 704 blessés de l'intersection BAAC-Registre, Rhône, 1996-2004 93
Tableau 36 : prédiction du niveau NISS en fonction des profils de blessés des BAAC (profil : sexe masculin, âge 20-29 ans, accident contre voiture, en agglomération, sur voie communale, en 2004).. 94
Tableau 37 : comparaison des NISS prédits et observés, parmi les blessés identifiés dans les BAAC seulement, Rhône, 2001 95
Tableau 38 : paramètres du modèle de capture-recapture, construit sur 93 298 blessés, BAAC et/ou Registre, Rhône, 1996-2004 107
Tableau 39 : coefficients de correction du sous-enregistrement des données des forces de l'ordre pour quelques profils de blessés (accidents sur routes nationales, en zone police, en 2004) 109
Tableau 40 : effectif et incidence des blessés (légers et graves), France, source = données des forces de l'ordre 113
Tableau 41 : effectif et incidence des blessés (NISS 1+), France, source = résultats de la projection 113
Tableau 42 : effectif et incidence prédits des blessés hospitalisés, France, source = données des forces de l'ordre 121
Tableau 43 : effectif et incidence des blessés hospitalisés, France, source = résultats de la projection 121
Tableau 44 : effectif et incidence des blessés graves (nécessitant plus de 6 jours d'hospitalisation), France, source = données des forces de l'ordre 125
Tableau 45 : effectif et incidence des blessés graves (NISS 9+), France, source = résultats de la projection 125
Tableau 46 : taux de gravité (basé sur les hospitalisés plus de 6 jours), France, source = données des forces de l'ordre 131
Tableau 47 : taux de gravité (basé sur le NISS 9+), France, source = résultats de la projection 131
Tableau 48 : effectif et incidence, des tués et des blessés avec séquelles majeures, France, moyennes annuelles sur 1996-2004 134
Tableau 49 : effectif et incidence de tués selon le type d'usager, moyennes annuelles sur 1996-2004, France, source = données des forces de l'ordre 135
Tableau 50 : effectif de blessés avec séquelles majeures, selon le type d'usager, moyennes annuelles sur 1996-2004, France, source = résultats de la projection 135

Sommaire des annexes

Table des matières

Annexes du chapitre 1
Synthèse bibliographique sur le sous-enregistrement des données des forces de l'ordre et les biais de sélection associés 3
Extrait du programme SAS : régression de Poisson modifiée, et calcul des probabilités marginales 11
Article publié dans le journal Accident Analysis and Prevention, version « pré-print » 12
Exemples de lettre et questionnaire envoyés dans le cadre de l'enquête « Registre seulement » 25
Extraits de programme SAS : analyse des réponses au questionnaire en tenant compte du plan de sondage 29

Annexes du chapitre 2
Synthèse bibliographique sur les erreurs de classement de la gravité des blessures par les forces de l'ordre 33
Extrait de programme SAS : régression de Poisson modifiée et calcul des probabilités marginales (erreurs de classement) 39
Article publié dans Journal of Trauma, version « pré-print » 40

Annexes du chapitre 3
Estimation du nombre de faux négatifs et de faux positifs, sur le résultat du chaînage des données de 2001 52
Résultats détaillés de l'analyse capture-recapture stratifiée, scénarios 1, 2 et 3 56
Article d'application de capture-recapture publié dans Accident Analysis and Prevention, version "pré-print" 58

Annexes du chapitre 4
Revue bibliographique sur la méthode de projection 69
Estimation du nombre de faux négatifs et de faux positifs sur le résultat du chaînage des données des années 1996 à 2004. 74
Résultats du modèle de capture-recapture (multinomial logit) ajusté sur 5 facteurs de biais, sur les données du Rhône 1996-2004 78
Extrait de programme SAS : modèle de capture-recapture 80
Traitement des valeurs manquantes : imputation simple 82
Adéquation des modèles de prédiction du NISS en trois catégories : NISS 1-3, NISS 4-8, NISS 9+ 83
Autres critères de santé étudiés : modèles de la procédure d'estimation 84
Détail du calcul sur la région Aquitaine de nombre de traumatisés crâniens graves d'après la littérature 97
Résultats supplémentaires du bilan France métropolitaine 98
Article soumis pour publication (estimation du bilan national) 131

Lexique et liste des abréviations

Lexique

Type d'usager= piétons, cyclistes, usagers de 2RM, automobilistes, autres

Policier (adjectif) = relevant des forces de l'ordre (police, gendarmerie, CRS, police aux frontières et préfecture de police de Paris)

Blessés= blessés non-décédés dans les six jours suivant l'accident

Registre seulement= sous-ensemble du Registre, restreint aux blessés identifiés dans le Registre seulement (c'est-à-dire pas identifiés dans les données policières)
BAAC seulement= sous-ensemble des BAAC, restreint aux blessés identifiées dans les BAAC seulement (c'est-à-dire pas identifiés dans le Registre)
Intersection BAAC-Registre= sous-ensemble formé par les blessés identifiés comme communs aux BAAC et au Registre

Liste des abréviations

2RM=deux-roues motorisé

AIS= Abbreviated Injury Scale
Par abus de langage, AIS= échelle de gravité AIS d'une blessure, allant de 1 à 6 (1=gravité mineure, 2= modérée, 3=sérieuse, 4=sévère, 5=critique, 6=au-delà des ressources thérapeutiques).

BAAC=Bulletin d'Analyse d'Accident Corporel de la circulation routière. Cela signifie : données policières informatisées

IC= Intervalle de confiance

IIS : Injury Impairment Scale : échelle de mesure des séquelles prévues à 1 an, comprise entre 1 et 6.
IIS 3 + : lésion avec séquelle majeure

MAIS= Maximum AIS : score de gravité maximum, de 1 à 6

NISS= New Injury Severity Score : échelle de gravité basée sur les scores AIS, comprise entre 1 et 75
NISS 1-3 : blessé de gravité mineure
NISS 4-8 : blessé de gravité modérée
NISS 9+ : blessé de gravité sérieuse et au-delà
NISS 1+ : blessé au sens AIS

PV= procès-verbal de l'accident (rapport des forces de l'ordre sur l'accident et les personnes impliquées)

OR= odds ratio

RR=risque relatif

INTRODUCTION

Afin de réduire l'insécurité routière, et dans ce but définir les priorités de prévention, il est nécessaire d'en connaître l'ampleur et les caractéristiques. Dans la plupart des pays, les connaissances épidémiologiques sur les victimes d'accident de la circulation routière sont basées sur les données recueillies par les forces de l'ordre. Les tués sont plutôt bien recensés, mais ce n'est pas le cas des blessés (Elvik and Mysen 1999). Le sous-enregistrement des blessés existe dans tous les pays et son ampleur est souvent grande. Cette non-exhaustivité pose la question de la représentativité des données recueillies. Plusieurs études ont mis en évidence des biais de sélection en fonction de la gravité des blessures, du type d'usager et du nombre de véhicules impliqués (un seul versus plusieurs) (James 1991; Hvoslef 1994; Elvik and Mysen 1999). Par ailleurs, les données des forces de l'ordre ne contiennent en général aucune information médicale ; le seul classement policier des blessés en blessés légers et graves ne peut être qu'imprécis. Cette situation générale est valable aussi en France.

Notre but est d'obtenir la meilleure estimation possible du bilan exhaustif de l'insécurité routière, sur la France entière. Pour l'atteindre, nous nous basons sur la coexistence des données des forces de l'ordre au niveau national et du Registre des victimes d'accident de la circulation routière dans le Rhône.

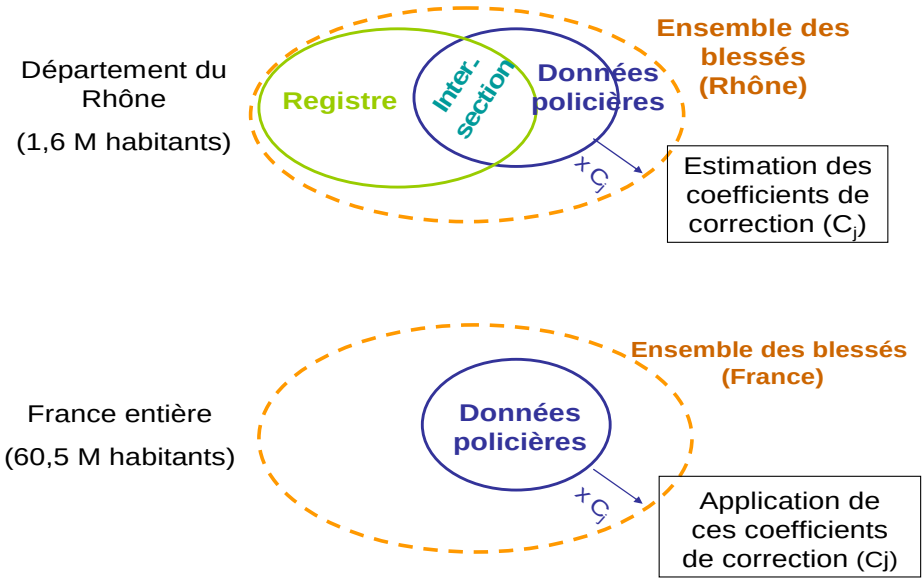
Ce registre a pour critère géographique d'inclusion le département du Rhône comme lieu d'accident. Il est basé sur la participation de l'ensemble des services hospitaliers publics et privés, du Rhône et des proches alentours, susceptibles d'accueillir des victimes de la route. Il recense l'ensemble des blessés traités dans un service hospitalier, qu'ils soient hospitalisés ou non. Le bilan lésionnel de chaque victime est codé selon la classification AIS (Abbreviated Injury Scale), échelle de traumatologie reconnue.

La coexistence des deux sources de recensement des blessés dans le département du Rhône depuis 1996, et l'identification des blessés communs aux deux sources (par chaînage) nous permettent, en prenant le Registre comme référence, de bien cerner les limites des données des forces de l'ordre : les biais de sélection et les erreurs de classement sur la gravité.

Cette coexistence permet aussi, à l'aide de la méthode de capture-recapture et sous certaines conditions, d'estimer le nombre de blessés ayant échappé aux deux sources de recensement. Il est ainsi possible d'estimer le bilan exhaustif de l'insécurité routière sur le département du Rhône.

Il s'agit alors d'estimer le bilan exhaustif au niveau national, par projection, en s'appuyant sur l'existence des données des forces de l'ordre sur l'ensemble du territoire français. L'hypothèse sous-jacente est celle d'homogénéité sur l'ensemble du territoire français des pratiques des forces de l'ordre d'enregistrement des victimes de la circulation routière, par force de l'ordre et par strate ; chaque strate est définie par la combinaison des facteurs de biais de sélection précédemment identifiés.
À partir des informations fournies par la comparaison des deux sources, et grâce à l'approche capture-recapture, nous construisons un modèle permettant de prédire le bilan exhaustif (du Rhône) à partir des données des forces de l'ordre (du Rhône). Nous appliquons ensuite ce modèle aux données nationales des forces de l'ordre pour obtenir une estimation du bilan exhaustif national. Plus précisément, le modèle de capture-recapture nous permet d'estimer au niveau du Rhône des coefficients de correction du sous-enregistrement des données des forces de l'ordre, définis en fonction des facteurs de biais de sélection. Nous appliquons ensuite ces coefficients de correction aux données nationales des forces de l'ordre, en standardisant sur les facteurs de biais de sélection, de façon analogue à la méthode de standardisation indirecte de taux d'incidence, méthode classique en épidémiologie. Le principe de la projection est illustré par la Figure 1.

Figure 1 : principe de la projection



Les deux premiers chapitres de ce mémoire portent sur la comparaison des données des forces de l'ordre avec les données du Registre, au niveau du département du Rhône. Le Registre, bien plus complet et disposant d'informations médicales, est pris comme référence.
Le premier chapitre porte sur les biais de sélection : essentiellement ceux des données des forces de l'ordre mais aussi ceux du Registre. Ce chapitre se compose de trois parties. La première partie est une analyse multivariée des biais de sélection des données des forces de l'ordre. Parmi l'ensemble des blessés recensés par le Registre, nous modélisons la probabilité qu'ils soient également recensés dans les données policières. Nous analysons ainsi le « filtre » d'enregistrement dans les fichiers des forces de l'ordre. Cette partie du mémoire a fait l'objet d'une publication dans la revue Accident Analysis and Prevention.
Cette partie est complétée par une étude des différences de couverture des données des forces de l'ordre et du Registre. En effet, en moyenne sur la période 1996-2004, les données des forces de l'ordre dénombrent 4000 blessés par an, le Registre 9000 blessés, et un chaînage entre les deux sources permet d'identifier 2500 blessés communs. Cela signifie qu'un certain nombre de blessés ne sont identifiés que par une source : 6500 par an dans le Registre seulement et 1500 par an dans les données des forces de l'ordre seulement.
Deux études complémentaires sont donc menées pour mieux comprendre pourquoi. La première enquête « Registre seulement » cherche à quantifier les situations possibles : absence d'alerte des forces de l'ordre par les blessés, ou absence de rédaction d'un procès-verbal par les forces de l'ordre, ou dysfonctionnement du chaînage entre les données des forces de l'ordre et le Registre. La deuxième étude complémentaire porte sur les blessés identifiés dans les données des forces de l'ordre seulement. Elle cherche à quantifier les situations possibles : blessés de gravité mineure telle qu'ils n'ont pas consulté de service hospitalier, ou dysfonctionnement du Registre, ou dysfonctionnement du chaînage entre données des forces de l'ordre et du Registre.

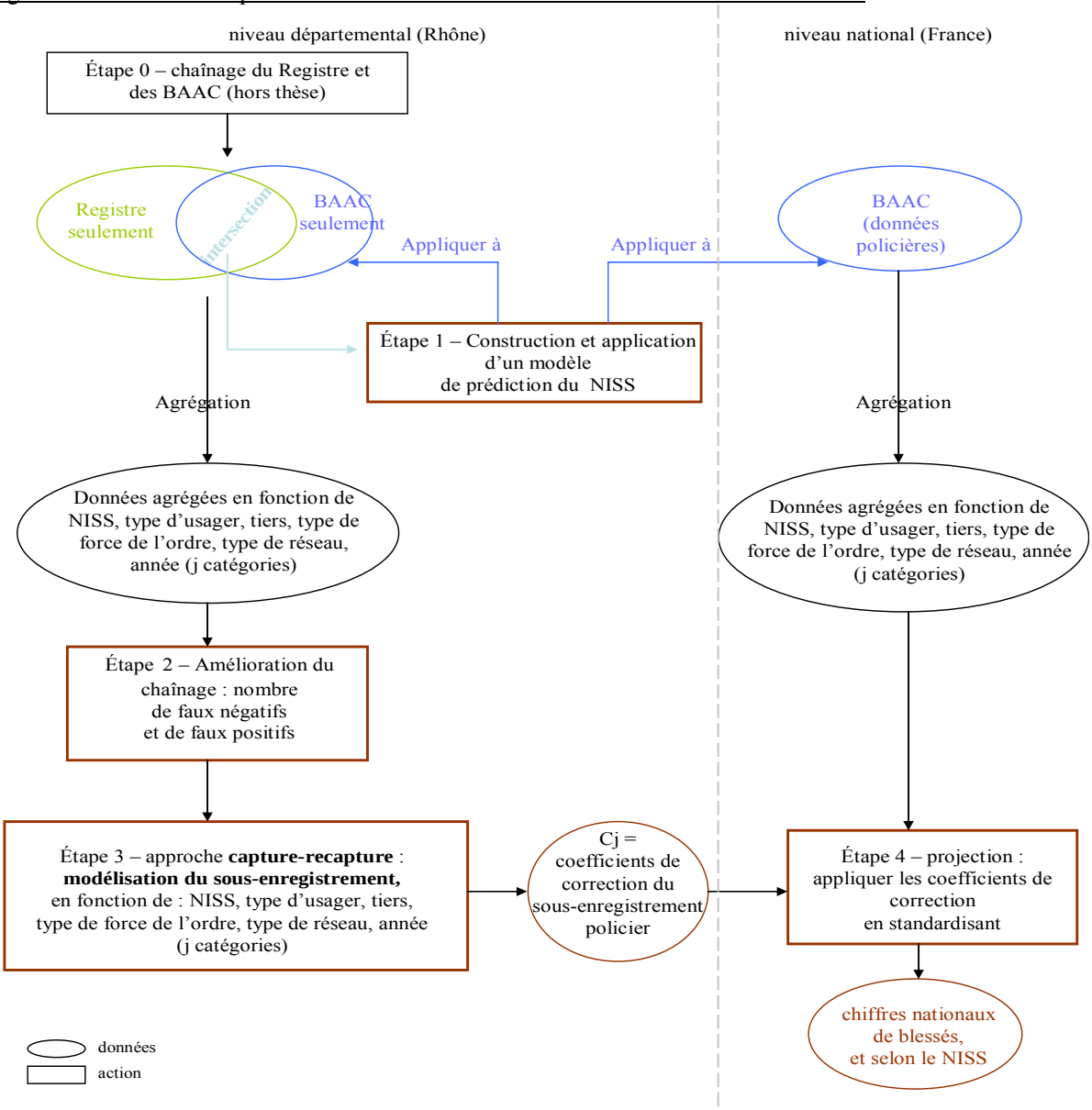
Le deuxième chapitre porte sur le classement de la gravité effectué par les forces de l'ordre. Celles-ci classent les blessés en deux catégories : « blessés légers » ou « blessés graves ». En prenant une classification médicale comme référence, nous quantifions les « erreurs » de classement par les forces de l'ordre. Nous explorons à quel point ces erreurs sont différentielles : nous identifions les caractéristiques associées à des erreurs de sous-classement ou de sur-classement de la gravité par les forces de l'ordre. Ce chapitre a fait l'objet d'une publication dans la revue Journal of Trauma. La connaissance des discordances entre classifications policière et médicale de la gravité nous sert pour la

suite (chap. 4) où nous avons besoin de prédire la gravité selon cette classification médicale, en nous appuyant en partie sur la classification policière.

Le troisième chapitre est une première application de la méthode de capture-recapture aux données des blessés de la route dans le Rhône. Nous avons utilisé une application simple, avec stratification, sans modélisation explicite, et restreinte aux données de 2001, pour lesquelles nous disposons d'informations supplémentaires grâce aux études complémentaires (chap. 1). Nous explorons l'application de l'approche capture-recapture sur trois scénarios, qui diffèrent selon le nombre estimé de blessés communs aux données des forces de l'ordre et au Registre. La stratification ne permet de prendre en compte que trois des facteurs majeurs de biais de sélection des données des forces de l'ordre, mis en évidence dans le chapitre 1. L'application de l'approche capture-recapture sera améliorée au chapitre 4, par une modélisation explicite. Ce chapitre 3 a fait l'objet d'une publication dans la revue Accident Analysis and Prevention.

Le quatrième chapitre couvre l'ensemble de la procédure d'estimation, sur la période 1996-2004. Cette procédure est schématisée dans la Figure 2 ; elle est présentée pour un des critères de santé étudiés, à savoir la gravité lésionnelle mesurée par le NISS (New Injury Severity Score).

Figure 2 : ensemble de la procédure d'estimation du bilan national de l'insécurité routière



Ce dernier chapitre est composé de quatre étapes. La première étape consiste à prédire la gravité des blessés enregistrés dans les données des forces de l'ordre selon une échelle médicale, en l'occurrence le NISS. Nous avons besoin de disposer de la gravité lésionnelle pour plusieurs raisons. D'une part, nous voulons un bilan national en fonction d'une échelle médicale de gravité (le chapitre 2 nous a confirmé que le classement de la gravité par les forces de l'ordre n'était pas satisfaisant) ; d'autre part, le chapitre 1 montre que la gravité lésionnelle est un biais de sélection majeur, et le chapitre 3 indique qu'il faut donc le prendre en compte pour appliquer la méthode de capture-recapture.
La deuxième étape consiste à améliorer le résultat du chaînage entre le Registre et les données des forces de l'ordre, selon le scénario 3 construit au chapitre 3. Il s'agit d'estimer le nombre de faux négatifs et de faux positifs et de les utiliser pour corriger le résultat du chaînage.
La troisième étape est une application améliorée de la méthode de capture-recapture aux données rhodaniennes. Nous modélisons explicitement le sous-enregistrement par un modèle multinomial logistique, ce qui nous permet de prendre en compte tous les facteurs majeurs de biais de sélection mis en évidence au chapitre 1. Cette modélisation nous fournit également les coefficients de correction du sous-enregistrement entre les données des forces de l'ordre et le bilan complet, définis en fonction des facteurs de biais de sélection.
La quatrième étape est la projection proprement dite : les coefficients de correction, estimés au niveau rhodanien, sont appliqués aux données nationales des forces de l'ordre, en standardisant sur les facteurs de biais de sélection.

Nous obtenons ainsi une estimation du nombre de blessés sur le territoire français, globalement, par type d'usager, par âge et sexe, et par niveau de gravité. Nous donnons aussi l'évolution de ces divers effectifs de 1996 à 2004. Les résultats et la discussion d'ensemble sont inclus dans ce quatrième et dernier chapitre. Un article correspondant à ce chapitre est soumis pour publication dans une revue anglophone de santé publique.

MATÉRIEL

Nous présentons ici les données utilisées dans l'ensemble de ce travail : données des forces de l'ordre, le Registre des victimes d'accident de la circulation routière du Rhône, et les données rapprochées : blessés identifiés comme communs aux deux sources de données, sur le Rhône, par chaînage. Selon l'objectif de tel ou tel chapitre, tout ou partie des données sera utilisée.

Données des forces de l'ordre

Les forces de l'ordre sont tenues d'établir un procès-verbal (PV) pour tout accident de la circulation routière causant au moins un blessé, dès qu'elles en ont connaissance. La définition policière d'un accident corporel est la suivante : il provoque au moins une victime, c'est-à-dire soit un usager « tué sur le coup » soit ayant nécessité des soins médicaux ; il survient sur une voie ouverte à la circulation publique ; il implique au moins un véhicule.
Le procès-verbal doit recenser tous les individus impliqués dans l'accident, et les classer en tués, blessés graves, blessés légers ou indemnes. Jusqu'en 2004 inclus, les définitions sont les suivantes : les tués sont les victimes décédées « sur le coup » ou dans les six jours suivant l'accident. Les blessés graves sont les blessés dont l'état nécessite plus de six jours d'hospitalisation ; les blessés légers sont ceux dont l'état nécessite de un à six jours d'hospitalisation (inclus) ou un soin médical sans hospitalisation. Les indemnes sont les impliqués dont l'état ne nécessite aucun soin médical.

Pour chaque procès-verbal, une partie des informations contenues est codée et saisie afin de constituer un enregistrement « BAAC » (Bulletin d'Analyse d'Accident Corporel de la circulation). Les informations enregistrées sont les caractéristiques des lieux de l'accident, de l'accident lui-même, des véhicules impliqués et des personnes impliquées. L'ensemble des BAAC forme ainsi une base de données nationale, le « fichier national des accidents corporels de la circulation routière ». Les données couvrent tout le territoire français. Elles sont également disponibles au niveau départemental ; la sélection se fait sur le département en tant que lieu d'accident.

Nous appelons données policières ou données des forces de l'ordre les données contenues dans les BAAC. Nous utilisons l'adjectif policier au sens large ; si nous parlons d'une force de l'ordre en particulier, cela sera spécifié. Les différentes forces de l'ordre sont la police, la gendarmerie, les CRS (Compagnies Républicaines de Sécurité), la police aux frontières, et la préfecture de police de Paris. Dans le département du Rhône, seules les trois premières sont présentes : la police, la gendarmerie, et les CRS.

Le Registre des victimes d'accidents de la circulation routière

Un registre des victimes de la circulation routière dans le Rhône existe depuis 1995 (Laumon, Martin et al. 1997; Charnay and Laumon 2002; Laumon 2002). Il est « qualifié » par le Comité national des Registres. Le Registre est géré par l'ARVAC (Association pour le Registre des Victimes d'Accidents de la Circulation).

Le Registre est basé sur la participation de toutes les structures hospitalières (publiques et privées) dans le Rhône, et de ses proches alentours, qui sont susceptibles d'accueillir des victimes d'accidents de la route. Ainsi il inclut environ 220 services : SMUR, SAMU, services d'urgence, services de réanimation, de chirurgie, … jusqu'aux services de réadaptation. Il recense tous les patients, hospitalisés et non-hospitalisés.
Le critère d'inclusion géographique du Registre est le département en tant que lieu d'accident (et non en tant que lieu de résidence comme c'est le cas de la plupart des registres médicaux).

Pour chaque victime, les informations suivantes sont collectées : quelques caractéristiques de l'accident (date, lieu, type d'usager, d'antagoniste,…) et surtout les caractéristiques de la victime (sexe, âge, date de naissance, adresse, service consulté, durée d'hospitalisation et transfert si pertinent, et bien sûr, bilan lésionnel). Pour chaque victime, le bilan lésionnel est basé sur l'ensemble des diagnostics posés par les différents services que la personne a consulté. Les diagnostics en clair sont codés par le médecin du Registre, selon la codification AIS (Abbreviated Injury Scale), révision de 1990 (AAAM 1990). Chaque blessure est ainsi codée selon la région corporelle, la structure anatomique, le type d'atteinte lésionnelle. Un niveau de gravité lui est attribué. Il s'agit de la gravité immédiate en fonction de la blessure elle-même, et non en fonction de l'état de santé de la personne, ou de sa capacité de récupération. Cette échelle va de 1 à 6 (1=gravité mineure, 2=modérée, 3=sérieuse, 4=sévère, 5=critique, 6=au-delà des ressources thérapeutiques, et 9=indéterminé, lorsque la description lésionnelle est imprécise).
Le Registre n'inclut que les blessés au sens de la classification AIS. Cela signifie par exemple que les personnes en état de choc psychologique ou gardées en observation ou présentant des symptômes, mais sans lésion anatomique, ne font pas partie des victimes dénombrées par le Registre.

Pour estimer la gravité globale des blessures d'une personne, plusieurs critères ont été développés : le MAIS, l'ISS et le NISS. Le MAIS est le Maximum AIS, c'est-à-dire le score de gravité maximum parmi les blessures de la victime. L'ISS (Baker, O'Neill et al. 1974) est le Injury Severity Score, et correspond à la somme des carrés des scores AIS des trois blessures les plus graves de trois régions corporelles distinctes, parmi six définies. Le NISS (Osler, Baker et al. 1997) est le New Injury Severity Score, et correspond à la somme des carrés des scores AIS des trois blessures les plus graves, indépendamment des régions corporelles atteintes. Il y a eu débat sur le choix du critère à utiliser (Brenneman, Boulanger et al. 1998; Sacco, MacKenzie et al. 1999; Samin and Civil 1999; Stevenson, Segui-Gomez et al. 2001; Kilgo, Osler et al. 2003; Lavoie, Moore et al. 2004; Tay, Sloan et al. 2004; Frankema, Steyerberg et al. 2005). L'ISS et le NISS tiennent mieux compte des lésions multiples que le MAIS, et le NISS n'a pas de contrainte de prise en compte de trois régions distinctes. Il a l'inconvénient de dépendre plus de la qualité de description lésionnelle. Le NISS semble être mieux corrélé avec la probabilité de survie (Osler, Baker et al. 1997; Brenneman, Boulanger et al. 1998; Lavoie, Moore et al. 2004; Frankema, Steyerberg et al. 2005). Nous privilégions donc le NISS comme score de gravité individuelle.

Données rapprochées

Une procédure de chaînage est appliquée chaque année pour identifier les victimes communes au Registre et aux BAAC. Cette procédure est basée sur une approche probabiliste semi-automatisée. Les variables de chaînage sont : la date de l'accident (année, mois, jour), le lieu de l'accident, le type d'usager, l'année de naissance, le mois de naissance, et le sexe de la victime. La procédure n'est pas complètement automatisée car la variable lieu de l'accident est laissée en texte libre afin de ne perdre aucune information : elle est en effet très riche ; elle contient la commune et souvent la rue ou la route. La décision de chaîner une paire d'enregistrements est prise lorsqu'au moins un certain nombre de variables concordent. Cette procédure a été utilisée de 1996 à 2001. À partir de 2002, le chaînage entre le Registre et les BAAC se fait lors de la saisie des victimes dans le Registre afin que le Registre puisse bénéficier de certaines informations contenues dans les BAAC ; cela implique que la décision de relier deux enregistrements est plus conservatrice.

Le registre a reçu l'autorisation de la CNIL (Commission Nationale de l'informatique et des Libertés). Les BAAC ont une date de création antérieure à la CNIL. Ce sont une base de données institutionnelle, mise à disposition de l'INRETS ; c'est ainsi que nous y avons accès.
Le résultat du chaînage entre le Registre et les BAAC est conservé au sein du Registre. Ainsi les contrôles et restrictions d'accès aux données du Registre s'appliquent à ces données. Le personnel travaillant sur ces données est bien entendu tenu à la confidentialité. D'une manière générale, nous

sommes soucieux de respecter la déontologie et les bonnes pratiques de l'épidémiologie (Goldberg, Chevalier et al. 2007).

Exclusions

Toute l'étude porte sur les blessés uniquement. Les tués sont plutôt bien enregistrés donc ils ne sont pas considérés ici. Les indemnes (présents dans les BAAC quand ils sont impliqués dans un accident corporel) sont exclus.

Nous utilisons les critères BAAC de définition des accidents de la circulation routière. Selon cette définition, les usagers de trottinette, rollers et planches à roulette sont considérés comme des piétons, et donc considérés comme victimes de la circulation routière que s'ils ont été heurtés par un véhicule. Nous excluons donc du Registre les victimes suivantes : les utilisateurs de roller, skateboard et trottinette s'ils se sont blessés seuls, ou contre un piéton.

Période d'étude

L'analyse des biais de sélection des données des forces de l'ordre (chap. 1) et celle des erreurs de classement sur la gravité (chap. 2) sont basées sur la période 1997-2001. L'année 1996 a été exclue (car elle a déjà fait l'objet d'une précédente analyse). Les données des années ultérieures à 2001 n'étaient pas disponibles lorsque ces deux analyses ont été conduites (en 2004).
Les deux études complémentaires (chap. 1) portent sur une seule année, la plus récente lors de leur mise en place, à savoir l'année 2001.
La première application de la méthode de capture-recapture aux données des blessés du Rhône porte sur les données de l'année 2001. Elle utilise des informations fournies par les deux études complémentaires.
Pour l'estimation du bilan national, nous travaillons sur la période la plus longue possible, soit 1996-2004 (analyse conduite en 2006). Le changement de stratégie dans le chaînage entre les BAAC et le Registre à partir de 2002 est pris en compte dans la procédure d'estimation.

Chapitre 1
Biais de sélection des données d'accidentologie

1.1 Biais de sélection des données des forces de l'ordre (analyse des blessés recensés par le Registre)

Introduction

Ce chapitre reprend essentiellement un article que nous avons publié (cf. annexe), et le complète.

Dans la plupart des pays, les données nationales sur les accidents de la route et leurs victimes sont fournies par les forces de l'ordre. Ces données constituent ainsi la base de la plupart des connaissances épidémiologiques sur les victimes d'accidents de la route. Il importe donc de connaître leur validité. L'enregistrement des tués est quasi-exhaustif (Elvik and Mysen 1999; Laumon 2002). En revanche, les blessés sont largement sous-enregistrés (Hauer and Hakkert 1988; James 1991; Hvoslef 1994; Elvik and Mysen 1999). Cette non-exhaustivité pose la question de la représentativité des données collectées, autrement dit la question des biais de sélection.

Un certain nombre d'études, dans divers pays, ont été conduites sur le sujet (Hauer and Hakkert 1988; James 1991; Hvoslef 1994; Stutts and Hunter 1998; Aptel, Salmi et al. 1999; Elvik and Mysen 1999; Alsop and Langley 2001; Cryer, Westrup et al. 2001; Dhillon, Lightstone et al. 2001; Rosman 2001; Laumon and Martin 2002; Langley, Dow et al. 2003). Un tableau est donné en annexe, listant ces études, leurs principales caractéristiques (pays, taille de l'étude, restrictions, type d'étude et type d'analyse), et leurs résultats. Les principaux facteurs de biais de sélection mis en évidence sont la gravité des blessures, le type d'usager et le nombre de véhicules impliqués dans l'accident (un seul versus plusieurs).

Nous conduisons une étude des biais de sélection des données des forces de l'ordre françaises, similaire à une étude précédente restreinte aux données de 1996 (Laumon and Martin 2002). Nous évaluons d'abord le taux de sous-enregistrement, puis comment il varie en fonction de caractéristiques du blessé et de l'accident. Une telle étude est possible grâce à l'existence du Registre des victimes d'accidents de la circulation routière du Rhône, et aux qualités de ce Registre ; notamment, et c'est ce qui nous importe ici, son taux de couverture est bien plus élevé que celui des données des forces de l'ordre et ses biais de sélection sont sans doute limités. Les données du Registre et les données des forces de l'ordre ont été chaînées afin d'identifier les victimes qui sont communes aux deux sources d'enregistrement. Il est ainsi possible d'analyser le « filtre » des forces de l'ordre, i.e. le filtre à travers lequel les victimes du Registre sont « sélectionnées » pour être enregistrées dans les données des forces de l'ordre. Nous conduisons une analyse multivariée : nous modélisons, parmi les blessés du Registre, la probabilité d'être enregistré dans les BAAC, en fonction de caractéristiques du blessé et de l'accident.

Dans cette analyse le sous-enregistrement des données des forces de l'ordre couvre à la fois le sous-signalement de la part des blessés vers les forces de l'ordre, et le sous-enregistrement de la part des forces de l'ordre elles-mêmes. Le premier correspond à la situation où les personnes impliquées dans un accident corporel n'alertent pas les forces de l'ordre. Le second correspond à la situation où, bien que les forces de l'ordre aient été alertées, elles n'établissent pas de procès-verbal, ou omettent certains blessés, ou n'enregistrent pas le procès-verbal dans la base de données des BAAC. Nous ne pouvons pas distinguer ici si le sous-enregistrement provient des impliqués ou des forces de l'ordre. La distinction sera possible dans l'enquête auprès des blessés identifiés dans le Registre seulement (cf. section 1.2)

Matériel

Cette analyse porte sur les données du Registre de 1997 à 2001. Ces données ont été chaînées avec les données des forces de l'ordre, et ainsi les blessés communs au Registre et aux données des forces de l'ordre ont été identifiés.

S'agissant d'étudier les biais de sélection des données des forces de l'ordre (mais pas les différences de définition entre BAAC et Registre), nous appliquons les critères BAAC de définition des accidents de la circulation routière à l'ensemble des victimes du Registre. Nous excluons donc du Registre les utilisateurs de roller, skateboard et trottinette s'ils se sont blessés seuls, ou contre un piéton.
Nous n'excluons pas ici les victimes d'un accident ayant eu lieu hors-réseau car ce critère n'est ni clairement défini ni appliqué systématiquement par les forces de l'ordre (l'un étant sans doute la conséquence de l'autre).

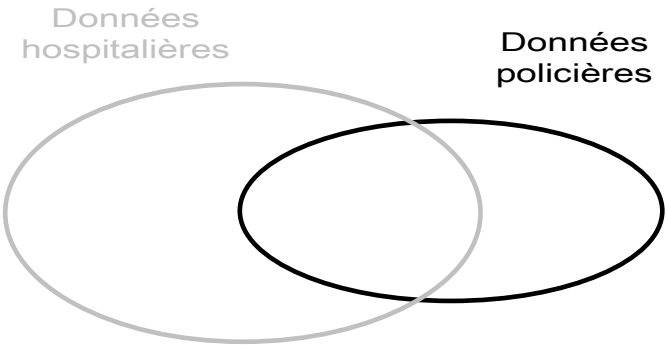
L'analyse est basée sur la période 1997-2001 (l'année 1996 a été exclue car elle a déjà fait l'objet d'une analyse ; l'année 2001 était la dernière année de données disponibles lorsque l'analyse a été conduite, en 2004).

Méthode

Sous-enregistrement

La première partie de l'analyse consiste simplement à estimer le degré du sous-enregistrement des blessés dans les données des forces de l'ordre (ou taux de couverture). Nous calculons le taux d'enregistrement policier, en utilisant la définition proposée par Elvik et Mysen (Elvik and Mysen 1999) comme la seule correcte (n°3 dans leur article). Cette définition est le ratio entre le nombre de blessés dans les fichiers policiers et le nombre total d'après les données hospitalières (ici le Registre) et les données des forces de l'ordre ; ce total est la somme des effectifs de ceux identifiés par les forces de l'ordre seulement, ceux identifiés par les hôpitaux (Registre) seulement, et ceux identifiés par les deux.

Figure 3 : schématisation des deux sources d'enregistrements des blessés dans le département du Rhône



Biais de sélection

La seconde partie de l'analyse a pour but d'identifier et de quantifier les biais de sélection. Nous considérons les caractéristiques suivantes comme potentiels biais de sélection (parmi les variables dont nous disposons dans le Registre) :

1) les caractéristiques du blessé ou de l'accident : la gravité lésionnelle, l'âge, le sexe, le type d'usager, la présence de tiers dans l'accident (oui/non),
2) les caractéristiques du lieu de l'accident : l'environnement (Grand Lyon versus hors Grand Lyon), la luminosité (jour versus nuit), et le type de jour (début versus fin de semaine),
3) les caractéristiques générales : la force de l'ordre contrôlant la zone où l'accident a lieu, et l'année calendaire.
Les caractéristiques sont fournies par les variables du Registre.

La gravité lésionnelle est mesurée par le NISS, regroupé en cinq catégories : NISS 1-3, 4-8, 9-15, 16-24, 25-75. Les seuils correspondent au changement de niveau de gravité d'une blessure isolée, et nous pouvons ainsi labelliser ces catégories en gravité mineure, modérée, sérieuse, sévère, critique.
Le type d'usager est décliné en six catégories : piétons / cyclistes / usagers de 2RM (deux-roues motorisé) / automobilistes / occupants de camionnettes, camions, bus, tram / non renseigné et autre. La présence/absence d'au moins un tiers impliqué dans l'accident est notée avec / sans tiers.
La variable Grand Lyon vs hors Grand Lyon est la seule variable dont nous disposons dans le Registre sur 1997-2001 pour mesurer un environnement urbain ou rural. C'est une approximation car le Grand Lyon (communauté urbaine de Lyon) comprend quelques communes plutôt rurales, et réciproquement, la zone hors Grand Lyon contient des villes d'une certaine taille (Villefranche : 31 000 habitants, Givors : 19 000, Tarare : 11 000).

Les forces de l'ordre sur le département du Rhône sont la police, la gendarmerie et les CRS. Ces trois forces de l'ordre nous intéressent pour deux raisons : 1) il se peut que le taux de sous-enregistrement soit différent entre ces trois forces de l'ordre, 2) ces trois forces de l'ordre sont aussi des indicateurs du caractère urbain ou rural du lieu de l'accident. Schématiquement, le secteur de compétence de la gendarmerie est celui de la rase campagne et des agglomérations de moins de 7000 habitants, celui de la police : les agglomérations de plus de 7000 habitants, et celui des CRS : les autoroutes et voie rapides en zone urbaine. Dans le Rhône, il faut noter qu'un certain nombre de communes du Grand Lyon sont couvertes par la gendarmerie.
Nous rappelons que nous travaillons ici sur les données du Registre, et donc que la force de l'ordre en charge du secteur où a eu lieu l'accident n'est pas connue a priori. Nous la déterminons en fonction de la commune et du type de réseau où a eu lieu l'accident. Pour chaque commune du département du Rhône, nous disposons de l'information sur la force de l'ordre en charge de la commune : police ou gendarmerie. Pour le réseau des autoroutes et voies rapides, nous disposons de l'information sur la force de l'ordre en charge, selon les sections : gendarmerie en rural et CRS en agglomération. Si le lieu d'accident est manquant, nous ne pouvons pas déterminer la force de l'ordre.

Parmi les blessés du Registre, nous comparons les caractéristiques mentionnées ci-dessus, entre les blessés identifiés seulement dans le Registre et ceux identifiés à la fois dans le Registre et dans les données des forces de l'ordre. Afin de conduire une analyse multivariée, nous modélisons la probabilité, parmi l'ensemble des blessés du Registre, d'être identifié dans les BAAC, et nous l'exprimons comme une fonction des caractéristiques du blessé et de l'accident. L'unité statistique est le blessé (et non l'accident). Nous considérons comme candidates à l'inclusion dans le modèle multivarié les variables mises en évidence dans la littérature ou celles significatives (à 5 %) en univarié. Nous considérons aussi comme candidates à l'inclusion dans le modèle les interactions qui sont structurelles. Il s'agit de l'interaction entre le type d'usager et la présence/absence d'un tiers (les piétons ne sont considérés comme victimes de la route que si heurtés par un véhicule, autrement dit avec tiers) et de l'interaction entre le type de réseau, l'environnement Grand Lyon vs hors Grand Lyon et la force de l'ordre (puisque la force de l'ordre dépend des deux premiers).

Nous souhaitons quantifier les biais de sélection à l'aide des risques relatifs (RR) et non à l'aide d'odds ratios (OR), qui sont facilement estimables par une régression logistique mais difficilement interprétables. Nous ne sommes pas dans la situation où le risque relatif peut être approximé par l'odds ratio, car l'événement étudié (être identifié dans les données des forces de l'ordre) n'est pas un événement rare. Dans cette situation, différents modèles existent pour estimer directement les risques relatifs (Greenland 2004; Spiegelman and Hertzmark 2005), tous dans le cadre des modèles linéaires

généralisés : modèles log-binomial, logistique avec des transformations, et Poisson modifié. Le modèle log-binomial présente souvent des difficultés de convergence. Le modèle logistique et ses transformations afin de passer des odds ratios aux risques relatifs peut s'avérer laborieux (et donc source d'erreurs). Nous choisissons d'utiliser la régression de Poisson modifiée. Cela consiste à appliquer un modèle de Poisson à des données binomiales, ce qui conduirait à une sur-estimation de l'écart-type de l'estimateur du risque relatif. On remédie à ceci grâce à une procédure d'estimation robuste de la variance, l'estimateur sandwich (Zou 2004). L'implémentation dans SAS a été décrite par Linquist (Lindquist 2004). La transcription sur nos données est fournie en annexe.

Outre les risques relatifs ajustés, nous fournissons les résultats du modèle multivarié sous forme de probabilités ajustées (standardisées). Elles correspondent aux probabilités marginales pour une caractéristique, « toutes choses égales par ailleurs », c'est-à-dire, si les autres caractéristiques étaient réparties de façon équilibrée. Ces probabilités sont calculées en standardisant sur les autres variables du modèle, plus précisément en moyennant sur les autres variables, en imposant des poids égaux aux différentes modalités de chaque variable catégorielle. Ceci est possible avec l'option LSMEANS de la procédure GENMOD de SAS (détail en annexe).

Valeurs manquantes :
Les observations où le NISS n'est pas renseigné (i.e. au moins une lésion AIS 9) représentent 1,3 % des observations ; elles sont exclues de l'analyse des biais de sélection. Les observations où l'âge et le sexe sont non-renseignés sont peu nombreuses (0,18 et 0,22 % respectivement). Les valeurs sont imputées en fonction des variables âge et sexe des BAAC pour les blessés identifiés dans les deux sources ; sinon, elles sont obtenues par imputation simple : l'âge en fonction du sexe et du type d'usager, et le sexe en fonction de l'âge et du type d'usager. Pour les variables type d'usager, type de réseau, et jour/nuit, les observations avec valeur manquante sont nombreuses ; nous constituons une catégorie spécifique (pour le type d'usager, elles sont regroupées avec « autre »).

Résultats

Sur la période 1997-2001, le Registre recense 53 988 blessés (non décédés dans les 6 jours). Nous excluons 2374 utilisateurs de rollers, skate-board ou trottinette s'étant blessés seuls, ce qui laisse 51 614 observations côté Registre. Sur la même période, les données des forces de l'ordre recensent 22 498 blessés. Parmi ceux-ci, 14 398 sont aussi identifiés dans le Registre. Le nombre total de blessés recensés par le Registre et les données des forces de l'ordre s'élève à 59 714 (effectif de Registre seulement + effectif de BAAC seulement + effectif de l'intersection = (51 614-14 398) + (22 498-14 398) + 14 398). Le taux d'enregistrement des blessés dans les données des forces de l'ordre est donc de 37,7 % (=22 498 / 59 714), et son intervalle de confiance à 95% est [37,3% -38,1%].

Figure 4 : répartition du nombre de blessés (non-décédés) selon la source d'enregistrement, Rhône 1997-2001

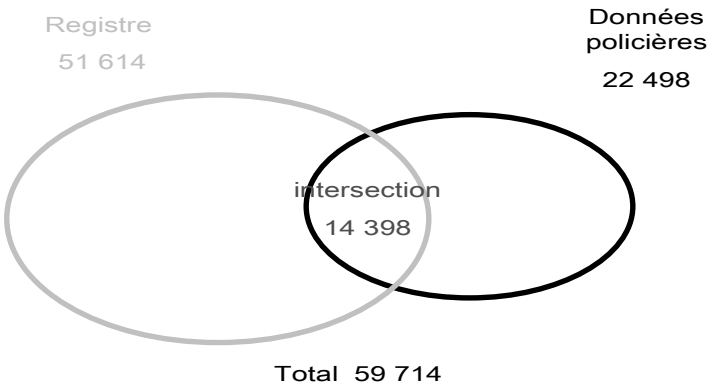


Tableau 1 : risques relatifs ajustés et probabilités ajustées (standardisées) d'enregistrement dans les données des forces de l'ordre; analyse multivariée (50 936 blessés du Registre, Rhône, 1997-2001)

		n	probabilité	RR	IC à 95 %	
gravité lésionnelle (p<0,0001)						
NISS 1-3		37 873	15,1 %	0,38	0,36	0,41
NISS 4-8		8 632	24,3 %	0,62	0,58	0,66
NISS 9-15		2 822	33,1 %	0,84	0,79	0,90
NISS 16-24		997	37,6 %	0,96	0,89	1,03
NISS 25-75 (référence)		612	39,3 %	1,00	-	-
âge du blessé (p<0,0001)						
0-13 ans		5 712	27,1 %	0,95	0,90	1,00
14-17		5 116	26,8 %	0,93	0,89	0,98
18-24		12 564	27,8 %	0,97	0,94	1,00
25-64 (référence)		25 111	28,6 %	1,00	-	-
65 ans et plus		2 433	30,9 %	1,08	1,03	1,13
sexe (p=0,0001)						
féminin		19 441	27,5 %	0,95	0,92	0,97
masculin (référence)		31 495	29,0 %	1,00	-	-
type d'usager (p<0,0001)	tiers					
piétons	avec	4 742	44,4 %	0,99	0,95	1,03
cyclistes	avec	1 651	33,6 %	0,75	0,69	0,81
usagers de 2RM	avec	4 924	46,5 %	1,03	0,99	1,07
automobilistes (référence)	avec	20 744	45,0 %	1,00	-	-
camionnettes, camions, bus, tram	avec	476	52,7 %	1,17	1,07	1,29
non renseigné et autre	avec	44	55,0 %	1,22	0,90	1,67
interaction type d'usager × tiers (p<0,0001)						
cyclistes	sans	5 284	2,1 %	0,06	0,05	0,08
cyclistes (référence)	avec	-	-	1,00	-	-
usagers de 2RM	sans	5 127	14,7 %	0,32	0,29	0,35
usagers de 2RM (référence)	avec	-	-	1,00	-	-
automobilistes	sans	6 870	34,9 %	0,78	0,75	0,81
automobilistes (référence)	avec	-	-	1,00	-	-
camionnettes, camions, bus, tram	sans	675	25,5 %	0,48	0,41	0,57
camionnettes, camions … (référence)	avec	-	-	1,00	-	-
non renseigné et autre	sans	399	36,3 %	0,66	0,46	0,96
non renseigné et autre (référence)	avec	-	-	1,00	-	-
force de l'ordre (zone) × type de réseau × environnement (p<0,0001)						
CRS, autoroutes		4 508	55,1 %	1,16	1,12	1,21
gendarmerie, autoroutes, hors Grand Lyon		117	46,3 %	0,98	0,81	1,18
gendarmerie, RN et RD, hors Grand Lyon		3 027	37,7 %	0,80	0,76	0,83
gendarmerie, RN et RD, Grand Lyon		474	29,9 %	0,63	0,56	0,71
gendarmerie, voies communales		7 221	22,5 %	0,48	0,45	0,50
gendarmerie, hors réseau		810	7,6 %	0,16	0,11	0,24
police, RN et RD, Grand Lyon		266	58,1 %	1,23	1,10	1,37
police, tout sauf autoroute, hors Grand Lyon		1 523	19,6 %	0,41	0,37	0,47
police, voies communales, Grand Lyon (ref.)		18 764	47,3 %	1,00	-	-
police, hors réseau		837	18,4 %	0,39	0,32	0,48
non renseigné		13 749	18,3 %	0,39	0,37	0,40
conditions lumineuses (p<0,0001)						
jour		27 059	28,4 %	0,92	0,89	0,95
nuit (référence)		10 107	30,8 %	1,00	-	-
non renseigné		13 770	25,7 %	0,83	0,80	0,86
année (p<0,0001)						
1997		9 921	26,9 %	0,98	0,94	1,02
1998		10 379	28,8 %	1,05	1,01	1,09
1999		10 704	29,1 %	1,06	1,02	1,10
2000		9 595	28,9 %	1,05	1,01	1,10
2001 (référence)		10 337	27,4 %	1,00	-	-

d'après une régression de Poisson avec estimation robuste de la variance

Pour l'analyse multivariée, 678 victimes où le NISS est manquant sont exclues. L'analyse multivariée porte donc sur 50 936 observations.
Lors des analyses univariées, toutes les variables potentielles facteurs de biais de sélection apparaissent comme significativement associées avec la probabilité d'enregistrement dans les données des forces de l'ordre. En multivarié, une variable -le type de jour- n'est plus significative. Le modèle multivarié comprend donc les variables suivantes : gravité lésionnelle (NISS), âge, sexe, type d'usager et présence/absence de tiers (interaction), jour/nuit, type de réseau, Grand Lyon vs hors Grand Lyon, force de l'ordre (interaction des trois), et année calendaire (sous forme qualitative).

Les résultats du modèle multivarié sont donnés dans le Tableau 1. Par comparaison aux probabilités ajustées, la probabilité moyenne observée d'être enregistré dans les données des forces de l'ordre, parmi les blessés du Registre est de 27,9 % (=14 398/51 614). L'analyse multivariée fait ressortir les biais de sélection suivants.
L'enregistrement dans les données des forces de l'ordre est fortement associé à la gravité lésionnelle : il y a un gradient très clair de probabilité décroissante d'être recensé dans les données des forces de l'ordre avec une gravité lésionnelle décroissante. Par rapport aux blessés NISS 25-75 (gravité critique), les victimes très légèrement blessées (NISS 1-3) ont 2,6 fois moins de chances d'être enregistrées dans les données des forces de l'ordre.
L'enregistrement dans les données des forces de l'ordre est légèrement associé à l'âge des blessés : plus les blessés sont jeunes, moins ils ont de chance d'être enregistrés dans les données des forces de l'ordre.
L'enregistrement dans les données des forces de l'ordre est légèrement associé avec le sexe : les femmes sont 1,05 fois moins susceptibles d'être enregistrées que les hommes.
L'enregistrement est fortement associé au type d'usager et à la présence de tiers dans l'accident. Le premier groupe de lignes du tableau sur le type d'usager donne les risques relatifs d'être enregistré dans les données des forces de l'ordre, parmi les accidents avec tiers, par rapport aux automobilistes (catégorie de référence). Parmi les accidents avec tiers, les automobilistes, les usagers de 2RM et les piétons ont la même probabilité d'être enregistrés dans les données des forces de l'ordre. Les cyclistes ont 1,33 (=1/0,75) fois moins de chance d'être enregistrés. Les occupants de camionnettes, camions, bus, trams ont 1,17 fois plus de chance d'être enregistrés que les automobilistes. Le deuxième groupe de lignes du tableau sur le type d'usager, donne le risque relatif d'être enregistré dans les données des forces de l'ordre pour un accident sans tiers par rapport à un accident avec tiers, pour un type d'usager donné. Les blessés dans un accident sans tiers ont moins de chance d'être enregistrés dans les données des forces de l'ordre que les blessés dans un accident avec tiers : 1,3 fois moins pour les automobilistes, 3 fois moins s'il s'agit des usagers de 2RM, et 17 fois moins s'il s'agit des cyclistes (inverse des risques relatifs donnés dans le tableau). La probabilité ajustée des cyclistes se blessant seuls, d'être enregistrés dans les données des forces de l'ordre, est de 2,1 %.
La probabilité d'enregistrement dans les données des forces de l'ordre est légèrement associée avec la caractéristique jour/nuit : les blessés impliqués dans un accident le jour ont moins de chances d'être enregistrés que ceux impliqués dans un accident la nuit.
Enfin, la probabilité d'enregistrement dans les données des forces de l'ordre est légèrement associée avec l'année calendaire.
L'enregistrement dans les données des forces de l'ordre est fortement associé avec la combinaison type de réseau, environnement (Grand Lyon vs hors Grand Lyon) et force de l'ordre ; la probabilité d'enregistrement dans les données des forces de l'ordre diminue avec le type de réseau : des autoroutes aux routes nationales et départementales, aux voies communales et jusqu'au hors-réseau. Le gradient est très clair dans les zones contrôlées par la gendarmerie. Il l'est moins dans les zones contrôlées par la police. Dans celles-ci il y a une forte différence de probabilité d'enregistrement entre Grand Lyon et hors Grand Lyon, avec une plus forte probabilité d'enregistrement pour les blessés accidentés dans le Grand Lyon. Cela ne se retrouve pas dans les zones contrôlées par la gendarmerie ; il semblerait que ce soit en fait l'inverse.
Finalement, quelle que soit la caractéristique étudiée, la catégorie 'non renseignée' est associée à une probabilité plus faible d'enregistrement dans les données des forces de l'ordre.

Discussion

Le taux d'enregistrement (37,7 %) est plutôt bas. Il est cependant cohérent avec la littérature sur le sujet ; il est même très proche de la moyenne estimée sur 13 pays (dont la plupart industrialisés) (Hauer and Hakkert 1988; James 1991; Hvoslef 1994; Elvik and Mysen 1999). En outre, le Registre couvre les hospitalisés et les non-hospitalisés alors que dans beaucoup d'études, les données hospitalières se limitent aux hospitalisés, ce qui tend à gonfler artificiellement le taux d'enregistrement des données des forces de l'ordre.

Le deuxième résultat de cette étude concerne les biais de sélection eux-mêmes : leur identification, et leur ampleur, ajustée sur les autres biais. Les facteurs de biais les plus importants en termes d'ordre de grandeur du risque relatif (autrement dit en termes de variation sur la probabilité d'enregistrement) sont la gravité lésionnelle, le type d'usager combiné avec l'absence / présence de tiers, et la combinaison type de réseau, Grand Lyon/hors Grand Lyon et force de l'ordre en charge du secteur.

La probabilité d'être enregistré dans les données des forces de l'ordre augmente avec la gravité lésionnelle : c'est en quelque sorte une évidence ; plus la gravité lésionnelle est élevée, plus les personnes impliquées dans l'accident, et les forces de l'ordre si elles sont présentes, vont considérer que le blessé est un vrai blessé et doit être déclaré. Ce facteur de biais a été largement rapporté dans la littérature (Hauer and Hakkert 1988; James 1991; Hvoslef 1994; Elvik and Mysen 1999).

La probabilité d'être enregistré dans les données des forces de l'ordre dépend du type d'usager. C'est aussi un facteur de biais communément relevé dans la littérature (Hauer and Hakkert 1988; James 1991; Hvoslef 1994; Elvik and Mysen 1999). Notamment, les victimes cyclistes ont un faible taux d'enregistrement ; il est extrêmement bas pour les cyclistes qui se blessent seuls (sans tiers), et bien loin de celui des autres usagers blessés dans un accident sans tiers : les accidents de vélo sont rarement considérés comme des accidents de la circulation, même lorsqu'il y a un tiers, que ce soit par les cyclistes eux-mêmes, les autres usagers ou les forces de l'ordre. C'est aussi le cas dans les autres pays (Hvoslef 1994; Elvik and Mysen 1999), même dans ceux d'Europe du Nord où pourtant l'utilisation du vélo comme moyen de transport est très répandue.

La probabilité d'être enregistré dans les données des forces de l'ordre est plus forte pour les accidents avec tiers (que sans tiers) ; ceci s'explique sans doute par la notion de responsabilité vis-à-vis des assurances et de la justice. Parmi les accidents avec tiers la probabilité d'enregistrement est encore plus forte pour les blessés occupants de camions, camionnettes, bus : la plupart sont sans doute aussi des accidents de travail (accidents de mission), et donc avec une raison supplémentaire pour qu'ils soient déclarés. Le biais de sélection selon la présence/absence de tiers est cohérent avec ce que montrent d'autres études, à savoir que les accidents à plusieurs véhicules sont mieux enregistrés que les accidents à un seul véhicule (Hauer and Hakkert 1988; James 1991; Hvoslef 1994; Elvik and Mysen 1999).

L'interaction entre le type de réseau, l'environnement (Grand Lyon vs hors Grand Lyon), et la force de l'ordre en charge de la zone et la probabilité d'enregistrement dans les données des forces de l'ordre paraît complexe mais elle est liée à la dépendance structurelle entre ces trois caractéristiques. En agglomération, les différents types de route ne sont pas très distincts ; cela peut expliquer qu'il n'y ait pas de différence de probabilité d'enregistrement entre ces types pour la police dans le Grand Lyon. En revanche, hors Grand Lyon (soit en grande partie en zone rurale), la plus grande probabilité d'enregistrement en fonction du « niveau » de route (autoroutes, routes nationales et départementales, voies communales, hors réseau) vient sans doute d'une gravité attendue plus élevée, liée à une vitesse plus élevée (la gravité réelle est prise en compte par l'ajustement sur la gravité mesurée par le NISS ; mais elle ne recouvre pas exactement la gravité attendue par les impliqués et par les forces de l'ordre). Dans la littérature, le type de réseau n'apparaît pas (soit il n'a pas été considéré, soit il n'est pas apparu comme significativement associé à la probabilité d'enregistrement) ; seul le hors-réseau a été mis en évidence comme associé à de plus faibles probabilités d'enregistrement (James 1991). Cela s'explique

soit par le flou autour de la définition d'un accident de la circulation routière et la notion de « réseau ouvert à la circulation publique » et/ou la perception, la représentation des impliqués et des forces de l'ordre de ce qu'est un accident de la circulation routière.
Il y a une différence de probabilité d'enregistrement entre hors Grand Lyon et Grand Lyon dans les zones « police », mais pas dans les zones « gendarmerie » ni dans les zones « CRS ». L'association entre taux d'enregistrement et zone urbaine/rurale a semble-t-il été très peu explorée ; une étude en Australie (Lopez, Rosman et al. 2000) n'a trouvé aucune différence significative entre l'enregistrement des blessés accidentés dans une grande métropole versus accidentés en zone rurale. L'association entre force de l'ordre (dans les pays où cela aurait un sens) et taux d'enregistrement n'a apparemment pas été étudiée.

Il y a plusieurs facteurs de biais de sélection mineurs, i.e. des caractéristiques (âge, sexe, jour/nuit, année) qui sont légèrement mais significativement associées à la probabilité d'enregistrement dans les données des forces de l'ordre ; cela est favorisé par le très grand nombre d'observations sur lequel se base cette étude, autrement dit par une puissance statistique élevée.
L'un de ces facteurs de biais mineur est l'âge, avec une augmentation de la probabilité d'enregistrement avec un âge plus élevé. Ceci est cohérent avec la littérature : une étude indique le même gradient avec l'âge (Hauer and Hakkert 1988) ; une autre (James 1991) indique que les enfants sont moins souvent enregistrés, mais probablement du fait qu'ils sont plus souvent blessés à vélo. Ceci est confirmé dans notre étude : le risque relatif associé à l'âge est assez élevé en univarié mais devient faible en multivarié, lorsqu'on ajuste entre autres, sur le type d'usager. Le biais de sélection sur l'âge peut être dû à une plus grande fragilité attendue (de la part des impliqués et/ou des forces de l'ordre) envers les personnes plus âgées.
La probabilité légèrement plus faible d'enregistrement des blessés dans un accident de jour par rapport à un accident de nuit peut s'expliquer là aussi par une gravité attendue plus élevée pour les accidents se produisant la nuit.
La gravité attendue (par les impliqués mais surtout par les forces de l'ordre) est sans doute prise en compte en partie par la gravité « réelle » incluse dans le modèle mais en partie seulement, car ce n'est pas la même notion.
L'association entre le sexe et la probabilité d'enregistrement ne s'explique pas aisément, étant donné que la gravité lésionnelle et le type d'usager sont déjà pris en compte. La plupart des études n'ont pas mis en évidence un tel biais. Il est de toute façon très faible (RR=0,95).

Les catégories « non renseigné » sont associées à de plus faibles probabilités d'enregistrement. Ceci est sans doute un artefact : il faut rappeler que les variables utilisées ici pour caractériser les victimes (du Registre, et parmi celles-ci, celles qui sont identifiées dans les BAAC) sont les variables du Registre. Si des informations manquent il est plus difficile d'identifier la victime correspondante dans les fichiers des forces de l'ordre (si elle y est enregistrée), et donc le taux d'enregistrement est artificiellement plus faible. Les informations manquantes concernent essentiellement le type d'usager, le type de réseau, la zone de force de l'ordre, et le moment dans la journée/nuit. Des valeurs manquantes pour le type de réseau et la force de l'ordre indiquent en effet (en général) que le lieu d'accident est manquant, et le lieu d'accident manquant est effectivement un obstacle pour le chaînage.

Les biais de sélection signifient que les données des forces de l'ordre donnent une image déformée des victimes d'accidents de la route. Les blessés dans un accident sans tiers sont moins souvent enregistrés que ceux dans un accident avec tiers, et encore bien moins parmi les blessés usagers de 2RM ou cyclistes, ce qui conduit à des comparaisons faussées, en multivarié et a fortiori en univarié. De la même façon le biais de sélection mis en évidence sur le type de réseau, la force de l'ordre et l'environnement (Grand Lyon vs hors Grand Lyon) implique que les comparaisons urbain/rural ou entre régions géographiques basées sur les données des forces de l'ordre sont faussées (que ce soit des comparaisons en analyse multivariée ou univariée). Les comparaisons entre hommes et femmes ou entre classes d'âge sont aussi faussées (en analyse univariée) car même s'il y a des biais de sélection (mineurs) sur l'âge et le sexe, ces caractéristiques sont très liées avec le type d'usager, qui est un

facteur majeur de biais de sélection. Elles ne seront pas (ou peu) faussées dans une analyse multivariée qui tiendrait compte du type d'usager.

L'hypothèse sous-jacente de l'étude est que chaque type de force de l'ordre enregistre de la même façon les blessés sur toute la France, autrement dit que les pratiques des forces de l'ordre d'enregistrement sont les mêmes partout en France, pour une force de l'ordre donnée. Cette hypothèse repose sur le fait que chaque force de l'ordre a une structure centralisée et reçoit des instructions nationales. Sous cette hypothèse, les résultats de cette étude, basés sur les données du département du Rhône sont généralisables à la France entière.

Résumé

Contexte :
Les données policières d'accidentologie, qui constituent la base des connaissances épidémiologiques sur les victimes de la circulation routière dans la plupart des pays, sont incomplètes et biaisées. Dans le Rhône, une deuxième source de recensement existe depuis 1995 : le Registre des victimes d'accidents de la circulation routière. En prenant les données du Registre comme référence, nous quantifions le sous-enregistrement des données des forces de l'ordre, nous identifions et quantifions les biais de sélection.

Méthode :
Nous comparons les données des forces de l'ordre aux données du Registre après identification par chaînage probabiliste des blessés communs. L'étude porte sur la période 1997-2001, et sur 59 714 blessés. Une analyse multivariée est conduite, qui modélise, parmi les blessés du Registre, la probabilité d'être (aussi) enregistré dans les données des forces de l'ordre, en fonction de caractéristiques du blessé et de l'accident. Les résultats sont exprimés sous forme de risques relatifs (RR) ajustés et de probabilités ajustées.

Résultats :
Le taux d'enregistrement des forces de l'ordre, par rapport à l'ensemble formé par la réunion des données des forces de l'ordre et du Registre, est de 37,7 % (IC à 95% : 37,3% -38,1%). La probabilité d'enregistrement dans les données des forces de l'ordre varie en fonction inverse de la gravité lésionnelle : par rapport aux blessés NISS 25-75 (gravité critique), les victimes légèrement blessées (NISS 1-3) ont 2,6 fois moins de chances d'être enregistrées dans les données des forces de l'ordre. La probabilité varie selon le type d'usager et la présence/absence de tiers, avec une interaction entre les deux. Parmi les accidents avec tiers, les occupants de camions, camionnettes, ou bus sont les mieux enregistrés (probabilité ajustée=52,7 %) ; viennent ensuite, au même niveau, les automobilistes, piétons et usagers de 2RM (probabilité ajustée autour de 45 %), et enfin les cyclistes (probabilité ajustée=33,6 %). Parmi les accidents sans tiers, les automobilistes sont les mieux enregistrés (probabilité ajustée=34,9 %) ; viennent ensuite les occupants de camions, camionnettes, bus (probabilité ajustée=25,5 %), puis les usagers de 2RM (probabilité ajustée=14,7 %) et enfin les cyclistes (probabilité ajustée=2,1 %). La probabilité d'enregistrement varie aussi selon la combinaison du type de réseau, de l'environnement (Grand Lyon vs hors Grand Lyon) et du type de force de l'ordre, les trois étant structurellement liés.

Conclusions :
Les données des forces de l'ordre ne sont pas représentatives de l'ensemble des victimes d'accidents de la circulation routière. Les facteurs de biais de sélection sont nombreux et l'ampleur de ces biais peut être grande.

1.2 Enquête auprès des blessés identifiés dans le Registre seulement

Introduction

Cette étude complémentaire a pour but de mieux comprendre le sous-enregistrement des blessés dans les BAAC, en distinguant notamment le sous-enregistrement du fait des impliqués dans l’accident, et le sous-enregistrement du fait des forces de l’ordre. Les impliqués ont obligation de déclarer auprès des forces de l’ordre tout accident corporel de la circulation routière (article R231-1 du Code de la route). Aucune sanction pénale n’est prévue, et cette obligation semble très peu connue. Les forces de l’ordre ont obligation d’établir un procès-verbal pour tout accident corporel, dès lors qu’elles en ont connaissance.

Une seconde raison possible de l’identification de blessés uniquement par le Registre serait un manque d’efficacité de la procédure de chaînage, suite à des données manquantes ou erronées.

Il s’agit d’une enquête auprès des blessés identifiés dans le Registre seulement. Les personnes sont contactées par courrier, et il leur est demandé de remplir un court questionnaire.

Matériel

Population d’étude

Population cible

La population cible est celle des blessés enregistrés par le Registre, mais non identifiés dans les BAAC à travers la procédure de chaînage, et qui satisfont aux critères d’enregistrement des BAAC (puisque le but est de comprendre pourquoi ils ne sont pas dans les données des forces de l’ordre).
Cela signifie que les usagers de rollers et skate-boards se blessant seuls ou en collision avec un piéton sont exclus (5% des blessés du Registre).

Population source

Nous nous restreignons aux blessés impliqués dans un accident ayant eu lieu en 2001 (effectif=7756). Il s’agit de ne pas solliciter les personnes trop longtemps après l’accident (enquête réalisée en 2004).

Les personnes suivantes sont exclues de la population source pour des raisons pratiques :
-les personnes décédées (n=25) : il n’est pas envisagé de solliciter un proche. L’information sur le décès est celle fournie par le Registre, avec un suivi limité à la période de soins liés à l’accident.
-les blessés ne résidant pas en France (n=52 et 142 d’adresse inconnue) : il s’agit simplement d’éviter des difficultés de langue.
-les blessés sans lésion décrite (n=118) : nous avons en effet besoin de connaitre la gravité lésionnelle ; c’est un des paramètres du plan de sondage.
-les blessés pour lesquels l’année de naissance est inconnue (n=16) : en effet, le questionnaire et la lettre sont fonction de l’âge de la personne sollicitée.

La population source inclut 7412 blessés.

Méthode

Objectifs détaillés

Les objectifs sont de quantifier et de caractériser les sous-groupes suivants :
- les forces de l’ordre n’ont pas été alertées,
- elles ont été alertées mais ne sont pas intervenues,
- elles sont intervenues mais n’ont pas rédigé de procès-verbal,
- elles sont intervenues et ont rédigé un procès-verbal, mais elles n’ont pas enregistré le BAAC correspondant au PV,
- un enregistrement BAAC a été réalisé, mais des informations erronées ou incomplètes ont rendu le chaînage avec le Registre inefficace.

(NB : il n’est cependant pas possible des distinguer les deux derniers sous-groupes par cette enquête auprès des blessés).
Nous voulons caractériser les sous-groupes ci-dessus en fonction des principaux facteurs de biais de sélection : la gravité lésionnelle, le type d’usager, la présence de tiers (oui/non). Lorsque cette enquête a été mise en place, la gravité lésionnelle d’une victime était mesurée par le MAIS.

Nous voulons aussi évaluer :
- les causes de non-alerte des forces de l’ordre, et notamment leur répartition,
- les taux d’erreurs des variables de chaînage.

Nous voulons mieux comprendre les causes de non-alerte des forces de l’ordre ; nous envisageons les différentes situations :
- personne n’a pensé à le faire,
- parce qu’il n’y avait pas de problème de responsabilité (pas de tiers, ou présence de tiers mais constat amiable rempli),
- parce qu’il n’y avait apparemment pas de blessures graves,
- à cause d’infraction(s) au code de la route (et donc par crainte de sanction) ; afin de ne pas rencontrer un fort taux de non-réponse, ceci a été reformulé selon les termes « réticence à faire appel aux forces de l’ordre ».

Nous voulons estimer le taux d’erreur des variables utilisées dans le chaînage, à savoir, pour le blessé : année, mois de naissance, sexe, type d’usager ; pour l’antagoniste : type d’usager ; pour l’accident : année, mois, jour, heure, et localisation (commune). Ce que l’on va mesurer c’est en fait un taux de discordance entre l’information disponible dans le Registre et l’information donnée par le blessé dans le questionnaire.

Construction de l’échantillon (plan de sondage)

Nous imposons comme taille de l’échantillon un effectif total de l’ordre de 800 à 1000 individus, pour que cette enquête donne des estimateurs avec des précisions du même ordre de grandeur que l’étude symétrique « BAAC seulement » (cf. section 1.3) où la population source est limitée à 1300 personnes.

Deux des objectifs de cette enquête sont l’estimation de proportions globales i.e. sur l’ensemble des blessés identifiés par le Registre et non identifiés dans les BAAC. Un troisième objectif de l’enquête est de caractériser les sous-groupes : forces de l’ordre alertées / présentes / avec PV établi, en fonction de la gravité, du type d’usager et de tiers (oui / non). Inversement, nous pouvons travailler sur les sous-groupes de MAIS, de type d’usager et de la présence / absence de tiers, et les caractériser en fonction des proportions d’alerte des forces de l’ordre, de présence des forces de l’ordre et de procès-verbaux établis.

Nous avons tout intérêt à mettre en œuvre un plan de sondage stratifié. En effet, ce type de plan de sondage permet d’augmenter la précision de l’estimateur grâce à des strates homogènes vis-à-vis du caractère étudié ; cela implique que la variable de stratification soit liée à la variable étudiée (ce qui est le cas). Par ailleurs, les résultats par strate sont intéressants en soi.

Nous décidons donc de réaliser un plan de sondage stratifié, en fonction des facteurs majeurs de biais de sélection : le type d’usager, l’implication de tiers (oui/non), et la gravité. Le type d’usager est regroupé en cinq catégories : les quatre les plus fréquentes (piétons / cyclistes / usagers de 2RM / automobilistes) et une catégorie « autre ». La gravité lésionnelle est mesurée par le MAIS regroupé en deux catégories : MAIS 1 et MAIS 2+ (pour des questions d’effectifs).

Il faut s’assurer que les effectifs à la marge soient suffisants (pour avoir une bonne précision), c’est-à-dire, pour les deux catégories de MAIS, pour les types d’usagers piétons / cyclistes / usagers de 2RM / automobilistes (la catégorie autre n’étant pas intéressante en soi car trop hétérogène), et pour les catégories avec ou sans tiers. Il y a là une hiérarchie des objectifs : il nous importe plus d’avoir une bonne précision des proportions par niveau de MAIS, que selon le type d’usager ou l’absence / présence de tiers. La précision dans une strate est celle d’un sondage aléatoire simple. Pour une strate donnée, il faut, pour estimer une proportion de l’ordre de 50 % (situation la plus coûteuse), un effectif de 400 pour obtenir une précision à 5 points (de pourcentage).

Nous choisissons de répartir la taille de l’échantillon selon une allocation proportionnelle (aux effectifs observés), mais en imposant le même effectif dans les strates MAIS 1 et MAIS 2+, d’au moins 400 pour avoir une précision correcte. En d’autres termes, dans chacune des deux strates MAIS 1 et MAIS 2+, nous fixons l’effectif total à 400 sujets, et nous réalisons un tirage selon une répartition proportionnelle (aux effectifs observés dans la population dans la strate considérée).

Les effectifs observés sont détaillés dans le Tableau 2.

Tableau 2 : répartition des effectifs observés des blessés selon le type d’usager, l’absence/présence de tiers, et la gravité lésionnelle, Rhône, 2001, Registre seulement.

plan de sondage	piétons*	cyclistes		usagers de 2RM		automobilistes		autres		total
tiers	oui	oui	non	oui	non	oui	non	oui	non	
MAIS 1	391	177	605	450	672	2420	844	57	256	5872
MAIS 2+	124	49	311	156	328	254	220	11	87	1540
total	515	226	916	606	1000	2674	1064	68	343	7412

* rappelons que les accidents de piétons sans tiers ne sont pas considérés comme des accidents de la route

Le plan de sondage proposé est donné dans le Tableau 3.

Tableau 3: répartition des effectifs de blessés selon un plan de sondage stratifié et proportionnel, pour l’enquête « Registre seulement »,

plan de sondage	piétons	cyclistes		usagers de 2RM		automobilistes		autres		total
tiers	oui	oui	non	oui	non	oui	non	oui	non	
MAIS 1	27	12	41	31	46	165	57	4	17	400
MAIS 2+	32	13	81	41	85	66	57	3	23	400
total	59	25	122	71	131	231	115	7	40	800

Nous vérifions que ce plan de sondage nous assure un effectif supérieur à 50 dans les quatre strates « à la marge » des principaux types d’usager, ainsi que dans les deux strates « à la marge » avec ou sans tiers.

Par ailleurs, nous gonflons les effectifs de tirage en appliquant un taux de réponse de 30 % chez les blessés MAIS 1 et de 50 % chez les blessés MAIS 2+ (taux estimés sur un échantillon de 80 personnes de test du questionnaire, cf. ci-dessous) ; nous obtenons donc le plan de sondage décrit dans le Tableau 4.

Tableau 4 : répartition des effectifs de blessés selon un plan de sondage stratifié et proportionnel, avec gonflement des effectifs tenant compte du taux de réponse attendu, pour l’enquête « Registre seulement »

	piétons	cyclistes		usagers de 2RM		automobilistes		autres		total
tiers	oui	oui	non	oui	non	oui	non	oui	non	
MAIS 1	89	41	138	103	153	550	192	13	59	1338
MAIS 2+	65	26	162	82	171	132	115	6	46	805
total	154	67	300	185	324	682	307	19	105	2143

Recueil des données par questionnaire

Nous envoyons un auto-questionnaire limité à un recto-verso, accompagné d’une lettre introductive. Nous incluons des enveloppes T (prépayées) pour la réponse. Les questionnaires ont été envoyés en avril-mai 2004. Une relance a été effectuée un mois après.

Les enfants n’étant pas exclus de ces études, le questionnaire et la lettre sont alors adressés aux parents. Nous distinguons les adolescents des enfants, et pour les adolescents, et nous adressons la lettre aux parents, en leur demandant de laisser à l’adolescent le soin de remplir lui-même le questionnaire. Nous avons fixé la limite d’âge entre ces deux groupes à 14 ans, âge à la date de l’étude.

Les lettres et questionnaires sont légèrement différents selon la gravité des lésions de la personne sollicitée. Il nous a en effet paru important de solliciter différemment les personnes les plus gravement atteintes. Des exemples de lettre et de questionnaire sont fournis en annexe.

Afin de vérifier la validité de l’information contenue dans le Registre (pour les variables utilisées dans le chaînage), nous demandons à nouveau aux personnes toutes ces informations, plutôt que de les donner et de demander confirmation (en effet, si les personnes ne se rappellent plus très bien de la date d’accident par exemple, elles auront tendance à accepter la date d’accident qu’on leur indique ; nous voulons éviter cette situation).

Test des questionnaires

Avant de lancer cette étude, le questionnaire est testé sur un petit échantillon afin de s’assurer que les questions sont bien comprises. Concernant les questions ouvertes (lieu de l’accident, date et heure d’accident), il faut s’assurer que l’on obtient des réponses suffisamment précises. Concernant les questions fermées, il faut regarder si certaines questions ont un taux de réponse très faible (ce qui pourrait indiquer une question mal comprise), et vérifier que les réponses « autre » ne sont pas trop nombreuses, ce qui pourrait indiquer une situation fréquente que nous aurions oubliée.

Nous avons tiré au sort 80 personnes. Les lettres et questionnaires étant légèrement différents selon la gravité des blessures, et l’âge, nous avons stratifié sur le MAIS et sur les classes d’âge (adultes adolescents, enfants).

Le test du questionnaire a eu lieu en janvier-février 2004.

Il est apparu que certaines formulations n’étaient pas très claires. Cela concerne les raisons possibles de non-alerte des forces de l’ordre, et en particulier lorsqu’il n’a pas été jugé utile de les appeler. Nous remplaçons la formulation « il n’y avait pas de problème de responsabilité » par « il n’y avait pas d’autre personne impliquée dans l’accident (pas d’antagoniste) ». Nous remplaçons les deux formulations « personne n’avait de blessures visibles dans l’accident » et « personne n’avait de blessures graves » par une seule formulation « personne ne paraissait avoir de blessure grave ».

Analyse statistique

Il s’agit d’une analyse descriptive. Les estimations de proportions et d’effectifs sont accompagnées de leur intervalle de confiance à 95 %. L’analyse prend en compte le plan de sondage stratifié. En pratique nous utilisons pour cela les procédures SURVEYFREQ, SURVEYMEANS et SURVEYLOGISTIC du logiciel SAS (détails présentés en annexe).

Nous explorons d’abord les éventuelles différences entre les répondants et les non-répondants. Étant donné que nous trouvons (cf. résultats) des différences significatives entre ces deux groupes, nous ne pouvons pas généraliser les observations sur les répondants à l’ensemble de la population cible, à savoir les blessés identifiés dans le Registre seulement. Cela signifie en particulier que les estimations des effectifs des différents sous-groupes (forces de l’ordre non-alertées, non-présentes et PV non rédigé) au niveau de la population cible ne seraient pas pertinentes. Nous nous limitons donc à des estimations de proportions : celle d’alertes des forces de l’ordre, celle de PV rédigés, (en fonction du type d’usager, de la présence/absence de tiers, du MAIS), et celles des différentes raisons de ne pas avoir alerté les forces de l’ordre.

Résultats

Répartition des retours de questionnaires

Un certain nombre de courriers ne sont pas parvenus à destination pour cause d’adresse manquante ou invalide (soit elle était erronée soit les personnes ont déménagé depuis).
Parmi les courriers arrivés à destination (que l’on peut supposer comme tels car non-retournés avec la mention « NPAI » signifiant « N’habite Pas à l’Adresse Indiquée ») le taux de réponse est de 39,4 %.

Le tableau ci-dessous donne le détail des retours.

Tableau 5 : répartition des retours de questionnaires, enquête « Registre seulement »

retours	effectif	proportion
questionnaire complété	547	25,3 %
questionnaire non complété	840	39,8 %
« non concerné(e) »	*17*	*0,8 %*
refus	*3*	*0,1 %*
décédé(e)	*2*	*0,1 %*
aucune réponse	*818*	*38,8 %*
adresse invalide	756	34,9 %
nom et/ou adresse non renseigné	*304*	*14,0 %*
« n’habite pas à l’adresse indiquée »	*452*	*21,0 %*
total	2143	100,0 %

Comparaison des répondants et des non-répondants

Nous avons comparé les répondants aux non-répondants au sens large (c’est-à-dire incluant les personnes qui ne répondent pas et les personnes avec adresse non renseignée ou invalide).
L’analyse multivariée, incluant les variables significatives (à 5 %) en analyse univariée, indique que les répondants diffèrent des non-répondants sur les caractéristiques suivantes : âge, sexe, gravité lésionnelle, type de réseau où s’est produit l’accident, port du casque (usagers de 2RM) ou de la ceinture lors de l’accident, et lieu d’habitation. Plus précisément, les non-répondants (et ceux avec adresse invalide) sont plus souvent de sexe masculin, âgés de 18 à 24 ans, blessés légers ou modérés (MAIS 1 ou 2), ne portant pas la ceinture ou le casque lors de l’accident, blessés dans un accident sur le réseau secondaire (voies communales) et habitant dans une zone urbaine plutôt que rurale.

Blessés s’avérant hors critères BAAC

Parmi les questionnaires complétés (547), il s’avère que quelques accidents ne satisfont pas aux critères BAAC. Ces blessés hors critères BAAC représentent 7,8 % des répondants.

Tableau 6 : répartition des réponses à l’enquête « Registre seulement » en fonction des critères BAAC (parmi les 547 questionnaires complétés)

	effectif	proportion	IC à 95 %	
questionnaires complétés et exploitables	495			
confusion sur l’accident ou sur la personne	3			
hors critères BAAC :	49	7,8 %	6,0 %	10,0 %
hors Rhône	*13*	*2,0 %*	*1,0 %*	*3,0 %*
hors réseau	*16*	*2,4 %*	*1,2 %*	*3,7 %*
hors « accident de la route »	*8*	*1,2 %*	*0,5 %*	*2,0 %*
trottinette seule	*7*	*1,1 %*	*0,4 %*	*1,8 %*
autre, divers	*5*	*1,0 %*	*0,1 %*	*2,1 %*
total	547	100,0 %	100,0 %	100,0 %

La plupart de ces blessés «hors critères BAAC » correspondent à des accidents qui se sont produits en dehors du département du Rhône, ou clairement hors réseau (typiquement cour d’école, jardin autour de la maison, terrain de sports).

Les blessés « hors accident de la route » correspondent à :
- deux accidents du travail exclusivement : déchargement d’un véhicule, accident de tracteur dans la vigne
- deux accidents de loisirs exclusivement : accident de cheval dans un centre équestre, à pied dans un stade
- deux accidents de la vie courante exclusivement : dans la maison, chute sur la voie publique,

- et enfin un accident dans le métro et une douleur sans accident de la route.

La catégorie « autre, divers » correspond à des accidents qui sont à la limite du « hors Rhône », ou à la limite de la définition d’un accident de la circulation routière.

Alerte et intervention des forces de l’ordre

Alerte des forces de l’ordre

Le Tableau 7 donne la répartition des répondants selon qu’ils aient alerté ou non les forces de l’ordre.

Tableau 7 : proportion d’alerte des forces de l’ordre
(parmi les 495 questionnaires exploitables de l’enquête « Registre seulement »)

alerte des forces de l’ordre ?	proportion	IC à 95 %	
oui	47,3 %	43,5 %	51,1 %
non	39,4 %	35,7 %	43,1 %
non-renseigné ou ne sait pas	13,3 %	10,6 %	16,0 %

Parmi les répondants, une petite moitié dit avoir alerté les forces de l’ordre.

Le Tableau 8 donne la proportion de ceux qui ont alerté les forces de l’ordre en fonction du type d’usager, de la présence/absence d’un tiers dans l’accident et de la gravité des blessures.

Tableau 8 : proportion d’alerte des forces de l’ordre, en fonction du type d’usager, de la présence/absence de tiers et de la gravité lésionnelle
(parmi les 495 questionnaires exploitables de l’enquête « Registre seulement »)

alerte des forces de l’ordre=oui	proportion	IC à 95 %	
par type d’usager			
piétons	43,4 %	27,4 %	59,4 %
cyclistes	10,7 %	5,4 %	16,0 %
usagers de 2RM	42,9 %	34,6 %	51,2 %
automobilistes	60,2 %	54,6 %	65,8 %
autres et non renseigné	49,9 %	24,6 %	75,2 %
type d’accident			
avec tiers	55,3 %	50,2 %	60,4 %
sans tiers	34,7 %	29,3 %	40,1 %
gravité lésionnelle			
MAIS 1	47,8 %	43,2 %	52,4 %
MAIS 2+	45,3 %	40,7 %	50,0 %

La proportion d’alerte des forces de l’ordre varie fortement selon le type d’usager. Seuls 11 % des cyclistes blessés les ont alertées alors que 60 % des automobilistes blessés l’ont fait. Les blessés dans un accident avec tiers les ont alertées plus souvent que les blessés dans un accident sans tiers. On ne note pas de différence du taux d’alerte entre les blessés légers (MAIS 1) et les blessés modérés et graves (MAIS 2+).

Présence des forces de l’ordre

Une fois que les forces de l’ordre ont été alertées, nous observons qu’elles sont toujours venues sur les lieux de l’accident (sauf quelques rares exceptions). Il n’y a donc pas lieu de distinguer la situation d’alerte/non alerte de celle de la présence/non présence des forces de l’ordre.

Procès-verbal rédigé

Les tableaux ci–dessous donnent, parmi les répondants pour lesquels les forces de l’ordre étaient présentes, la proportion de procès-verbaux établis.

Tableau 9 : taux de procès-verbaux apparemment établis
(parmi les 270 répondants de l’enquête « Registre seulement » avec présence des forces de l’ordre)

procès-verbal établi ?	proportion	IC à 95 %	
oui	49,6%	44,4%	54,9%
non	23,6%	18,9%	28,2%
non-renseigné ou ne sait pas	26,8%	22,0%	31,6%

Parmi les répondants avec présence des forces de l’ordre sur les lieux de l’accident, la moitié estiment qu’un procès-verbal a été rédigé ; environ un quart ne répond pas à cette question.

Le tableau ci-dessous donne la proportion de PV apparemment établis, selon le type d’usager, la présence / absence de tiers et la gravité lésionnelle, parmi les répondants pour lesquels les forces de l’ordre étaient présentes.

Tableau 10 : taux de procès-verbaux apparemment établis, en fonction du type d’usager, de la présence/absence de tiers et de la gravité lésionnelle
(parmi les 270 répondants de l’enquête « Registre seulement » avec présence des forces de l’ordre)

procès-verbal établi=oui	proportion	IC à 95 %	
par type d’usager			
piétons	85,8%	76,9%	94,6%
cyclistes	35,2%	11,5%	58,9%
usagers de 2RM	54,8%	43,5%	66,2%
automobilistes	46,8%	40,0%	53,7%
autres et non renseigné	40,7%	9,9%	71,5%
type d’accident			
avec tiers	55,2%	48,9%	61,6%
sans tiers	36,4%	27,0%	45,8%
gravité lésionnelle			
MAIS 1	48,3%	41,9%	54,6%
MAIS 2+	55,2%	48,2%	62,2%

La proportion de PV apparemment établis est très différente selon le type d’usager : elle varie entre 35% de PV établis chez les cyclistes blessés, et jusqu’à 85% de PV établis chez les piétons blessés, parmi les répondants avec présence des forces de l’ordre. La forte proportion chez les piétons est en partie due à ce qu’ils sont par définition des accidents avec tiers. La proportion de PV apparemment établis est en effet différente selon que l’accident impliquait un tiers ou non : 55% versus 36%. La proportion de PV apparemment établis semble différente selon la gravité des blessés : 55% de PV établis parmi les blessés MAIS2+ et 48% parmi les blessés MAIS1.

Le Tableau 8 mesure les biais d’enregistrement (dans les données des forces de l’ordre) de la part des blessés, alors que le Tableau 10 ci-dessus mesure les biais d’enregistrement de la part des forces de l’ordre.

Le tableau ci–dessous donne la répartition sur l’ensemble des répondants selon qu’un procès-verbal a été établi ou non.

Tableau 11 : taux de procès-verbaux apparemment établis
(parmi les 495 questionnaires exploitables de l’enquête « Registre seulement »)

procès-verbal établi ?	proportion	IC à 95 %	
oui	27,6 %	24,1 %	31,1 %
non	14,6 %	11,7 %	17,4 %
non-renseigné ou ne sait pas	58,8 %	54,0 %	61,6 %

Parmi l’ensemble des répondants, plus d’un sur quatre estime que les forces de l’ordre ont dressé un procès-verbal.

Raisons de non-alerte

Le Tableau 12 donne les proportions des différentes raisons de non alerte des forces de l’ordre.

Tableau 12 : distribution des raisons données pour NE PAS avoir alerté les forces de l’ordre
(parmi les 230 répondants dans ce cas, enquête « Registre seulement »)

	n	proportion	IC à 95 %	
cela n’a pas été jugé utile	167	74,7 %	69,7 %	79,7 %
car :				
il n’y avait pas d’autre personne impliquée dans l’accident	79	32,5 %	28,1 %	36,9 %
un constat amiable d’accident avait été rempli	46	23,4 %	19,4 %	27,4 %
personne ne paraissait avoir de blessure grave	66	30,8 %	25,2 %	36,4 %
personne n’a pensé à le faire	22	8,5 %	5,3 %	11,6 %
réticence à les appeler	8	3,5 %	1,3 %	5,7 %
autre raison	9	4,4 %	1,8 %	6,9 %
« vous ne savez pas »	19	7,1 %	4,3 %	9,9 %

La somme dépasse 100 % car il était possible de donner plusieurs réponses

Si les forces de l’ordre n’ont pas été alertées c’est principalement (74,7 % des non-alertes) parce que cela n’a pas été jugé utile.

Taux de discordance sur les variables de chaînage

Le tableau ci-dessous donne les taux de discordance entre les informations dans le Registre et les informations fournies par les répondants au questionnaire, pour les variables utilisées dans le chaînage entre le Registre et les données des forces de l’ordre.

Tableau 13 : discordance entre les informations du Registre et les réponses au questionnaire, enquête « Registre seulement » (parmi les 495 questionnaires exploitables)

discordance sur :	effectif	proportion
mois de naissance	7	1,4 %
année de naissance	18	3,6 %
jour d’accident	82	16,6 %
jour d’accident (>1 jour)[a]	56	11,3 %
mois d’accident	11	2,2 %
année d’accident	8	1,6 %
heure d’accident	163	32,9 %
heure d’accident (>1 h)[b]	54	10,9 %
sexe	4	0,8 %
type d'usager	13	2,6 %
commune (lieu d'accident)	23	4,6 %
communes discordantes et non adjacentes	7	1,4 %

a) discordance sur jour accident au-delà d’un jour
b) discordance sur heure accident au-delà d’une heure

Le taux de discordance ne correspond pas nécessairement au taux d’erreur des variables du Registre, car il n’est pas assuré que les réponses au questionnaire soient correctes, d’autant moins que ces réponses sont données quatre ans après l’accident.

Discussion

Les non-répondants (incluant ceux avec adresse manquante ou invalide) diffèrent des répondants : ils sont plus souvent de sexe masculin, âgés de 18 à 24 ans, blessés légers ou modérés, portaient moins souvent leur ceinture ou leur casque lors de l’accident, et habitent plutôt en zone urbaine.
Du fait de cette différence, nous ne pouvons pas généraliser les résultats basés sur les répondants à l’ensemble des blessés identifiés dans le Registre seulement. En outre, nous pouvons supposer que parmi les non-répondants, la proportion d’alerte des forces de l’ordre est plus faible et que les raisons de non-alerte diffèrent, avec sans doute une plus grande proportion de personnes réticentes à les alerter, notamment pour non-respect de la réglementation. En effet, les non-répondants portaient bien moins souvent leur ceinture ou leur casque[1] lors de l’accident, ce qui laisse supposer qu’ils sont moins enclins en général à respecter la réglementation, et donc à alerter les forces de l’ordre. De plus, s’ils étaient réticents à alerter les forces de l’ordre lors de l’accident, il se peut qu’ils aient aussi été réticents à donner leur adresse pour la fiche de notification du Registre et plus tard à répondre à notre questionnaire (quelle que soit sa formulation et l’assurance donnée dans la lettre d’accompagnement que les informations ne seraient transmises à quiconque). La proportion d’alerte des forces de l’ordre, parmi l’ensemble des blessés identifiés dans le Registre seulement, doit donc être plus faible que celle observée chez les répondants. Cette proportion est de 47,3%, si on fait l’hypothèse, assez raisonnable, que les non-réponses à cette question (alerte des forces de l’ordre ?), 13%, correspondent à « alerte=non ».
Il en découle que le taux de procès-verbaux établis parmi l’ensemble des blessés identifiés dans le Registre seulement est vraisemblablement inférieur à celui observé chez les répondants. La proportion de répondants indiquant un procès-verbal établi est de 27,6%. Parmi les répondants avec présence des forces de l’ordre sur les lieux de l’accident (environ la moitié des répondants), la proportion de ceux qui estiment qu’un PV a été établi est de 49,6%, mais la proportion de ceux qui ne savent pas ou ne répondent pas est de 26,8%. Il n’est pas plausible de supposer ici que les non-réponses à cette question correspondent à « PV établi=non ». Les non-réponses peuvent ici souvent correspondre à la situation où les personnes ne savent réellement pas si un PV a été établi ou pas.

[1] Port du casque exploré chez les usager de 2RM uniquement

Si on généralisait les résultats de l’enquête à la population cible, on pourrait appliquer le taux de 27,6 % de procès-verbaux établis à l’ensemble des blessés identifiés dans le Registre seulement (accidentés en 2001) ; son effectif est de 7823, on obtiendrait alors 2159 blessés de « Registre seulement » avec un PV établi. On devrait donc les retrouver dans les données des forces de l’ordre. Or le nombre de blessés identifiés dans les BAAC seulement (c’est-à-dire non encore reliés à un blessé du Registre) pour la même année est de 1319, soit un effectif bien inférieur. Cela indique que la proportion de PV établis est bien plus faible que 27,6% ou alors qu’une proportion non négligeable de PV n’ont pas été enregistrés sous forme de BAAC.

Les raisons données de non-alerte et les proportions correspondantes n’ont rien de surprenant. Les résultats sont cohérents avec la seule enquête similaire que nous avons trouvée dans la littérature (Simpson 1996), dans laquelle les principales raisons de non-alerte sont que l’accident était bénin ou qu’il n’y avait pas de tiers impliqué dans l’accident. Il aurait été intéressant de prévoir dans l’enquête une question sur la connaissance par les impliqués de l’obligation de déclarer tout accident corporel aux forces de l’ordre.

Il s’avère que parmi les 547 répondants, 7,8 % ne satisfont pas aux critères des BAAC. Il s’agit de blessés dans des accidents hors Rhône, hors réseau, ou bien de trottinette seul, ou encore des accidents de travail, domestique ou de loisirs exclusivement.
Cela se produit alors que la population cible de l’enquête a été restreinte aux blessés satisfaisant aux critères BAAC (puisque le but était de comprendre pourquoi ces blessés n’étaient pas dans les données des forces de l’ordre). Il est apparu que l’information du Registre sur ces blessés était dans la plupart des cas incomplète, ce qui n’avait pas permis de les identifier comme hors critères. Cette proportion n’est pas négligeable et la question qui se pose est celle de sa généralisation à l’ensemble des blessés identifiés dans le Registre seulement : est-ce que sa valeur est la même chez les non-répondants ? Elle pourrait être plus élevée : cela pourrait être une cause de non-réponse à l’enquête (ne pas se sentir concerné par le questionnaire car ce n’était pas un accident de la route).

Le taux de discordance entre les informations contenues dans le Registre et les réponses aux questionnaires sur les variables servant au chaînage n’est pas vraiment exploitable. L’idée était d’estimer le taux d’erreur sur les variables de chaînage, afin d’estimer le nombre de faux négatifs (Howe 1998). Une autre méthode sera utilisée (cf. chap. 3).

Résumé

Contexte :
Il s’agit de mieux comprendre pourquoi un grand nombre de blessés ne sont recensés que par le Registre : combien n’ont pas alerté les forces de l’ordre ? Pour quelles raisons ? Combien les ont alertées mais sans que cela donne lieu à un procès-verbal ?

Méthode :
La population cible est l’ensemble des blessés identifiés dans le Registre seulement, restreint aux accidents de l’année 2001. Un échantillon de 2143 blessés, stratifié sur la gravité lésionnelle (mesurée par le MAIS), sur le type d’usager et sur l’absence / présence de tiers est constitué. Un auto-questionnaire est envoyé demandant aux personnes si elles ont alerté les forces de l’ordre suite à l’accident, et si non, pour quelles raisons.

Résultats :
35 % des blessés de l’échantillon avaient des adresses non renseignées ou invalides, 39 % n’ont pas répondu et 26 % ont répondu. Parmi les personnes qui ont pu être contactées, le taux de réponse est de 39 %. Cela correspond à 547 personnes.
Les non-répondants (incluant ceux avec adresse manquante ou invalide) diffèrent des répondants : ils sont plus souvent de sexe masculin, âgés de 18 à 24 ans, blessés légers ou modérés, portaient moins souvent leur ceinture ou leur casque lors de l’accident, et habitent plus souvent en zone urbaine.
Parmi les 547 répondants, il s’avère que 7,8 % (IC à 95% : 6,0% - 10,0%) ne satisfont pas aux critères des BAAC : accident hors Rhône, hors réseau, ou de trottinette seul, ou accident d’une autre nature exclusivement : accident de travail ou domestique ou de loisirs exclusivement.
Parmi les 495 répondants (avec un questionnaire exploitable), 47 % (IC à 95%: 43,5% – 51,1%) disent avoir alerté les forces de l’ordre, et 27,6 % (IC à 95% : 24,1% - 31,1%) estiment qu’un procès-verbal a été établi.

Conclusion :
Les répondants étant différents des non-répondants, nous ne pouvons pas généraliser les taux d’alerte des forces de l’ordre et de PV établis à l’ensemble des blessés identifiés seulement dans le Registre. Les non-répondants portaient moins souvent la ceinture (ou le casque pour les usagers de 2RM) lors de l’accident ; ils sont donc vraisemblablement moins enclins à respecter la réglementation routière que les répondants. Ainsi, la proportion d’alerte des forces de l’ordre est sans doute bien plus basse chez les non-répondants que chez les répondants, et donc plus basse que 47% dans l’ensemble des blessés identifiés seulement dans le Registre.

1.3 Étude des blessés identifiés dans les BAAC seulement

Introduction

Il s’agit de mieux comprendre pourquoi un certain nombre de blessés sont identifiés dans les BAAC et pas dans le Registre. Trois raisons sont envisageables : la première est la différence de définition d’inclusion des blessés : les BAAC incluent tous les blessés alors que le Registre n’inclut que ceux qui ont consulté une structure hospitalière. La deuxième raison possible est celle d’un dysfonctionnement du Registre, et la troisième un dysfonctionnement du chaînage entre le Registre et les BAAC.

Contrairement à l’enquête « Registre seulement », les blessés ne sont pas contactés. La consultation des procès-verbaux desquels sont issus les BAAC suffit à cette étude.

Matériel

Population d’étude

La population cible est l’ensemble des blessés des BAAC du Rhône non identifiés dans le Registre par la procédure de chaînage. La population source est restreinte aux blessés impliqués dans un accident ayant eu lieu en 2001. Son effectif est de 1319 blessés.

Accès aux procès-verbaux

Les procès-verbaux ont été fournis par TRANSPV (organisme qui centralise tous les PV d’accidents corporels pour les fournir aux compagnies d’assurance). Ils ont été transmis sous forme de fichiers électroniques images, qui sont des numérisations (« scan ») des procès-verbaux papiers.

Un procès-verbal correspond à un accident et à ses impliqués (décédés, blessés et indemnes). Il n’y a pas de numéro d’identifiant unique de l’accident, ni de la victime, ce qui aurait permis de faire aisément la correspondance entre un enregistrement BAAC (un accident et une victime) et le procès-verbal duquel il est issu. Pour établir la correspondance entre BAAC et PV au niveau de l’accident, nous avons utilisé les informations : date de l’accident, force de l’ordre ayant rédigé le procès-verbal, et numéro de PV, puis nous avons vérifié que le lieu d’accident concordait. Une fois le PV identifié, l’identification de la victime parmi les impliqués recensés par le PV se fait sur le mois et l’année de naissance, le sexe, et éventuellement la lettre identifiant les véhicules et la place dans le véhicule (en sachant que la lettre identifiant le véhicule est peu fiable, car souvent modifiée lors de la saisie des PV vers les BAAC). Nous rappelons que le jour de naissance, les nom et prénom ne sont pas disponibles dans les BAAC et ne peuvent donc pas servir à l’identification des blessés entre BAAC et PV.

Les fichiers contenant les procès-verbaux sont stockés sur une zone disque à accès très restreint. Les personnes qui y ont accès sont soumises aux règles de confidentialité des données.

Un certain nombre de procès-verbaux manquait (pour des raisons techniques : « scan » illisible ou « scan » non réalisé car le PV était encore en cours de rédaction). Il faut donc vérifier la comparabilité des blessés avec PV exploitable et blessés sans PV exploitable.

Méthode

La première question à laquelle cette étude souhaite répondre porte sur les critères d’inclusion du Registre, qui sont légèrement différents de ceux des BAAC : est-ce que les blessés identifiés dans les BAAC, mais pas dans le Registre, le sont tout simplement parce qu’ils auraient des blessures de gravité mineure, telle qu’elle n’aurait pas entraîné de consultation hospitalière ? Nous voulons donc

connaître la gravité des blessés et leur type de recours aux soins. Secondairement, nous explorons la distance du lieu de leur accident à l’hôpital, comme possible facteur de bais de sélection du Registre.

Caractéristiques étudiées

Recours aux soins

Parmi les blessés identifiés seulement dans les BAAC, nous voulons identifier et quantifier les sous-groupes suivants :
- les personnes qui sont en fait indemnes,
- les blessés qui n’ont pas reçu de soin,
- les blessés qui se sont soignés eux-mêmes, ou qui ont été soignés par un proche,
- les blessés qui sont allés consulter en médecine de ville,
- les blessés qui sont allés consulter à l’hôpital ou à la clinique (et qui, si l’hôpital est dans le Rhône ou dans ses proches alentours, auraient dû être recensés par le Registre).

Gravité des blessures

La gravité lésionnelle d’une victime est mesurée par le NISS (New Injury Severity Score). Il suffit pour cela de connaître les codes AIS de gravité des trois lésions les plus graves. Le codage est réalisé par deux médecins du Registre à partir des descriptions lésionnelles fournies par les certificats médicaux, souvent présents dans les procès-verbaux.

Distance à l’hôpital

Il s’agit de la distance du lieu d’accident à l’hôpital le plus proche. L’ensemble des 310 communes du Rhône a été classé en deux catégories : loin ou proche d’un hôpital, en fonction de la localisation géographique de la commune par rapport à celle de l’hôpital le plus proche disposant d’un service d’urgences et en fonction du relief où se trouve la commune (plaine ou monts du Lyonnais).

Le deuxième aspect de cette étude concerne l’efficacité de la procédure de chaînage entre les BAAC et le Registre. Quelle est la part des blessés identifiés seulement dans les BAAC et non identifiés dans le Registre à cause d’un manque d’efficacité de la procédure de chaînage ? i.e. à cause de valeurs manquantes ou erronées dans les variables utilisées pour le chaînage. Nous cherchons à évaluer les taux d’erreur de ces variables : date, heure, et lieu de l’accident, mois, année de naissance, sexe et type d’usager des blessés.
Nous procédons aussi à un chaînage supplémentaire entre les BAAC et le Registre en utilisant un complément d’information : éventuelles corrections et compléments d’information sur les variables utilisées dans le chaînage, et surtout utilisation des nom et prénom, que nous extrayons des procès-verbaux (côté Registre, ces variables sont renseignées depuis mars 2000). Le résultat de ce chaînage supplémentaire nous renseignera sur l’efficacité du chaînage effectué en routine.

Enfin, l’existence de blessés identifiés seulement dans les BAAC peut provenir d’un dysfonctionnement du Registre : fiches de notification non–établies pour cause d’oubli, ou de surcharge de travail des services hospitaliers, et/ou de non-identification des circonstances de blessures comme étant un accident de la circulation. La possibilité d’un dysfonctionnement de ce type pourra être traitée en fonction des réponses apportées aux deux questions précédentes.

Les données recueillies dans les procès-verbaux sont les suivantes :
- nom (nom de jeune fille et nom d’épouse si pertinent) et prénom,
- adresse postale,
- si renseigné, type de recours aux soins,
- si renseigné, description des blessures.

L’idée première était de réaliser une enquête postale auprès des blessés. Or il est apparu que les procès-verbaux contenaient un grand nombre d’informations médicales : recours aux soins (hôpital ou médecin indiqué), description des blessures, copie de certificats médicaux. Ces informations nous ont paru largement suffisantes. En outre, étant donné les différents délais, l’enquête auprès des blessés aurait eu lieu cinq ans après leur accident. Cela impliquait des difficultés à contacter les personnes (changement d’adresse), et des difficultés pour les personnes à se souvenir de l’accident. Nous avons donc décidé de ne pas faire l’enquête auprès des blessés, et de nous concentrer sur l’étude des procès-verbaux.

Analyse statistique

Comparabilité des blessés avec PV exploitable et blessés sans PV exploitable

Nous vérifions d’abord si les blessés diffèrent sur des caractéristiques qui pourraient être liées à la procédure de rédaction, ou d’envoi des procès-verbaux (telles que mois d’accident et force de l’ordre). En effet, nous pensons qu’un certain nombre de procès-verbaux n’ont pas été fournis dans le jeu de données des PV de 2001 du Rhône à cause du délai entre l’établissement du procès-verbal, sa numérisation et sa mise à disposition, et donc que les procès-verbaux d’accidents ayant eu lieu vers la fin de l’année 2001 ont été reportés dans le jeu de données des PV de 2002 (que nous n’avons pas pu obtenir malgré nos demandes).
Dans un deuxième temps, nous vérifions si les deux groupes de blessés (avec ou sans PV exploitable) diffèrent sur les caractéristiques d’intérêt, mesurables pour les deux groupes : gravité et distance à l’hôpital. Pour comparer la gravité des blessés (de « BAAC seulement ») avec PV exploitable à celle des blessés sans PV exploitable, seule la gravité au sens BAAC est disponible.
Nous présentons les distributions univariées, et les résultats du test du Chi-deux de comparaison de deux distributions.

Étude des caractéristiques d’intérêt : gravité, type de recours aux soins, distance à l’hôpital

Parmi les blessés avec procès-verbal exploitable, nous identifions et quantifions les blessés selon leur recours aux soins. Parmi les blessés dont le procès-verbal inclut un certificat médical informatif sur les blessures, nous présentons la distribution de leur gravité et nous la comparons à celles des blessés des BAAC qui ont également été identifiés dans le Registre. Parmi les blessés identifiés seulement dans les BAAC, nous comparons la gravité entre les blessés qui sont allés consulter en médecine de ville et ceux qui sont allés consulter en médecine hospitalière.
Nous présentons la distance du lieu d’accident à l’hôpital des blessés identifiés seulement dans les BAAC et nous la comparons à celle des blessés des BAAC qui ont aussi été identifiés dans le Registre. Parmi les blessés identifiés seulement dans les BAAC, nous comparons aussi la distance à l’hôpital entre les blessés qui sont allés à l’hôpital et ceux qui n’y sont pas allés.
La comparaison des caractéristiques entre les blessés identifiés seulement dans les BAAC et les blessés identifiés dans l’intersection BAAC-Registre est une façon d’analyser le « filtre » opéré par le Registre ; cela permet d’explorer les biais de sélection du Registre.

Estimation du taux d’erreur des variables BAAC utilisées dans la procédure de chaînage

Il s’agit des variables date d’accident, lieu d’accident, mois et année de naissance, sexe, type d’usager des victimes. Étant donné qu’il n’y a finalement pas eu d’enquête auprès des blessés eux-mêmes, mais uniquement retour aux procès-verbaux, il s’agit de comparer les valeurs dans les PV aux valeurs enregistrées dans les BAAC.

Chaînage supplémentaire

Grâce à l’information supplémentaire recueillie dans les PV, notamment les nom et prénom des blessés, nous avons lancé une nouvelle procédure de chaînage entre les blessés non rapprochés, autrement dit entre les blessés identifiés seulement dans les BAAC et les blessés identifiés seulement dans le Registre. Les noms et prénoms ne sont pas comparés tels quels mais ont été encodés afin que les comparaisons ne soient pas pénalisées par des erreurs d’orthographe, de saisie, ou de transcription pour les noms et prénoms en caractères non-latins. Le code Soundex a été utilisé (fonction SOUNDEX dans le logiciel SAS) ; il a été développé pour les noms anglo-saxons et il est donc optimisé pour ce type de noms mais il semble donner des résultats corrects sur d’autres types de noms européens (Quantin, Binquet et al. 2004). Dans la saisie des procès-verbaux, les noms d’épouse et de jeune fille ont été saisis le cas échéant. Dans le Registre, seul le nom d’épouse est disponible. Les comparaisons ont donc essentiellement porté sur le nom d’épouse.

Ce chaînage supplémentaire s’est basé sur les variables suivantes : date de l’accident, lieu de l’accident (commune + précision : lieu-dit, n° route ou nom de la rue), nom et prénom, mois et année de naissance du blessé. La stratégie consiste à faire une sélection de paires d’enregistrements en fonction de la concordance sur quelques-unes de ces variables puis, pour chacune des paires sélectionnées, de décider de rapprocher ou non en fonction de la concordance sur les autres variables d’identification :
1) même prénom (code Soundex) ET même nom (code Soundex)
2) même prénom (code Soundex) ET même année de naissance ET date accident +/- 3 jours
3) même prénom (code Soundex) ET même mois de naissance ET date accident +/- 3 jours
4) même nom (code Soundex) ET même mois de naissance ET date accident +/- 3 jours
5) même nom (code Soundex) ET même année de naissance ET date accident +/- 3 jours
6) inversion nom et prénom : prénom Registre = nom PV ET nom Registre = prénom PV (en code Soundex)
7) comme le n° 1 mais nom de jeune fille des PV comparé au nom d’épouse du Registre
8) comme les n° 2 à 5 mais au lieu de date d’accident +/-3 jours : différence de 29-31 jours (erreur de 1 mois)
9) comme les n° 2 à 5 mais au lieu de date d’accident +/-3 jours : jour et mois accident inversés
10) sans la date d’accident (puisque non-modifiée par la lecture des procès-verbaux) :
même nom (code Soundex) ET même mois de naissance ET même année de naissance
même prénom (code Soundex) ET même mois de naissance ET même année de naissance
11) sans nom ni prénom (puisque le code Soundex a été conçu pour les noms anglo-saxons)
date de l'accident +/- 3 jours, même mois ET année naissance
(les informations contenues dans les PV sont parfois différentes de celles des BAAC)

Le chaînage supplémentaire sur les noms et prénoms a été effectué par une seule personne, soumise aux règles de confidentialité ; le résultat du chaînage a été stocké sans les noms et prénoms.

Résultats

Population d’étude

Les blessés recensés par les BAAC et non identifiés dans le Registre sont au nombre de 1319, sur l’année 2001. Parmi les PV reçus, 872 correspondent à des accidents de ces 1319 victimes. Parmi ceux-ci, on note 4 cas où le PV n’apporte aucune information par rapport aux BAAC (PV réduit à une seule page). Finalement, 868 victimes ont un procès-verbal exploitable (soit 65,8 %), au sens où le blessé a pu être identifié dans le PV, et où le PV est informatif.
Un doute subsiste sur l’identification de 5 blessés (parmi ces 868), car il y a incohérence entre BAAC et PV sur l’année et le mois de naissance, principale information commune permettant d’identifier une personne une fois que le procès-verbal a été identifié.

Comparaison des blessés « BAAC seulement » avec PV exploitable et ceux sans PV exploitable

Nous comparons les blessés avec PV exploitable et sans PV exploitable en fonction du mois de l’accident.

Tableau 14 : répartition des blessés avec ou sans PV exploitable en fonction du mois de l’accident, étude « BAAC seulement »

	PV exploitable		PV non exploitable	
mois d’accident	n	proportion	n	proportion
janvier	67	7,7 %	28	6,2 %
février	74	8,5 %	9	2,0 %
mars	94	10,8 %	22	4,9 %
avril	94	10,8 %	15	3,3 %
mai	70	8,1 %	20	4,4 %
juin	88	10,1 %	20	4,4 %
juillet	113	13,0 %	14	3,1 %
aout	78	9,0 %	20	4,4 %
septembre	96	11,1 %	45	10,0 %
octobre	94	10,8 %	55	12,2 %
novembre	0	0,0 %	142	31,5 %
décembre	0	0,0 %	61	13,5 %
ensemble	868	100,0 %	451	100,0 %

Test du Chi-deux de comparaison des deux répartitions : $p<0,0001$

Les blessés avec PV exploitable correspondent plutôt à des accidents en début d’année. Notamment aucun PV n’a pu être exploité pour les accidents ayant eu lieu en novembre et décembre 2001.

Nous comparons les blessés avec ou sans PV exploitable en fonction de la force de l’ordre qui a rédigé le procès-verbal.

Tableau 15 : répartition des blessés avec ou sans PV exploitable en fonction de la force de l’ordre, étude « BAAC seulement »

	PV exploitable		PV non exploitable	
force de l’ordre	n	proportion	n	proportion
gendarmerie	235	27,1 %	146	32,4 %
CRS	158	18,2 %	47	10,4 %
police	475	54,7 %	258	57,2 %
ensemble	868	100,0 %	451	100,0 %

Test du Chi-deux de comparaison des deux répartitions : $p=0,0006$

Il y a une différence de répartition des forces de l’ordre desquelles sont issus les procès-verbaux, entre blessés avec ou sans PV exploitable. Les blessés avec PV exploitable comptent une plus grande proportion de PV issus des CRS que les blessés sans PV exploitable.

Nous comparons les blessés avec ou sans PV exploitable sur les caractéristiques d’intérêt : gravité et distance à l’hôpital.

Tableau 16 : répartition des blessés avec ou sans PV exploitable en fonction de la gravité (BAAC), étude « BAAC seulement »

gravité (BAAC)	PV exploitable		PV non exploitable	
	n	proportion	n	proportion
« blessé grave »	64	7,4 %	40	8,9 %
« blessé léger »	804	92,6 %	411	91,1 %
ensemble	868	100,0 %	451	100,0 %

Test du Chi-deux de comparaison des deux répartitions : p=0,3

Le test de comparaison des proportions de blessés graves entre les deux groupes de blessés, avec ou sans procès-verbal exploitable ne met pas en évidence de différence de gravité (classement effectué par les forces de l'ordre). Cependant la puissance associée à ce test est très faible (16%).

Tableau 17 : répartition des blessés avec ou sans PV exploitable en fonction de la distance lieu d'accident - hôpital, étude « BAAC seulement »

	PV exploitable		PV non exploitable	
lieu de l'accident :	effectif	proportion	effectif	proportion
proche d'un hôpital	782	90,1 %	405	89,8 %
loin d'un hôpital	86	9,9 %	46	10,2 %
	868	100,0 %	451	100,0 %

Test du Chi-deux de comparaison des deux répartitions : p=0,9

Il n'apparaît pas de différence en termes de distance entre le lieu de l'accident et l'hôpital le plus proche (avec service d'urgences), entre les deux groupes de blessés : avec ou sans PV exploitable.

Caractéristiques des blessés « BAAC seulement »

Type de recours aux soins

Parmi les 1319 blessés identifiés seulement dans les BAAC, les 868 blessés avec procès-verbal exploitable se répartissent en fonction du type de recours aux soins, de la manière suivante :

1) les personnes qui s'avèrent non blessées :
Nous dénombrons 36 personnes pour lesquelles le procès-verbal indique que la personne n'a pas été blessée. Un certificat médical le confirme pour 4 personnes (NISS 0). La plupart de ces personnes sont allées consulter. Il s'agit essentiellement de femmes enceintes ou de nourrissons gardés en observation, et de personnes en état de choc psychologique.

2) les blessés qui n'ont pas reçu de soin, et
3) les blessés qui se sont soignés eux-mêmes, ou par un proche :
Les procès-verbaux ne contiennent pas explicitement ce type d'information. Nous pouvons seulement identifier l'ensemble constitué de ces deux sous-groupes par non-appartenance aux autres sous-groupes. Cela correspond à 30 personnes.

4) les blessés qui ont consulté en médecine de ville :
D'après les procès-verbaux, 83 personnes ont consulté en médecine de ville (dont 5 personnes où est seulement indiquée l'intention de s'y rendre). Parmi ces 83 blessés, 34 seraient également allés à l'hôpital. Il y a donc 49 personnes qui ont consulté au plus en médecine de ville, dont 2 indemnes.
Parmi les 83 personnes ayant déclaré une consultation d'un médecin de ville, 54 (soit 65 %) présentent un certificat médical correspondant.

5) les blessés qui ont consulté en médecine hospitalière :
D'après les procès-verbaux, 776 personnes sont allées consulter dans un service hospitalier. Parmi elles, 19 ne sont pas blessées, donc le nombre de blessés ayant consulté à l'hôpital est de 757.

La majorité des personnes ayant consulté un service hospitalier ont un certificat médical correspondant, c’est-à-dire émis par un service hospitalier (n=557 soit 64 % de l’ensemble, IC à 95% : 61,0%-67,4%, et 72 % de ceux déclarant s’être rendus à l’hôpital, IC : 68,6% - 74,9%)

Tous les services hospitaliers cités ou émetteurs de certificat médical sont inclus dans Registre.

Le tableau suivant résume la répartition des blessés avec PV exploitable en groupes disjoints, en ne retenant que la consultation médicale la plus « lourde » et en séparant les blessés et les non-blessés.

Tableau 18 : répartition des 868 personnes avec PV, selon le statut de blessé/non-blessé et selon le recours aux soins (groupes disjoints), étude « BAAC seulement »

groupes disjoints	n	proportion	IC à 95 %
non-blessés	36	4,1 %	2,8 % - 5,5 %
blessés, mais pas de soin, ou auto-soin ou soin par un proche (au plus)	30	3,5 %	2,2 % - 4,7 %
blessés, et médecine de ville (au plus)	47	5,4 %	3,9 % - 6,9 %
blessés, et consultation hospitalière (au plus)	757	87,1 %	84,9 %– 89,3 %
total	868	100,0 %	

Le tableau ci-dessous donne simplement la répartition selon le type de recours aux soins.

Tableau 19 : répartition des 868 personnes avec PV, selon le recours aux soins, étude « BAAC seulement »

groupes non-disjoints	n	proportion	IC à 95 %
pas de consultation	43	5,0 %	3,5 % - 6,4 %
consultation médecine de ville	83	9,6 %	7,6 % -11,5 %
consultation hospitalière	776	89,4 %	87,4 % - 91,4 %
ensemble	868	100,0 %*	

*la somme est supérieure à 100 % car les groupes ne sont pas tous disjoints.

Il est à retenir que la très grande majorité (89,4 %) des blessés identifiés seulement dans les BAAC (pour lesquelles le procès-verbal a été exploitable) déclare s’être rendue dans un hôpital ou une clinique.

En outre, parmi ces personnes, il est indiqué que 121 ont été hospitalisées (15,6 %, IC : 9,1% - 22,1%).

Gravité des blessures

Parmi les 616 certificats médicaux (ou similaire : arrêt de travail, acte de décès…), il a été possible de coder la gravité des blessures de 532 personnes (soit 86%). L’impossibilité de coder correspond à des certificats illisibles (« scan » de mauvaise qualité) ou non informatifs quant aux lésions, se limitant souvent à une évaluation du nombre de jours d’ITT (Incapacité Totale Temporaire).

Le tableau suivant compare la gravité des blessés de « BAAC seulement » avec procès-verbal et certificat médical, aux blessés des BAAC qui ont aussi été identifiés dans le Registre.

Tableau 20 : comparaison de la distribution du NISS
entre les blessés « BAAC seulement » et les blessés de l'intersection BAAC - Registre, Rhône, 2001

gravité	« BAAC seulement »		intersection BAAC - Registre	
	effectif	proportion	effectif	proportion
NISS 0	4	0,8 %	0	0 %
NISS 1-3	367	69,0 %	1763	62,9 %
NISS 4-8	115	21,6 %	578	20,6 %
NISS 9-15	38	7,1 %	255	9,1 %
NISS 16-24	7	1,3 %	115	4,1 %
NISS 25-75	1	0,2 %	91	3,2 %
total	532	100,0 %	2802	100,0 %

Test du Chi-deux de comparaison des deux répartitions : p<0,0001

Globalement, les blessés de « BAAC seulement » (avec un procès-verbal et un certificat médical) sont un peu moins gravement blessés que les blessés des BAAC qui ont aussi été identifiés dans le Registre. La proportion de blessés NISS 4+ est de 30,2% (IC : 26,3%- 34,1%) chez les premiers contre 37,1% (IC : 35,3% - 38,9%) chez les seconds. La proportion de blessés NISS 9+ est de 8,6% (IC : 6,2% - 11,0%) chez les premiers contre 16,4% (IC : 15,0% - 17,8%) chez les seconds.

Parmi les blessés de « BAAC seulement » (ceux avec procès-verbal et certificat médical), nous comparons la gravité des sous-groupes définis selon leur recours aux soins. Par définition les non-blessés sont NISS 0, mais cela n'a pas pu être vérifié car il n'y avait pas de certificat médical dans le procès-verbal pour la majorité d'entre eux.

La gravité NISS des blessés s'étant soignés eux-mêmes ou ayant été soignés par un proche n'est pas connue (puisque par définition ils ne disposent pas de certificat médical).

Le tableau suivant permet de décrire la gravité NISS des blessés de « BAAC seulement » (avec procès-verbal et certificat médical) qui ont au plus consulté en médecine de ville ou en service hospitalier.

Tableau 21 : répartition des blessés de l'étude « BAAC seulement »
selon le recours aux soins et selon la gravité NISS

consultation	médecine de ville		hôpital / clinique	
gravité	effectif	proportion	effectif	proportion
NISS 0	2	6,9 %	2	0,4 %
NISS 1-3	27	93,1 %	340	67,6 %
NISS 4-8	0	0,0 %	115	22,9 %
NISS 9-15	0	0,0 %	38	7,6 %
NISS 16-24	0	0,0 %	7	1,4 %
NISS 25-75	0	0,0 %	1	0,2 %
total	29	100,0 %	503	100,0 %

Test du Chi-deux de comparaison des deux répartitions : p<0,0001

La totalité des 29 blessés de « BAAC seulement » (avec procès-verbal et certificat médical) qui ont au plus consulté en médecine de ville sont des blessés de gravité mineure (aucune blessure ne dépassant un AIS 1).

Les 503 blessés de « BAAC seulement » (avec procès-verbal et certificat médical) qui sont allés à l'hôpital (ou à la clinique) sont essentiellement des blessés de gravité mineure ou modérée (90,6 %, IC : 88,0% - 93,2%). Une proportion non négligeable (9,4 %, IC : 6,8% - 12,0%) est de gravité sérieuse et au-delà (NISS 9+).

Distance du lieu d’accident à l’hôpital

Nous comparons la distance du lieu d’accident à l’hôpital le plus proche entre les blessés identifiés seulement dans les BAAC et les blessés des BAAC identifiés aussi dans le Registre.

Tableau 22 : comparaison de la distance du lieu d’accident à l’hôpital le plus proche, entre les blessés « BAAC seulement » et les blessés de l’intersection BAAC - Registre, Rhône 2001

lieu d’accident	BAAC seulement		intersection BAAC - Registre	
	effectif	proportion	effectif	proportion
proche d’un hôpital	1187	90,0 %	2680	95,2 %
loin d’un hôpital	132	10,0 %	136	4,8 %
total	1319	100,0 %	2816	100,0 %

Test du Chi-deux de comparaison des deux répartitions : p<0,001

Les blessés identifiés dans les BAAC seulement (avec procès-verbal exploitable) ont plus souvent été accidentés loin d’un hôpital que les blessés BAAC identifiés aussi dans le Registre.

Parmi les blessés identifiés dans les BAAC seulement nous comparons la distance à l’hôpital le plus proche entre ceux qui sont allés à l’hôpital et ceux qui n’y sont pas allés.

Tableau 23 : répartition des blessés « BAAC seulement » selon la consultation hospitalière, et la distance du lieu d’accident à l’hôpital le plus proche

lieu d’accident	consultation hospitalière :			
	oui		non	
proche d’un hôpital	effectif	proportion	effectif	proportion
loin d’un hôpital	698	90,0 %	84	91,3 %
lieu d’accident	78	10,0 %	8	8,7 %
	776	100,0 %	92	100,0 %

Test du Chi-deux de comparaison des deux répartitions : p<0,68

Parmi les blessés identifiés seulement dans les BAAC, il n’apparaît pas de différence de distance à l’hôpital entre les blessés qui sont effectivement allés consulter à l’hôpital et ceux qui n’y sont pas allés.

Taux d’erreur sur les variables servant au chaînage

Par comparaison des procès-verbaux aux BAAC, les erreurs suivantes ont été relevées :
- erreur sur la variable année de naissance : 9 cas soit 1,4 %
- erreur sur la variable mois de naissance : 5 cas soit 0,6 %

Dans trois cas l’erreur portait à la fois sur le mois et sur l’année de naissance.
- erreur sur la variable sexe : 8 cas soit 0,9 %.

Il n’a pas été possible d’identifier des erreurs sur les autres variables utilisées dans le chaînage entre les BAAC et le Registre :
- erreurs sur les variables mois, année et lieu de l’accident : erreurs non identifiables car variables servant à identifier les procès-verbaux correspondant aux BAAC,
- erreurs sur la variable type d’usager de la victime : erreur non identifiable car il n’avait pas été prévu de corriger les valeurs de cette variable lors de la saisie de données supplémentaires provenant des PV.

Chaînage complémentaire sur les noms et prénoms

Grâce à l’information supplémentaire recueillie dans les procès-verbaux, notamment les nom et prénom des blessés, nous avons lancé une nouvelle procédure de chaînage entre les blessés non encore rapprochés.

Parmi les 868 blessés identifiés seulement dans les BAAC et avec procès-verbal exploitable, dont 865 pour lesquels nom et prénom ont été saisis, 94 blessés (soit 10,8 %, IC : 8,7% - 10,9%) ont pu être

identifiés comme communs au Registre. Le détail de ce rapprochement est donné dans le tableau suivant.

Tableau 24: résultat détaillé du chaînage supplémentaire avec noms et prénoms, Rhône, 2001
(nombre de paires reliées parmi nombre de paires sélectionnées)

Stratégie de sélection de paires	résultat
1) même prénom ET même nom	79 parmi 116
2) même prénom ET même année de naissance ET date accident +/- 3 jours	3 parmi 17
3) même prénom ET même mois de naissance ET date accident +/- 3 jours	2 parmi 40
4) même nom ET même mois de naissance ET date accident +/- 3 jours	8 parmi 15
5) même nom ET même année de naissance ET date accident +/- 3 jours	1 parmi 5
6) inversion nom et prénom	2 parmi 6
7) comme 1 mais nom de jeune fille des PV comparé au nom d’épouse du Registre	infructueux
8) comme 2 à 5 mais au lieu de date d’accident +/-3 jours : différence de 29-31 jours (erreur de 1 mois)	infructueux
9) comme 2 à 5 mais au lieu de date d’accident +/-3 jours : jour et mois accident inversés	infructueux
10) même nom ET même mois de naissance ET même année de naissance	infructueux
10 bis) même prénom ET même mois de naissance ET même année de naissance	infructueux
11) date de l'accident +/- 3 jours, même mois ET année naissance	infructueux
ensemble	94 paires reliées

Discussion

Il était au départ prévu d’effectuer cette étude à travers une enquête par questionnaire auprès des blessés eux-mêmes. Cela a été remis en cause lorsque nous avons constaté, une fois l’extraction et la saisie d’informations des procès-verbaux réalisée, que ceux-ci contenaient la plupart du temps les informations qui nous intéressaient : hôpital ou médecin de ville consulté, certificats médicaux avec description des blessures. Cette description des blessures s’est révélée très informative, sans doute bien plus fiable que celle que nous aurions pu obtenir des blessés eux-mêmes, par questionnaire, surtout cinq ans après l’accident. En revanche, la procédure d’extraction et de saisie d’informations des procès-verbaux n’était pas conçue pour saisir notamment le jour de naissance des blessés. Nous aurions ainsi eu la date de naissance complète, et nous aurions pu l’utiliser dans le chaînage supplémentaire.

Le procès-verbal a pu être exploité pour 66 % des blessés identifiés dans les BAAC seulement. La non-disponibilité d’un certain nombre de procès-verbaux semble due à des problèmes techniques. Il ne semble pas y avoir de différence en termes de distance entre le lieu de l’accident et l’hôpital le plus proche entre les blessés avec procès-verbal exploitable et ceux sans. En termes de gravité, la proportion observée de blessés graves est de 7,4% parmi les blessés avec PV exploitable versus 8,9% parmi les blessés sans PV exploitable. Le test statistique ne met pas en évidence de différence, mais la puissance de ce test est très faible. En outre il s’agit de la gravité mesurée par les forces de l’ordre, faute de mieux. Une différence est mise en évidence en termes de répartition des types de force de l’ordre (et donc en termes de type de réseau et de répartition urbain/rural). Il n’est donc pas assuré que les résultats obtenus sur le sous-groupe des blessés avec procès-verbaux exploitables soient généralisables à l’ensemble des blessés identifiés dans les BAAC seulement.

Les blessés identifiés dans les BAAC seulement, avec procès-verbal exploitable, ont les caractéristiques suivantes : la très grande majorité d’entre eux (89,4 %) seraient allés consulter en hôpital ou en clinique. La consultation hospitalière est en général confirmée par un certificat médical issu d’un service hospitalier : au total 64 % des blessés identifiés seulement dans les BAAC (avec procès-verbal exploitable) présentent un certificat médical issu d’un service hospitalier.

La gravité des blessures a pu être évaluée grâce aux certificats médicaux inclus dans les procès-verbaux. Les blessés identifiés seulement dans les BAAC sont un peu moins gravement blessés que les

blessés des BAAC identifiés aussi dans le Registre. Cela indique un biais de sélection du Registre sur la gravité lésionnelle.
Parmi les blessés identifiés seulement dans les BAAC, ceux ayant au plus consulté un médecin de ville sont tous de gravité mineure (NISS 1-3), ce qui est conforme à l'attendu. Parmi ceux ayant consulté dans une structure hospitalière, ils sont en majorité de gravité mineure ou modérée, ce qui signifie aussi qu'un certain nombre (9 %) sont de gravité sérieuse ou au-delà (NISS 9 +).
Ce résultat mais surtout la forte proportion de blessés avec un certificat médical issu d'un hôpital (64%) invalide l'hypothèse de départ, à savoir que les blessés ne seraient identifiés que dans les BAAC parce que de gravité mineure ET sans consultation hospitalière. Les blessés de « BAAC seulement » ont majoritairement consultés une structure hospitalière.

Nous avons identifié un autre facteur de biais de sélection du Registre : la distance à l'hôpital. Les blessés dont l'accident s'est produit loin d'un hôpital vont un peu moins consulter un hôpital que les blessés dont l'accident s'est produit à proximité d'un hôpital. C'est un résultat tout à fait plausible pour les blessés de gravité légère (NISS 1-3), voire modérée (NISS 4-8), qui constituent la majorité des blessés. C'est un résultat d'analyse univariée, et il ne peut pas s'expliquer par une gravité moindre des blessés accidentés loin d'un hôpital, car les accidents en zone rurale sont en moyenne plus graves que les accidents en zone urbaine. Par ailleurs, il se pourrait qu'il y ait un léger biais de sélection du Registre sur le type d'usager : une étude (Sigal 2005) sur un hôpital du Registre a montré que les accidents de vélo étaient un peu sous-notifiés. Cela ne pouvait pas être exploré dans cette étude puisque les données des forces de l'ordre souffrent fortement de ce biais de sélection.

Concernant l'estimation des taux d'erreur, elle n'a été possible que pour les variables mois, année de naissance et sexe, et d'une façon très limitée car ces variables servent justement à l'identification des personnes au sein d'un PV. En outre, cela ne prend en compte que les erreurs de saisie des procès-verbaux vers les BAAC et pas les erreurs possibles lors de la rédaction des procès-verbaux. L'identification d'erreurs sur les autres variables ayant servi au chaînage de routine n'a pas été possible (date de l'accident et lieu de l'accident), car elles ont elles-mêmes servi à l'identification des procès-verbaux des blessés de l'étude.
L'estimation des taux d'erreur devait servir à estimer le nombre de faux négatifs et de faux positifs du chaînage. Une autre méthode existe et est présentée dans le chapitre 3, section « estimation du nombre de faux négatifs et de faux positifs ».

Le chaînage supplémentaire, utilisant les noms et prénoms comme identifiants supplémentaires a permis de rapprocher 10,8 % des blessés précédemment identifiés comme « seulement dans les BAAC ». L'encodage utilisé (Soundex) est optimisé pour les noms anglo-saxons ; des méthodes existent pour encoder les noms français (méthode Henry par exemple) (Armstrong 2000; Quantin, Binquet et al. 2004) ; elles ne sont cependant pas « prêtes-à-l'emploi » et il reste à prendre en compte la translittération des noms provenant d'autres alphabets (notamment arabes). Utiliser des encodages plus adaptés serait évidemment plus efficace pour chaîner les BAAC et le Registre, mais le taux de reliés ne serait sans doute pas changé fondamentalement.
Cela signifie que la majorité des blessés identifiés seulement dans les BAAC restent dans cette situation : « identifiés seulement dans les BAAC », alors que la majorité d'entre eux ont effectivement été blessés lors de l'accident et sont allés consulter dans une structure hospitalière, et qui plus est, couverte par le Registre.

Cela indique un dysfonctionnement du Registre : celui de fiches de notification non établies pour des personnes se présentant dans une structure hospitalière, blessées lors d'un accident de la route. On imagine que des fiches n'ont pas été établies par simple omission, ou par urgence / surcharge de travail, notamment dans les services d'urgence, ou encore par non-identification des circonstances de la blessure telles qu'un accident de la circulation routière.

Afin d'explorer les situations de sous-notification de certains services hospitaliers, nous avons établi deux distributions. Parmi les blessés identifiés seulement dans les BAAC (et qui restent comme tels après le chaînage supplémentaire sur nom et prénom) et avec un certificat médical émis par un hôpital,

nous avons établi la distribution des services hospitaliers émetteurs de ces certificats. En parallèle, parmi les blessés des BAAC qui sont aussi identifiés dans le Registre (autrement dit qui sont passées par le « filtre » Registre) nous avons aussi établi la distribution des services hospitaliers émetteurs des fiches de notification de ces blessés. En comparant les deux distributions, nous pouvons identifier quels services sont sur-représentés parmi les blessés identifiés dans les BAAC seulement, ce qui indique une sous-notification de leur part. Ce résultat a été transmis à la directrice du Registre : il a confirmé la connaissance empirique que l’ARVAC détient quant au dysfonctionnement, certaines années, de quelques services hospitaliers.

Résumé

Contexte :
Il s’agit de mieux comprendre pourquoi un certain nombre de blessés ne sont enregistrés que dans les données des forces de l’ordre (et pas dans le Registre). Est-ce que leurs blessures étaient légères et ne nécessitaient pas de consulter un service hospitalier ? ou bien s’agit-il d’un dysfonctionnement du Registre ? ou bien d’un dysfonctionnement du chaînage entre les données des forces de l’ordre et le Registre ?

Méthode :
La population d’étude est l’ensemble des blessés identifiés dans les données des forces de l’ordre seulement, restreint aux accidents de l’année 2001 (fichiers BAAC). Son effectif est de 1319 blessés, classés blessés « légers » ou « graves ». Nous avons consulté les procès-verbaux desquels les BAAC sont issus, et nous en avons extrait des informations supplémentaires : nom et prénom, type de recours aux soins (hôpital ou médecin de ville), nom du service ou du médecin, et certificat médical. Celui-ci nous permet d’évaluer la gravité lésionnelle (le NISS est utilisé). Les noms et prénoms nous permettent d’effectuer un chaînage supplémentaire entre ces blessés et ceux identifiés dans le Registre seulement.

Résultats :
Le procès-verbal est accessible et exploitable pour 868 blessés (66 %). Les deux groupes de blessés, avec ou sans procès-verbal exploitable) ne semblent pas différer en termes de gravité ni en termes de distance entre le lieu d’accident et l’hôpital le plus proche. Cependant il s’agit d’une part de la gravité estimée par les forces de l’ordre, et d’autre part, le test a une puissance très faible. Des différences entre les deux groupes de blessés (avec ou sans PV exploitable) sont mises en évidence en termes de mois de l’accident et de type de force de l’ordre émettant le procès-verbal.
Parmi les 868 blessés avec PV exploitable, 4,1 % ne seraient pas blessés (IC à 95% : 2,8% - 5,5%) ; parmi les blessés, 5,4 % (IC : 3,9% -6,9%) auraient au plus consulté en médecine de ville et 87,1 % (IC : 84,9% - 89,3%) auraient consulté un service hospitalier. Tous les services cités dans les procès-verbaux font partie du Registre. La consultation hospitalière est certaine pour 64 % (IC : 61,0%-67,4%) des « enquêtés » car attestée par un certificat médical. La gravité lésionnelle est un peu moindre que celle des blessés des BAAC qui sont également identifiés dans le Registre. On note cependant 9 % de blessés NISS 9+ (IC : 6,8% -12,0%) et parmi ceux s’étant rendus à l’hôpital, 16 % d’hospitalisés (IC : 15,0% -17,8%). Les blessés identifiés dans les BAAC seulement ont plus souvent eu un accident loin d’un hôpital que les blessés identifiés comme communs aux BAAC et au Registre (10 % vs 5 %). La gravité lésionnelle et la distance à l’hôpital sont ainsi identifiées comme des biais de sélection du Registre. Le chaînage supplémentaire utilisant les noms et prénoms permet d’identifier un enregistrement correspondant dans le Registre pour 11 % (IC : 8,7%- 12,9%) des 868 blessés avec PV exploitable.

Conclusion :
Notre hypothèse était que la grande majorité des blessés identifiés seulement dans les BAAC étaient de gravité lésionnelle mineure telle qu’ils n’avaient pas consulté de service hospitalier. Cette hypothèse est invalidée. Même s’ils sont majoritairement de gravité mineure ou modérée, la grande majorité de ces blessés ont consulté un service hospitalier. Les résultats indiquent que certains services hospitaliers ont, cette année-là, dysfonctionné dans la notification des accidentés vers le Registre, mais aussi que le chaînage de « routine » entre les BAAC et le Registre n’est pas optimal.

Chapitre 2
Erreurs de classement de la gravité lésionnelle dans les données des forces de l'ordre

Introduction

Ce chapitre reprend l'article publié dans Journal of trauma (cf. annexe) ; il apporte quelques points supplémentaires.

Dans la grande majorité des pays, les données nationales sur les victimes de la circulation routière proviennent des forces de l'ordre. Les définitions policières de la gravité lésionnelle sont souvent une catégorisation en tués, blessés graves et blessés légers. Les définitions policières du niveau de gravité : léger ou grave des blessés non-décédés sont en général basées sur l'hospitalisation (oui/non), la durée d'hospitalisation, le type de lésion, la nécessité d'un traitement médical ou sa durée, l'incapacité à reprendre le travail, la durée d'incapacité, ou encore sur une combinaison de certains de ces critères. Comme les données des forces de l'ordre constituent la plupart du temps les seules données sur les victimes de la circulation routière disponibles à l'échelle nationale, elles sont à la base des connaissances épidémiologiques actuelles. Leur qualité doit donc être évaluée.

Certaines études ont analysé la validité de la classification policière de la gravité (Bull and Roberts 1973; Shinar, Treat et al. 1983; Barancik and Fife 1985; Hansen, Popkin et al. 1991; Rosman and Knuiman 1994; Austin 1995; Aptel, Salmi et al. 1999). Elles ont comparé le résultat de la classification policière avec la même classification (ou une très similaire) mais appliquée par des médecins. D'autres études (Dove, Pearson et al. 1986; Tunbridge and Everest 1988; Nordentoft, Larsen et al. 1989; Agran, Castillo et al. 1990; Farmer 2003; Morris, Mackay et al. 2003) ont étudié la discordance entre la classification policière de la gravité et une classification traumatologique internationale et reconnue, en l'occurrence l'Abbreviated Injury Scale. Dans ces études, la gravité lésionnelle des blessés, parfois poly-traumatisés, a été mesurée par le MAIS (Maximum AIS) ou par l'ISS (Injury Severity Score). Ces études se sont penchées sur les erreurs de classement de la gravité des forces de l'ordre.
En outre, quelques chercheurs (Hansen, Popkin et al. 1991; Lopez, Rosman et al. 2000; Farmer 2003) ont étudié si la discordance entre classification policière et classification traumatologique était associée à des caractéristiques de la victime et/ou de l'accident, telles que le type d'usager, le sexe, ou la gravité lésionnelle elle-même. Autrement dit, ces études se sont penchées sur les erreurs différentielles et les caractéristiques associées à ces erreurs.

Un tableau est présenté en annexe, listant l'ensemble de ces études et donnant leurs principales caractéristiques (zone géographique de l'étude, durée de l'étude, nombre d'observations, type d'analyse, et résultats : variables mises en évidence comme étant liées aux erreurs de classement).

En France, jusqu'en 2004 inclus, la définition policière du niveau de gravité est basée sur la durée d'hospitalisation avec un seuil à six jours : sont classés « blessés graves » ceux qui nécessitent plus de six jours d'hospitalisation, et « blessés légers » ceux qui nécessitent de un à six jours (inclus) d'hospitalisation ou des soins sans hospitalisation. La validité de la classification policière française a déjà été analysée. Elle a été comparée avec la durée réelle d'hospitalisation dans deux études. La première étude a été conduite sur les blessés d'accident de la circulation routière à la Réunion (Aptel, Salmi et al. 1999) et a estimé la sensibilité à 77 % et la spécificité à 61 %, soit : 77 % des blessés réellement graves et 61 % des blessés réellement légers sont classés comme tels par les forces de l'ordre. La deuxième étude portait sur les blessés dans le département du Rhône entre 1998 à 2000, recensés par le Registre (Ovième 2003) ; la sensibilité a été estimée à 68 %, la spécificité à 95 %, la valeur prédictive positive à 64 %, et la valeur négative prédictive à 96 %. Ces deux dernières valeurs

signifient respectivement que, parmi les blessés classés « graves » par les forces de l'ordre, 64 % le sont réellement et parmi ceux classés « légers », 96 % le sont réellement. Certaines de ces valeurs sont loin de 100 %. Cela confirme une information non-écrite à savoir que les forces de l'ordre ne se renseignent pas toujours auprès des hôpitaux pour connaître la durée d'hospitalisation d'un blessé accidenté six jours plus tôt. On peut penser qu'ils estiment la gravité lésionnelle de la personne « au jugé » sur le lieu de l'accident, ou bien encore qu'ils s'appuient sur le nombre de jours d'ITT (incapacité totale temporaire) évalué par des médecins pour des raisons de procédure légale, et rapporté dans les procès-verbaux.

Par ailleurs, la durée d'hospitalisation n'est pas une très bonne mesure de la gravité lésionnelle. Elle est en effet liée à des caractéristiques autres que la gravité lésionnelle elle-même : des caractéristiques individuelles, telles que l'âge, la co-morbidité (Bergeron, Lavoie et al. 2005), mais aussi des caractéristiques générales comme les politiques hospitalières.
Par conséquent, nous ne nous intéressons pas plus longtemps à la relation entre classification policière et durée réelle d'hospitalisation. Nous nous focalisons sur la relation entre la classification policière de la gravité et la gravité lésionnelle elle-même ; la classification policière, c'est ce qui est disponible sur les données nationales ; la gravité lésionnelle, c'est ce qui nous importe. Nous avons besoin de savoir comment elles sont associées. Nous mesurons la gravité lésionnelle avec le NISS. Nous rappelons qu'il est égal à la somme des carrés des niveaux de gravité AIS des trois lésions les plus graves, indépendamment de la région corporelle où elles se situent. Nous groupons ensuite les scores NISS en deux catégories les plus proches possibles de la définition de la classification policière, afin de pouvoir étudier la discordance entre les deux classifications ; par simplification nous utilisons le terme d'erreurs de classement. La première partie de l'analyse évalue l'ampleur des erreurs de classement de la gravité par les forces de l'ordre. La deuxième partie porte sur la question des erreurs différentielles. Nous identifions et nous quantifions les associations entre les erreurs de classement et les caractéristiques du blessé et de l'accident. Plus précisément, il s'agit d'identifier et de quantifier à quel point certains groupes de blessés, et/ou d'accidents sont plus susceptibles d'avoir leur gravité lésionnelle sous-classée ou sur-classée par les forces de l'ordre.

Matériel et Méthode

L'étude se base sur les blessés par accident dans le Rhône, identifiés comme communs aux BAAC et au Registre, pour lesquels les deux classifications policière et médicale (NISS) sont disponibles. L'étude porte sur la même période que l'analyse des biais de sélection des données des forces de l'ordre, soit 1997-2001. Les données des forces de l'ordre (BAAC), le Registre et le chaînage entre les BAAC et le Registre ont été décrits au début de ce mémoire (paragraphe Matériel).

Pour les informations démographiques (âge, sexe) les variables du Registre sont utilisées, et pour les informations accidentologiques (type d'usager, type de réseau,…) les variables des BAAC sont utilisées. Pour les quelques observations où les variables âge et sexe, variables du Registre sont non renseignées (moins de 0,5 %), les informations fournies par les variables des BAAC sont alors utilisées. Nous excluons les quelques observations (0,7 %) où le NISS est manquant (ce qui correspond à une description lésionnelle insuffisante pour évaluer la gravité).

La première partie de l'analyse porte sur l'ampleur et la direction de la discordance entre la classification policière de la gravité et la classification NISS. Nous présentons d'abord pour chaque catégorie de NISS, la proportion de blessés qui ont été classés « blessés graves » par les forces de l'ordre. Nous utilisons les cinq catégories définies par Rosman (Rosman, Knuiman et al. 1996) : NISS 1-3, 4-8, 9-15, 16-24 et 25-75. Les valeurs seuils correspondent au changement du niveau de gravité pour une blessure isolée, élevé au carré (AIS 1^2, 2^2, 3^2, 4^2, 5^2). Ensuite, en regroupant les scores NISS en deux catégories, nous pouvons étudier la discordance entre les deux classifications et donc estimer la sensibilité, la spécificité, les valeurs prédictives négative et positive de la classification policière (en

considérant le NISS comme la classification de référence). Nous choisissons la valeur seuil NISS 16 pour former ces deux catégories, ce qui correspond au niveau de gravité AIS 4 d'une lésion isolée. Ce niveau de gravité nous semble correspondre le plus au seuil de six jours d'hospitalisation de la définition policière. En effet, d'après l'étude descriptive des lésions des victimes du Registre (Chiron, Gadegbeku et al. 2004), la plupart des blessures de gravité AIS 3 sont des fractures ouvertes déplacées, et la plupart des blessures AIS 4 sont des blessures à la tête ou au thorax, telles que des contusions pulmonaires bilatérales. Les secondes nécessitent plus de six jours d'hospitalisation, contrairement aux premières. Ceci est corroboré par une autre étude (Haddak, Filou et al. 2004) des données du Registre qui donne une médiane de quatre jours d'hospitalisation pour les victimes ISS[2] 9-15 et 13 jours pour les victimes ISS 16-24. Tout cela justifie le choix du seuil à 16 du NISS. Les catégories NISS 1-15 et NISS 16-75 sont ainsi les plus proches possible en termes de définition, des catégories « blessés légers » et « blessés graves » de la classification policière. Cela nous permet de passer du terme de discordance à celui d'erreur de classement, et donc d'erreurs de sous-classement et de sur-classement. Le sur-classement correspond aux blessés NISS 1-15 classés comme « blessés graves » par les forces de l'ordre ; le sous-classement correspond aux blessés NISS 16-75 classés « blessés légers » par les forces de l'ordre.

La seconde partie de l'analyse consiste à identifier les caractéristiques du blessé et/ou de l'accident qui sont associées à ces erreurs de classement, de sur-classement ou de sous-classement, et à quantifier ces associations. Pour cela nous modélisons respectivement :
1) la probabilité d'erreur de classement,
2) la probabilité parmi les blessés NISS 16-75 d'être classés « blessés légers » par les forces de l'ordre,
3) la probabilité parmi les blessés NISS 1-15 d'être classés « blessés graves ».
Nous modélisons chacune de ces probabilités en fonction des caractéristiques du blessé, de l'accident, et de caractéristiques « externes ». Les caractéristiques que nous considérons sont :
1) les caractéristiques du blessé : la gravité lésionnelle (NISS en cinq catégories), l'âge (cinq catégories), le sexe et le type d'usager ;
2) les caractéristiques de l'accident : la présence/absence de tiers, le type de réseau, le caractère urbain/rural, le type de jour (début / fin de semaine), la luminosité (jour / nuit), les conditions météorologiques (bonnes / mauvaises), et la configuration de l'accident en intersection (oui / non) ;
3) les caractéristiques externes : année calendaire et force de l'ordre.
Nous construisons les modèles multivariés en incluant les variables significatives dans les analyses univariées (au niveau d'erreur 5 %) ou les variables mises en évidence dans d'autres études. Nous retenons dans les modèles multivariés les variables significativement associées avec la variable à expliquer au niveau d'erreur de 5 %.

Afin de mesurer l'ampleur de l'association entre erreur de classement et une caractéristique, nous utilisons les risques relatifs et non les odds ratios, car les premiers sont plus simples à interpréter. On ne peut pas utiliser ici l'approximation des risques relatifs par les odds ratios (facilement disponibles par une régression logistique) car l'événement étudié (erreur de classement, sous ou sur-classement) n'est pas un événement rare. Dans cette situation, différentes méthodes de modélisation (Greenland 2004; Spiegelman and Hertzmark 2005) permettent d'estimer des risques relatifs : un modèle log-binomial, un modèle logistique avec des transformations ou encore un modèle de Poisson modifié (Zou 2004).

Nous choisissons (cf. chap. 1) la régression de Poisson avec une estimation robuste de la variance, et nous utilisons la procédure GENMOD de SAS (Lindquist 2004; Spiegelman and Hertzmark 2005).
En plus des risques relatifs, nous fournissons aussi les résultats sous forme de probabilités ajustées, qui correspondent à des probabilités marginales pour une caractéristique donnée si les données étaient équilibrées. Ces probabilités sont calculées en standardisant les données sur les autres facteurs de

[2] Nous rappelons que l'ISS (Injury Severity Score) est la somme des carrés des scores AIS des trois blessures les plus graves de régions corporelles distinctes.

biais, précisément en moyennant sur les modalités des autres variables catégorielles en leur attribuant des poids égaux. Ceci est obtenu avec l'option LSMEANS de la procédure GENMOD.

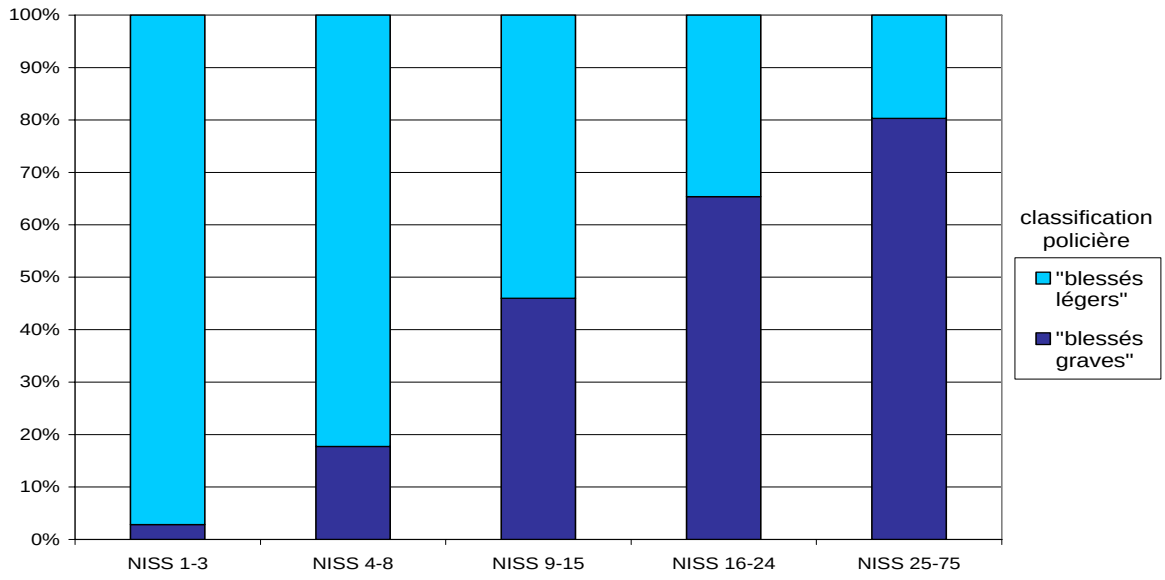
Résultats

Dans le département du Rhône, sur la période 1997-2001, les données des forces de l'ordre contiennent 23 080 victimes, le Registre 54 661. Le chaînage a permis d'identifier 14 974 victimes communes aux deux. Parmi ces victimes reliées, 14 446 sont classées comme blessées (« blessé léger » ou « blessé grave ») par les forces de l'ordre. Nous excluons alors 104 blessés (0,7 %) à cause de score NISS manquant. L'étude porte donc sur 14 432 blessés.

Erreurs de classement de la gravité

La Figure 5 montre la probabilité d'être classé « blessé grave » en fonction du NISS. Cette probabilité est de plus en plus grande en fonction du niveau du NISS (cinq catégories).

Figure 5 : proportions de blessés classés « légers » et « graves » par les forces de l'ordre, en fonction du NISS (14 342 blessés non-décédés, intersection BAAC - Registre, Rhône, 1997-2001)



Le Tableau 25 donne en termes d'effectifs le croisement des deux classifications.

Tableau 25 : distribution des blessés selon les classements policier et médical de la gravité, Rhône, 1997-2001

Classement des forces de l'ordre	Classement médical NISS 1-15	NISS 16-75	total
"blessé léger"	11 815	306	12 121
"blessé grave"	1 444	777	2 221
total	13 259	1 083	14 342

La mesure de concordance kappa est estimée à 0,41. La sensibilité est estimée à 71,7 % (IC à 95% : 69,0% - 74,4%). Autrement dit, un peu plus des deux tiers des blessés NISS 16-75 sont classés « blessés graves » par les forces de l'ordre. Inversement, la valeur prédictive positive est estimée à 35,0 % (IC : 33,0% - 37,0%) : parmi ceux classés « blessés graves » par les forces de

l'ordre, un tiers sont effectivement NISS 16-75. La spécificité est estimée à 89,1 % (IC : 88,6% - 89,6%), c'est-à-dire que la plupart des blessés NISS 1-15 sont classés « blessés légers » par les forces de l'ordre. Inversement, la valeur prédictive négative est estimée à 97,5 % (IC : 97,2% - 97,8%) : parmi ceux classés « blessés légers » par les forces de l'ordre, la quasi totalité sont effectivement des blessés NISS 1-15.

Caractéristiques associées aux erreurs de classement de la gravité

Dans les analyses univariées, la quasi-totalité des variables considérées sont significativement associées aux erreurs de classement de la gravité par les forces de l'ordre. Les exceptions sont la présence / absence de tiers et le type de jour (début / fin de semaine).
Dans la construction du modèle multivarié, les caractéristiques suivantes ne sont plus significatives : le sexe, le type de réseau, l'environnement (en / hors agglomération), les conditions lumineuses (jour / nuit), les conditions météorologiques et la configuration de l'accident (intersection oui/non). Les caractéristiques significativement associées aux erreurs de classement de la gravité par les forces de l'ordre, dans le modèle multivarié, sont la gravité lésionnelle (mesurée par le NISS en cinq catégories), l'âge du blessé, le type d'usager, le type de force de l'ordre enregistrant l'accident et l'année. Les risques relatifs ajustés et les probabilités ajustées sont donnés dans le Tableau 26.

Tableau 26 : caractéristiques liées aux erreurs de classement de la gravité par les forces de l'ordre, analyse multivariée sur 14 342 blessés (intersection BAAC - Registre, Rhône, 1997-2001)

	effectif	probabilité ajustée	RR ajusté	IC à 95 %	
gravité lésionnelle (p<0,0001)					
NISS 1-3	8 895	0,03	1,00		
NISS 4-8	2 883	0,16	5,28	4,55	6,12
NISS 9-15	1 481	0,36	12,23	10,57	14,15
NISS 16-24	621	0,25	8,64	7,19	10,39
NISS 25-75	462	0,15	4,97	3,90	6,34
âge (p=0,0003)					
0-13 ans	991	0,13	1,00		
14-17 ans	1 038	0,12	0,94	0,75	1,18
18-24 ans	3 493	0,14	1,11	0,91	1,35
25-64 ans	7 809	0,16	1,21	1,01	1,45
65 ans et plus	1 011	0,17	1,33	1,09	1,63
type d'usager (p<0,0001)					
piétons	2 038	0,18	1,36	1,20	1,53
cyclistes	485	0,13	1,02	0,82	1,25
usagers de 2RM	2 561	0,17	1,26	1,14	1,39
automobilistes	8 701	0,13	1,00		
autres usagers	557	0,11	0,85	0,64	1,13
force de l'ordre (p<0,0001)					
CRS	2 479	0,10	1,00		
gendarmerie	3 022	0,24	2,34	1,98	2,76
police	8 841	0,13	1,25	1,06	1,48
année (p=0,003)					
1997	2 612	0,17	1,28	1,12	1,45
1998	3 152	0,14	1,09	0,97	1,24
1999	3 016	0,14	1,05	0,92	1,20
2000	2 763	0,14	1,06	0,93	1,21
2001	2 799	0,13	1,00		

RR ajustés par une régression de Poisson avec estimation robuste de la variance

La relation entre la probabilité d'erreur de classement et le NISS (en cinq catégories) est en forme de U renversé, avec un risque d'erreur de classement qui est 12,2 fois plus élevé pour les blessés NISS 9-15 par rapport aux blessés NISS 1-3. La probabilité d'erreur de classement augmente légèrement en

fonction de l'âge des blessés, avec un risque relatif de 1,3 pour les blessés âgés de 65 ans et plus par rapport aux blessés âgés de 0 à 13 ans. Les blessés piétons et usagers de 2RM ont une plus grande probabilité d'être mal classés avec un risque relatif de 1,4 et 1,3 respectivement, par rapport aux blessés automobilistes. Par rapport au corps des CRS, le corps de police est un peu plus susceptible de faire des erreurs de classement de la gravité (RR=1,2), mais le corps de gendarmerie encore plus (RR=2,3). Enfin, la probabilité d'erreur de classement est plus forte en 1997 qu'en 2001 (RR=1,3).

Caractéristiques associées aux erreurs de sur-classement de la gravité

Nous rappelons que le sur-classement de la gravité correspond à la situation où les blessés de gravité NISS 1-15, sont classés blessés « graves » par les forces de l'ordre.
Dans les analyses univariées, toutes les variables explorées sont significativement associées avec la probabilité d'erreur de sur-classement de la gravité par les forces de l'ordre. Dans l'analyse multivariée, la probabilité de sur-classement de la gravité est significativement associée avec la gravité lésionnelle (mesurée par le NISS), l'âge du blessé, le type d'usager, le type de force de l'ordre et l'année (résultats donnés dans le Tableau 27).

Tableau 27 : caractéristiques liées au sur-classement de la gravité par les forces de l'ordre, analyse multivariée sur 13 259 blessés NISS 1-15 classés « blessés graves » (intersection BAAC - Registre, Rhône, 1997-2001)

	effectif	probabilité ajustée	RR ajusté	IC à 95 %	
gravité lésionnelle (p<0,0001)					
NISS 1-3	8 895	0,03	1,00		
NISS 4-8	2 883	0,15	4,89	4,22	5,68
NISS 9-15	1 481	0,33	10,94	9,43	12,68
âge (p=0,0003)					
0-13 ans	926	0,09	1,00		
14-17	948	0,10	1,06	0,82	1,35
18-24	3 254	0,12	1,25	1,00	1,55
25-64	7 226	0,12	1,33	1,08	1,63
65 ans et plus	905	0,14	1,47	1,17	1,84
type d'usager (p<0,0001)					
piétons	1 816	0,15	1,42	1,25	1,63
cyclistes	436	0,11	1,01	0,81	1,26
usagers de 2RM	2 219	0,13	1,24	1,11	1,38
automobilistes	8 260	0,11	1,00		
autres usagers	528	0,08	0,78	0,57	1,07
force de l'ordre (p<0,0001)					
CRS	2 376	0,07	1,00		
Gendarmerie	2 506	0,23	3,10	2,58	3,73
Police	8 377	0,09	1,16	0,96	1,40
année (p<0,0001)					
1997	2 409	0,14	1,49	1,30	1,71
1998	2 941	0,12	1,22	1,07	1,40
1999	2 782	0,11	1,16	1,00	1,34
2000	2 532	0,11	1,11	0,96	1,29
2001	2 595	0,10	1,00		

RR ajustés par une régression de Poisson avec estimation robuste de la variance

La probabilité de sur-classement de la gravité augmente avec le niveau de NISS : par rapport aux blessés NISS 1-3, les blessés NISS 4-8 ont cinq fois plus de risque d'être sur-classés et les blessés NISS 9-15 onze fois plus. La probabilité de sur-classement est liée au type d'usager : les piétons et usagers de 2RM sont un peu plus susceptibles d'avoir leur gravité sur-classée que les automobilistes. La probabilité de sur-classement de la gravité est associée au type de force de l'ordre qui a enregistré l'accident : la gendarmerie a trois fois plus de risque de sur-classer la gravité que les CRS et la police. Enfin, la probabilité de sur-classement de la gravité a diminué sur la période étudiée 1997-2001. Si

nous supposons une évolution linéaire en remplaçant dans le modèle la variable année catégorielle en une variable quantitative, nous obtenons un risque relatif de 0,91 (IC à 95% : 0,89-0,94) entre deux années consécutives.

Caractéristiques associées aux erreurs de sous-classement de la gravité

Nous rappelons que le sous-classement de la gravité correspond à la situation où les blessés de gravité NISS 16+, sont classés blessés « légers » par les forces de l'ordre.
Dans les analyses univariées, toutes les variables sauf l'âge, le sexe et les conditions météorologiques sont significativement associées avec la probabilité de sous-classement de la gravité. Dans l'analyse multivariée, la probabilité de sous-estimation est associée significativement à la gravité lésionnelle (mesurée par le NISS), le type de réseau, le type de force de l'ordre, et l'année. Nous donnons uniquement les résultats ajustés (Tableau 28).

Tableau 28 : caractéristiques liées au sous-classement de la gravité par les forces de l'ordre, analyse multivariée sur 1083 blessés NISS 16-75 classés « blessés légers » (intersection BAAC - Registre, Rhône, 1997-2001)

	effectif	probabilité ajustée	RR ajusté	IC à 95 %	
gravité lésionnelle (p<0,0001)					
NISS 16-24	621	0,29	1,69	1,39	2,07
NISS 25-75	462	0,17	1,00		
type de réseau (p=0,02)					
autoroutes	78	0,13	1,00		
routes nationales	135	0,24	1,84	1,02	3,32
routes départementales	422	0,28	2,16	1,12	4,14
routes communales	406	0,28	2,18	1,11	4,27
autres	42	0,22	1,69	0,72	3,93
force de l'ordre (p<0,0001)					
CRS	102	0,32	3,06	1,90	4,93
gendarmerie	516	0,10	1,00		
police	465	0,32	3,11	2,31	4,18
année (p=0,001)					
1997	203	0,18	1,00		
1998	211	0,18	1,03	0,72	1,46
1999	234	0,21	1,16	0,84	1,61
2000	231	0,26	1,46	1,06	1,99
2001	204	0,30	1,65	1,21	2,26

RR ajustés par une régression de Poisson avec estimation robuste de la variance

La probabilité de sous-estimation de la gravité par les forces de l'ordre est presque deux fois plus grande pour les blessés NISS 16-24 que pour les blessés NISS 25-75. La probabilité de sous-classement est liée au type de réseau où l'accident s'est produit : les blessés sur routes départementales et voies communales ont deux fois plus de risque d'avoir leur gravité sous-classée par rapport aux blessés sur autoroute. Les CRS et la police ont trois fois plus de risque de sous-classer la gravité que la gendarmerie. Enfin, la probabilité de sous-estimation de la gravité a augmenté sur la période 1997-2001. Si nous supposons une évolution linéaire, nous obtenons un risque relatif de 1,1 (IC à 95% : 1,08-1,23) entre deux années consécutives. Ces résultats sont ajustés sur les autres variables du modèle, et donc en particulier sur le NISS. La catégorie NISS 25-75 étant très large, nous avons comparé les médianes du NISS dans les sous-groupes constitués par les trois forces de l'ordre, ou les cinq types de réseau ou les cinq années ; elles sont très similaires (entre 33,5 et 35).

Discussion

Interprétation des résultats

Le premier résultat de cette étude est l'ampleur des erreurs de classement de la gravité par les forces de l'ordre (la sensibilité est acceptable mais la valeur prédictive positive est faible). Ce résultat est cohérent avec les études menées dans d'autres pays où la classification policière de la gravité a été comparée avec le MAIS ou l'ISS. Une étude en Grande-Bretagne (Dove, Pearson et al. 1986) a trouvé que, parmi les blessés classés « sérieux » par la police, un tiers était en fait ISS 1-3 (gravité mineure) et un tiers ISS 4-9 (gravité modérée). Une étude plus récente (Morris, Mackay et al. 2003), toujours en Grande-Bretagne, trouve ce même résultat d'un tiers de blessés ISS 1-3 parmi ceux classés comme « sérieux » par la police. Une petite étude aux États-Unis (Agran, Castillo et al. 1990) montre que parmi 52 blessés ISS 9+, la moitié n'est pas considérée comme étant de gravité « sévère » par la police.

Cependant, même s'il y a discordance entre classification policière et NISS, il y a une forte association entre les deux, comme le montrent la figure 1 et les résultats des modèles multivariés. Les blessés aux deux extrémités de l'échelle NISS (les plus légèrement et les plus gravement blessés) sont les moins à risque d'erreur de classement ; cela n'a rien de surprenant, c'est même plutôt rassurant sur la qualité du classement policier. La discordance, et en même temps la forte association entre classification policière et NISS ne sont pas très surprenantes, vus les critères utilisés par les forces de l'ordre. S'ils utilisent effectivement la durée d'hospitalisation ou s'ils utilisent le nombre de jours d'ITT (incapacité totale temporaire), ils utilisent des critères qui ne mesurent pas la gravité elle-même, mais qui y sont fortement corrélés, et donc corrélés au NISS. Dans les situations où les forces de l'ordre évaluent la gravité « à vue de nez » sur le lieu de l'accident, il est tout à fait vraisemblable que leur évaluation soit liée au nombre de blessures, aux régions corporelles atteintes, voire au type de blessures. Leur évaluation est donc bien liée à la gravité réelle (et donc au NISS).

Concernant la direction globale de ces erreurs de classement, il est très tentant de comparer les effectifs des blessés dont la gravité est sur-estimée et ceux où elle est sous-estimée (respectivement 1444 et 306) et de résumer le problème en concluant que globalement les forces de l'ordre sur-estiment la gravité, et ainsi amoindrir cette question des erreurs de classement. Ceci est erroné car ce serait oublier le problème du sous-enregistrement et surtout le fait qu'il est biaisé par rapport à la gravité lésionnelle. En France (cf. chap. 1), comme dans les autres pays (Hvoslef 1994), le sous-enregistrement est inversement lié à la gravité : plus les blessures sont légères, moins il est probable que le blessé soit enregistré dans les données des forces de l'ordre. Ce biais signifie que le nombre de blessés NISS 1-15 doit être multiplié par un facteur de correction du sous-enregistrement bien plus grand que le nombre de blessés NISS 16-75, et il est donc impossible, à cette étape, de dire quelle est globalement la direction de l'erreur.

Cette étude confirme que les erreurs de classement de la gravité par les forces de l'ordre sont différentielles, et identifie quelles sont les caractéristiques liées à ces erreurs de classement. Les piétons et usagers de 2RM blessés ont une probabilité plus élevée d'avoir leur gravité sur-estimée, toutes choses égales par ailleurs. Ils font partie des usagers vulnérables, c'est-à-dire ceux qui ne sont pas protégés par une carrosserie, et qui, à choc de violence égale, sont en moyenne plus gravement blessés que ceux à l'intérieur d'un habitacle. Il n'est pas surprenant que cette gravité plus élevée attendue joue un rôle dans l'évaluation faite par les forces de l'ordre. Dans une étude australienne (Lopez, Rosman et al. 2000), le type d'usager n'est pas significativement associé avec la concordance entre classement policier et admission à l'hôpital (oui/non), mais l'étude n'était basée que sur 406 sujets. L'association avec le type d'usager n'a pas été étudiée dans les deux autres études similaires (Hansen, Popkin et al. 1991; Farmer 2003). Une étude française (Laumon and Martin 2002), basée aussi sur le département du Rhône mais limitée aux données de 1996 avait déjà trouvé que les piétons blessés avaient une probabilité plus élevée d'être classés « blessés graves » par les forces de l'ordre (à MAIS égal).

L'âge croissant des blessés est associé à une plus grande probabilité de sur-estimation de la gravité par les forces de l'ordre. L'âge est un facteur de risque de blessure grave et aussi de plus longue durée d'hospitalisation (Bergeron, Lavoie et al. 2005). Dans les cas où les forces de l'ordre appliquent strictement la définition de leur critère de gravité (durée d'hospitalisation supérieure ou inférieure à six jours), il est logique que leur classification soit liée à l'âge. Dans les cas où les forces de l'ordre estiment la gravité au jugé, on peut penser que le sur-risque de gravité lié à l'âge soit perçu ou connu d'une certaine façon des forces de l'ordre et joue ainsi sur leur évaluation. Ce résultat est cohérent avec l'étude australienne (Lopez, Rosman et al. 2000) qui montre que l'âge croissant des usagers de véhicules motorisés est associé avec une probabilité d'erreur de classement de la gravité. En revanche, dans l'étude américaine (Farmer 2003), un âge au dessus de 65 ans est inversement associé avec un risque de sur-estimation de la gravité.

L'étude met en évidence de grandes différences dans les probabilités de sous- ou sur-estimation de la gravité entre les trois forces de l'ordre. Une étude aux États-Unis (Shinar, Treat et al. 1983) analysant la validité des données policières d'accidentologie a aussi trouvé des différences entre les différents corps de police (State/county/municipal). Les auteurs ont attribué une grande partie de ces différences aux différences entre leurs zones de contrôle. Nous avons exploré cette possibilité dans l'étude : lors de la construction des modèles multivariés, la variable environnement (en agglomération vs hors agglomération) devient non significative dès que la variable force de l'ordre est introduite dans les modèles de la sous- et sur-estimation de la gravité ; la variable type de réseau devient non significative dans le modèle de la sur-estimation de la gravité ; elle reste significative dans le modèle de sous-estimation de la gravité, et les risques relatifs ajustés (entre autres sur cette variable) comparant les forces de l'ordre restent élevés. Cela signifie que les caractéristiques de réseau, de caractère urbain/rural (qui sont différentes entre les zones attribuées aux trois forces de l'ordre) n'expliquent qu'une petite partie des différences (de classement de la gravité) entre les trois forces de l'ordre. Nous avons vérifié qu'il ne restait pas de différence en termes de gravité lésionnelle « réelle » dans le groupe des blessés NISS 25-75 entre les différentes forces de l'ordre (les trois médianes sont très proches). En Grande-Bretagne, des différences de classement de la gravité entre les forces de l'ordre ont aussi été observées : une étude (Lai 1980) a montré que l'ampleur de la différence entre les forces de l'ordre était plus grande que celle entre les caractéristiques d'environnement des accidents : urbain/rural (combiné avec jour/nuit) qui différencient les zones des forces de l'ordre.

Sur la période d'étude 1997-2001, la probabilité de sur-estimation de la gravité a diminué, mais la probabilité de sous-estimation a augmenté (ajustées sur les autres variables du modèle). Cela peut être vu comme une seule évolution : les blessés sont de plus en plus souvent classés « blessés légers » par les forces de l'ordre, à NISS donné. Il est vrai qu'il y a une tendance générale vers des durées d'hospitalisation plus courtes (changement de politiques hospitalières, mais aussi amélioration de l'efficacité des soins) mais cette tendance n'est pas visible sur une période de cinq ans et elle ne concernerait qu'une partie des blessés, ceux qui sont réellement classés en fonction de leur durée effective d'hospitalisation.

Forces et faiblesses de l'étude

Le NISS n'est certes pas le gold standard qui mesure parfaitement la gravité lésionnelle. Le NISS a été choisi car il est basé sur une échelle d'évaluation des blessures (Abbreviated Injury Scale, AIS) largement utilisée et reconnue. Les analyses ont été répliquées avec les autres mesures basées sur l'AIS, à savoir l'ISS et le MAIS. Nous avons vérifié que les résultats n'étaient pas sensibles à la mesure de gravité utilisée. Les scores NISS ont été regroupés en deux catégories, définies de façon à pouvoir évaluer la discordance entre la classification NISS et la classification policière ; cette discordance a été nommée erreur de classement, et séparée en sous-classement et sur-classement ; il s'agit de simplifications mais elles sont nécessaires et elles restent informatives.

Les points forts de l'étude sont liés à ceux du Registre. Ce Registre n'est pas limité à quelques hôpitaux (comme cela est parfois le cas) mais il couvre l'ensemble des hôpitaux du département du Rhône et des proches alentours. Il inclut les hospitalisés et les non-hospitalisés qui forment la majorité

des blessés (alors qu'ils ne sont pas inclus dans des bases de données hospitalières nationales qui existent dans d'autres pays, notamment en Europe du Nord). En outre, l'étude porte sur une période de cinq ans et donc sur un grand nombre de blessés.

L'hypothèse sous-jacente de l'étude est que chaque force de l'ordre a une pratique homogène de classement des blessés sur tout le territoire français. La structure des trois forces de l'ordre étant nationale et centralisée (et non régionale ou départementale), chaque force de l'ordre reçoit des instructions nationales. Sous cette hypothèse, les résultats de l'étude peuvent être généralisés à l'ensemble des données des forces de l'ordre des victimes de la circulation routière en France.

Conclusion

Les résultats de l'étude conduisent à une certaine prudence dans l'interprétation des études basées sur les données des forces de l'ordre. Le classement de la gravité en blessés légers / graves étant lié à certaines caractéristiques de l'accident et du blessé, la description et la comparaison de la gravité (des BAAC) entre différents sous-groupes peuvent être trompeuses. Aussi, quand on utilise les données des forces de l'ordre pour étudier l'association entre la gravité (au sens des BAAC) et de potentiels facteurs de risque de la gravité, le fait que les erreurs soient différentielles (i.e. liées à des caractéristiques de la victime et de l'accident) conduit en général à amplifier le risque relatif (ou l'odds ratio) qui mesure cette association.

La classification effectuée par les forces de l'ordre ne peut être qu'une approximation. Les forces de l'ordre n'ont ni les compétences ni les moyens d'établir un diagnostic des blessures. Le bilan lésionnel des victimes recensées par le Registre est basé sur le diagnostic de médecins, appuyés par des moyens techniques (radiologie, imagerie…). Le bilan est consolidé, c'est-à-dire qu'il est basé sur l'ensemble des diagnostics établis dans les différents services que le patient a consulté. Enfin, ce bilan est codé selon la classification AIS par des médecins spécifiquement formés à cette classification.

La classification policière de la gravité est positivement corrélée au NISS. Elle est liée à des caractéristiques de l'accident ou du blessé, caractéristiques qui sont en général prédictives d'une plus grande gravité (cf. chap. 2). Cependant, à gravité égale (catégorie de NISS), la probabilité de sous-classement augmente sur la période étudiée.

Pour toutes ces raisons, nous ne pouvons pas nous contenter de cette classification pour établir un bilan de la morbidité par accident de la circulation routière.

Résumé

Contexte :
Les données des forces de l'ordre étant les seules données disponibles à l'échelle nationale, il importe de connaître leur qualité. Nous nous intéressons ici à la classification policière de la gravité lésionnelle, en deux catégories, blessés « légers » ou « graves », et nous voulons savoir comment cette classification se situe par rapport à une échelle médicale de gravité.

Méthode :
L'étude porte sur le département du Rhône (1,6 million d'habitants) et sur la période 1997-2001. Les données des forces de l'ordre ont été chaînées avec le Registre des victimes de la circulation routière. Nous disposons ainsi des deux classifications de la gravité, policière et médicale (New Injury Severity Score, NISS), pour 14 342 blessés identifiés comme communs aux deux sources de données. La classification policière de la gravité en blessés « légers » et blessés « graves » est comparée à la classification NISS groupée en NISS 1-15 et NISS 16-75. Nous utilisons ensuite un modèle multivarié pour exprimer la probabilité d'erreur de classement de la gravité par les forces de l'ordre, de sur et de sous-classement, en fonction de caractéristiques du blessé et de l'accident. Autrement dit, nous explorons en quoi les erreurs de classement sont non-différentielles. Les résultats sont donnés sous forme de risques relatifs ajustés et de probabilités ajustées.

Résultats :
La sensibilité de la classification des forces de l'ordre est estimée à 72 % (IC à 95% : 69,0% - 74,4%) et la valeur prédictive positive à 35 % (IC : 33,0% - 37,0%). Les blessés de gravité modérée sont les plus susceptibles d'être mal classés : le risque relatif est de 12,2 entre les blessés NISS 9-15 et les blessés NISS 1-3 pris comme référence (IC : 10,6-14,2). Les piétons et usagers de 2RM sont les usagers les plus susceptibles d'avoir leur gravité sur-classée (RR ajusté=1,4 (1,2-1,6) et 1,2 (1,1-1,4) respectivement par rapport aux automobilistes). La gendarmerie est trois fois plus susceptible de sur-classer que les autres forces de l'ordre. Sur la période d'étude 1997-2001, la probabilité de sur-classer diminue mais la probabilité de sous-classer augmente.

Conclusion :
La classification de la gravité effectuée par les forces de l'ordre ne peut être qu'une approximation de la gravité lésionnelle réelle. Nous avons pu vérifier que la classification effectuée est positivement corrélée au NISS ; elle est influencée par des caractéristiques de l'accident ou du blessé, caractéristiques qui sont effectivement liées à la gravité. Cependant, la probabilité de classer un blessé, à gravité réelle donnée, comme « blessé léger » augmente sur la période étudiée. Nous ne pouvons pas nous contenter du classement des forces de l'ordre dans le but d'établir un bilan de la morbidité par accident de la circulation routière.

Chapitre 3
Bilan rhodanien : nombre de blessés et gravité lésionnelle

Introduction

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 ci-après.

Nous appliquons cette méthode sur le département du Rhône où coexistent les données des forces de l'ordre et le Registre. Parmi les conditions d'application figure l'homogénéité de capture. Celle-ci n'étant vérifiée que par sous-groupes, nous stratifions sur les trois principales variables caractérisant ces sous-groupes : la gravité lésionnelle, le type d'usager et l'absence / présence de tiers. Une autre condition d'application de la méthode est l'identification parfaite des individus présents dans les deux sources d'enregistrement. Afin de nous en approcher, nous tentons d'améliorer le chaînage effectué en routine entre les données des forces de l'ordre et le Registre. Nous construisons ainsi trois scénarios qui se distinguent par le nombre estimé de blessés communs aux deux sources, et nous appliquons la méthode de capture-recapture sur chacun des trois scénarios.

Ce chapitre reprend l'article publié dans Accident Analysis and Prevention (cf. annexe) et le complète. Il correspond à une étape intermédiaire d'application de la méthode de capture-recapture aux données d'accidentologie du Rhône. Nous utilisons ici une méthode simple (sans modélisation explicite) qui a ses limites, mais qui permet de bien appréhender le principe de l'approche capture-recapture. Par ailleurs, la publication dans Accident Analysis and Prevention, revue de référence en accidentologie, d'une méthode simple avec stratification sur trois variables est importante dans la mesure où, dans ce domaine, l'utilisation de capture-recapture n'est pas très développée (limitée à des sous-populations, et avec trop peu de prise en compte des facteurs de biais de sélection).

L'application de capture-recapture aux données d'accidentologie du Rhône sera améliorée par la suite (chapitre 4) par une modélisation multinomiale logistique. Elle nous permettra notamment de prendre en compte l'ensemble des variables liées aux probabilités d'enregistrement.

Synthèse bibliographique sur capture-recapture

Trois articles (Hook and Regal 1995; IWGDMF 1995; Gallay, Nardone 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, de marquer les individus capturés, de les relâcher, 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 1ère fois, seulement la 2ème 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.

Par transposition des captures à des listes d'enregistrement, l'approche a été utilisée dans d'autres domaines : en démographie, pour mesurer la couverture d'un recensement (Fienberg 1992; Garton, Abdalla 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, Bloor et al. 1991; Abeni, Brancato et al. 1994), personnes atteintes du SIDA (Bernillon, Lievre et al. 2000), toxicomanes (Hay, McKeganey et al. 1999; Maxwell 2000; Gemmell, Millar et al. 2004), prostituées (Bloor, Leyland et al. 1991), personnes sans domicile fixe (Fisher, Turner et al. 1994)…. 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 1995; IWGDMF 1995), malformations congénitales, cancer (Sherman 1981; Robles, Marrett et al. 1988; Hilsenbeck, Kurucz et al. 1992; Brenner, Stegmaier et al. 1994; Schouten, Straatman et al. 1994; Brenner, Stegmaier et al. 1995; Crocetti, Miccinesi et al. 2001), VIH et SIDA (Frischer, Leyland et al. 1993; Abeni, Brancato et al. 1994; Johri, Kaplan et al. 1999; Bernillon, Lievre et al. 2000), traumatologie (Roberts and Scragg 1994; LaPorte, Dearwater et al. 1995; Johnson, Gabella et al. 1997; Rosenman, Kalush et al. 2006).

La méthode de capture-recapture est aussi utilisée dans l'épidémiologie des traumatismes de la circulation routière. Elle a été appliquée sur les sous-populations suivantes : enfants (Roberts and Scragg 1994; Jarvis, Lowe et al. 2000), adolescents et jeunes adultes (Morrison and Stone 2000), piétons et cyclistes (Dhillon, Lightstone et al. 2001), conducteurs de poids lourds (Meuleners, Lee et al. 2006), mais aussi à des zones géographiques : villes (Razzak and Luby 1998; Tercero and Andersson 2004) ou une île (la Réunion) (Aptel, Salmi et al. 1999).

L'application de la méthode de capture-recapture en épidémiologie est sujette à des prises de position très tranchées. Certains chercheurs sont très enthousiastes et pensent qu'il vaut mieux 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; Hook and Regal 2000) 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, Lowe et al. 2000; Morrison and Stone 2000) ; cependant, leur discussion des conditions d'applications n'apparait pas complète.

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 calculette 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 1995; Chao, Tsay 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 2000) et autres qui n'ont pas été explorés ici.

Matériel

L'étude est restreinte aux blessés par accidents de la route de l'année 2001. Elle est basée sur les données des forces de l'ordre (BAAC), les données du Registre, les données issues du chaînage entre les deux, mais aussi sur les résultats des études complémentaires (blessés identifiés dans les BAAC seulement, ou identifiés dans le Registre seulement) qui sont restreintes à l'année 2001.

Données policières d'accidentologie

Elles ont été décrites précédemment (chap. 1). Sur l'année 2001, les données des forces de l'ordre comptent 4135 blessés, légers ou graves (et non-décédés dans les six jours).

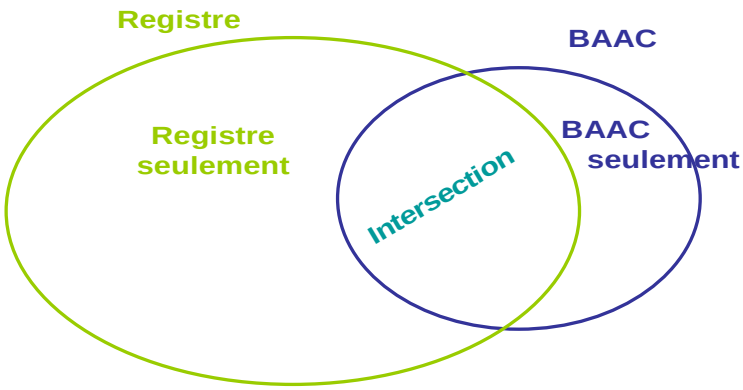
Registre des victimes de la circulation routière dans le Rhône

Il a été décrit précédemment (chap. 1). Sur l'année 2001, le Registre comptabilise 10 636 blessés (non décédés dans les six jours) et satisfaisant aux critères BAAC.

Données rapprochées

La méthode de capture-recapture repose sur les effectifs des trois sous-groupes « Registre seulement », « BAAC seulement » et « intersection BAAC-Registre », et sur l'hypothèse d'identification parfaite des individus présents dans les deux sources d'enregistrement. Cela dépend de l'efficacité du chaînage entre BAAC et Registre.

Figure 6 : répartition des blessés recensés par le Registre et par les BAAC



Lors de la procédure de rapprochement ou chaînage (record-linkage), on compare les victimes des BAAC 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 :

Tableau 29 : description d'un résultat de chaînage entre deux fichiers d'enregistrement (unité=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 réfère au résultat de la décision : positif signifie rapproché ou « relié » ou « chaîné », et négatif signifie non-rapproché. Faux et vrai qualifient la décision prise au regard de la réalité. 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

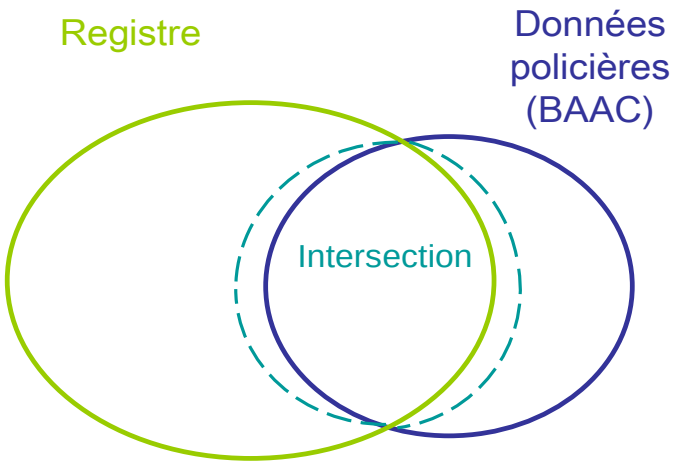
(réellement) à des individus distincts est noté U pour Unmatched. Enfin, « relié » se traduit par « linked ».

Tout comme dans les études diagnostiques, certains critères mesurent la qualité du record-linkage (Clark 2004) :
Sensibilité : effectif de vrais positifs / (effectif de vrais positifs + effectif de faux négatifs). Cela correspond à la probabilité qu'une paire d'enregistrements d'une même victime soit reconnue comme telle par la procédure.
Spécificité : vrais négatifs / (vrais négatifs + faux positifs). Cela correspond à la probabilité qu'une paire d'enregistrements de deux victimes distinctes soit reconnue comme telle par la procédure, i.e. soit non-reliée.
Valeur prédictive positive : vrais positifs / (vrais positifs + faux positifs). Cela correspond à la probabilité qu'une paire reliée soit réellement une paire, autrement dit de la même victime.
Valeur prédictive négative : vrais négatifs / (vrais négatifs + faux négatifs). Cela correspond à la probabilité qu'une paire non-reliée soit réellement une paire d'enregistrements de deux victimes distinctes.
Une victime est identifiée par une personne ET un accident.

Trois scénarios de chaînage

Par rapport à la condition requise pour l'approche capture-recapture, d'identification parfaite des individus communs aux deux sources d'enregistrement, nous construisons trois scénarios plausibles de la situation de recoupement entre les données des forces de l'ordre et le Registre ; ils se différencient par la taille de l'intersection BAAC-Registre.

Figure 7: redéfinition de la taille de l'intersection BAAC-Registre



Le scénario n° 1 est constitué par le résultat du chaînage « standard » (effectué chaque année en routine au sein de l'UMRESTTE). Le scénario n° 2 correspond à la situation standard améliorée par un chaînage supplémentaire basé sur les noms et prénoms. Le scénario n° 3 correspond à la situation standard (scénario n° 1) améliorée par l'estimation du nombre de faux positifs et de faux négatifs. Un quatrième scénario avait été envisagé ; nous pensions que les études complémentaires nous fourniraient directement le nombre de blessés manquant à l'intersection BAAC-Registre ; cela n'est pas le cas (cf. discussion).

Scénario n° 1 de chaînage des blessés BAAC et Registre

Une procédure de chaînage est appliquée chaque année pour identifier les victimes communes au Registre et aux BAAC. Cette procédure est basée sur une approche semi-automatisée et probabiliste.

Les variables de chaînage sont la date de l'accident (année, mois, jour), le lieu de l'accident, le type d'usager, l'année de naissance, le mois de naissance, et le sexe de la victime. La procédure n'est pas complètement automatisée car la variable lieu de l'accident est laissée en texte libre afin de ne perdre aucune information. Elle est en effet très riche car elle contient le nom de la commune et souvent la rue ou la route. La décision de chaîner une paire dépend de combien de variables concordent et lesquelles (cf. ci-dessous).

Cette procédure de routine a été utilisée sur les données de 2001. Nous rappelons que les blessés dans les BAAC sont au nombre de 4135, les blessés dans le Registre 10 636. Le nombre de blessés communs aux BAAC et au Registre est de 2813 dans ce scénario. Le nombre de blessés identifiés seulement dans les BAAC est de 1322[3].

Scénario n° 2 de chaînage des blessés BAAC et Registre

Ce scénario correspond au chaînage « de routine » amélioré par un chaînage supplémentaire basé sur les noms et prénoms. Ces informations sont présentes dans le Registre ; pour les données des forces de l'ordre, elles ont été extraites des procès-verbaux dans le cadre de l'étude « BAAC seulement » (cf. chap. 1). Ce chaînage a en outre bénéficié de quelques corrections des données BAAC à partir des procès-verbaux (PV), sur le mois, l'année de naissance, et sur le sexe pour un très petit nombre de blessés. Ce chaînage supplémentaire a été décrit au chapitre 1, partie 1.3.
Ce chaînage a permis d'identifier 94 victimes supplémentaires comme étant communes avec le Registre, soit 10,8 % de l'effectif des blessés de « BAAC seulement » (IC à 95% : 8,7% -12,9%). Nous estimons alors que sur l'ensemble des 1322 victimes identifiées seulement dans les BAAC, 143 d'entre elles seraient aussi dans le Registre. Autrement dit, cela ajoute 143 victimes à l'intersection BAAC - Registre (et autant de moins pour « BAAC seulement » et « Registre seulement ») par rapport au résultat du chaînage de routine. Le scénario n° 2 correspond donc à une situation où il y a 2956 victimes communes aux BAAC et au Registre.

Scénario n° 3 de chaînage des blessés BAAC et Registre

Il correspond au scénario « de routine » modifié par le nombre estimé de faux négatifs et de faux positifs. Les estimations des effectifs de ces groupes sont basées sur des calculs de probabilités et de recherche de valeurs approchées, et largement inspirées par les travaux de Brenner et Schmidtmann (Brenner 1994; Brenner and Schmidtmann 1996) ; la procédure détaillée est donnée en annexe.

Un faux négatif est une paire d'enregistrements non reliés, à tort, autrement dit une paire d'enregistrements du même blessé que l'on n'a pas rapprochés :

P(faux négatif)
= P(ne pas rapprocher alors qu'il s'agit d'enregistrements d'une même victime)
= P(ne pas rapprocher / même victime)
= 1-P(rapprocher / même victime)
= 1-P(concordance sur toutes ou certaines variables / même victime)

Concordance « sur toutes ou certaines variables de chaînage » est précisément défini. Les variables de chaînage sont : 1) le jour de l'accident 2) le mois de l'accident, 3) le lieu de l'accident 4) le mois de naissance de la victime, 5) l'année de naissance de la victime 6) le sexe de la victime 7) le type d'usager de la victime. L'année de l'accident n'est pas considérée comme une variable de chaînage car elle est utilisée comme variable de blocage, c'est-à-dire que la procédure de chaînage est conduite par blocs, définis par les années. Cela suppose l'absence d'erreur sur cette variable. Pour les accidents ayant lieu en fin d'année (d'après les BAAC), les enregistrements du Registre des jours du début de l'année suivante sont explorés.

[3] Dans le chapitre 1 ; étude « BAAC seulement » ce nombre est de 1319 ; la discordance vient d'un problème de doublons dans le chaînage, identifié et résolu entre-temps.

Si lors de la comparaison des modalités BAAC et Registre d'une variable de chaînage, une des deux valeurs est manquante, nous considérons que le résultat de la comparaison est discordant.

La décision de relier une paire est prise si :

- A) les sept variables de chaînage concordent, ou
- B) seulement six variables concordent, ou
- C) 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.

Ces trois situations (A, B, C) étant disjointes, la probabilité de l'ensemble est la somme des trois probabilités. Chacune de ces trois probabilités, sous l'hypothèse d'indépendance entre les variables de chaînage, est le produit des probabilités de concordance pour chaque variable de chaînage (cf. annexe).

D'après Brenner et Schmidtmann, si l'on dispose du « gold standard » i.e. si l'on connaît le sous-ensemble M des paires correspondant à une même victime, on estime 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. Mais cela donnera plutôt une sur-estimation des probabilités de concordance. Nous décidons de diminuer ces proportions de 5 % (cf. annexe).
La probabilité de faux négatifs est évaluée à 0,130 (IC à 95% : 0,118 -0,142), et le nombre de faux négatifs à 396 (IC à 95% : 358 – 434).

Un faux positif est une paire d'enregistrements reliés à tort, autrement dit une paire d'enregistrements de deux victimes distinctes que l'on a rapprochés :

P(faux positif)
= P(rapprocher alors qu'il s'agit d'enregistrements de deux victimes distinctes)
= P(rapprocher / victimes distinctes)
= P(concordance sur toutes ou certaines variables / victimes distinctes)

La concordance « sur toutes ou certaines variables » est celle définie précédemment, de façon précise en fonction des sept variables de chaînage.

La concordance sur une variable alors qu'il s'agit de victimes distinctes correspond à une concordance par hasard. Les proportions de concordance sont estimées par les proportions observées sur l'ensemble (U) des paires d'enregistrements de victimes distinctes si l'on dispose du gold standard. À défaut, cet ensemble est approximé par l'ensemble des paires non reliées, voire par l'ensemble de toutes les paires possibles, car la différence entre ces deux sous-ensembles est négligeable (leurs tailles respectives sont $N_A \times N_B - N_A$ et $N_A \times N_B$, avec N_A négligeable devant $N_A \times N_B$). Nous choisissons d'approximer par l'ensemble des paires non reliées, soit 47 220 438 paires. Le tableau des proportions observées de concordance est donné en annexe.

La probabilité de faux positifs est évaluée à $2{,}1.10^{-6}$ (IC à 95 % = $1{,}7.10^{-6}$ - $2{,}5.10^{-6}$) ; elle est très faible mais elle est définie sur l'ensemble des paires d'enregistrements de victimes distinctes, dont le nombre s'élève à 47 millions. Il est ainsi estimé un effectif de 97 faux positifs (IC à 95% : 77 - 116).

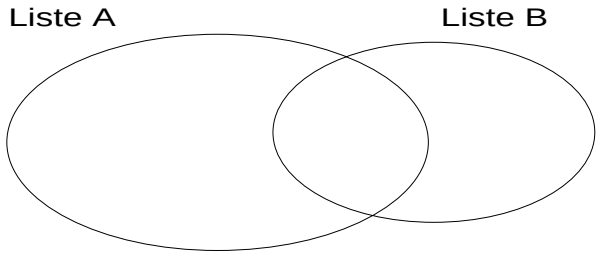
Le scénario n° 3 est le scénario « standard » amélioré par l'estimation de 97 faux positifs et 396 faux négatifs. L'intersection doit donc être augmentée de 299 blessés (396-97) et les sous-ensembles « BAAC seulement » et « Registre seulement » diminués d'autant. Le scénario n° 3 correspond donc à une situation où il y a 3112 blessés communs aux BAAC et au Registre.

Méthode

Capture-recapture

L'approche capture-recapture s'appuie sur au moins deux sources d'enregistrement (nommés listes) des sujets d'intérêt.

Figure 8: schéma de deux recensements (ou « listes »)



Le tableau ci-dessous précise les notations.

Tableau 30: 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 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 $\hat{n} = \frac{n_A \times n_B}{n_{AB}}$ qui est aussi l'estimateur du maximum de vraisemblance.

Comme nous allons stratifier (voir ci-dessous), et que certaines strates contiennent un faible nombre d'observations, nous utilisons l'estimateur NUE, Nearly Unbiased Estimator, (Wittes 1972) car l'estimateur du maximum de vraisemblance n'est qu'asymptotiquement non-biaisé. L'expression du NUE est la suivante :

$$\hat{n} = \frac{(n_A + 1) \times (n_B + 1)}{(n_{AB} + 1)} - 1$$

Cette formulation s'applique dans chaque strate.

Conditions d'application

La méthode de capture-recapture est basée sur quatre conditions clés :

- population fermée,
- indépendance d'enregistrement des deux sources,
- homogénéité de capture,
- identification parfaite des sujets communs aux deux sources.

Les conditions 1 et 2 seront discutées plus loin.

La condition d'homogénéité de capture signifie que toutes les victimes ont la même probabilité d'être enregistrées par une source donnée. 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, par stratification sur la variable définissant les sous-groupes ou par inclusion de la variable dans le modèle, si modélisation il y a.
Les variables définissant ces sous-groupes dans lesquels la probabilité d'enregistrement peut être considéré homogène sont en fait les facteurs de 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'usager, la présence / absence de tiers, le type de réseau, et le type de force de l'ordre (cf. chap. 1). En ce qui concerne les biais de sélection du Registre, 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'usager (sous-notification des cyclistes).
Nous décidons d'appliquer la méthode capture-recapture en stratifiant sur les facteurs de biais les plus importants : la gravité lésionnelle, le type d'usager et la présence/absence de tiers. Nous sommes limités par le nombre de variables sur lesquelles nous pouvons stratifier (trois est en général un maximum).
La gravité lésionnelle est mesurée par le NISS, regroupée en trois catégories : NISS 1-3 (gravité mineure), NISS 4-8 (gravité modérée) et NISS 9+ (gravité sérieuse et au-delà). Le type d'usager est défini en cinq catégories : piétons, cyclistes, usagers de 2RM, automobilistes, et autres usagers (camionnette, camion, bus…). Les strates présence/absence de tiers ne sont construites que pour les automobilistes et les usagers de 2RM ; elles ne sont pas construites pour les blessés cyclistes et usagers autres car les effectifs sont trop faibles.
La gravité lésionnelle mesurée par le NISS est disponible pour les blessés du Registre. Elle est aussi disponible pour une grande partie des blessés identifiés seulement dans les BAAC, grâce à l'étude « BAAC seulement ». Nous faisons ici l'hypothèse que la distribution du NISS évaluée sur les blessés avec procès-verbal exploitable et certificat médical présent est généralisable à celle des blessés sans certificat médical et/ou sans procès-verbal exploitable, c'est-à-dire à l'ensemble des blessés identifiés seulement dans les BAAC.

La quatrième condition suppose l'identification parfaite des blessés présents dans les deux sources. Ceci est difficile à réaliser, mais aussi à évaluer. Nous conduisons une sorte d'analyse de sensibilité, en appliquant la méthode capture-recapture sur les trois scénarios définis précédemment.

Agrégation des données et variables de stratification

Les scénarios n° 2 et n° 3 sont une amélioration du scénario n° 1, avec une redéfinition des tailles des sous-ensembles « BAAC seulement », « Registre seulement », et « intersection BAAC-Registre». Les individus correspondant ne sont pas nécessairement identifiés (pas du tout dans le scénario n° 3, et seulement en partie dans le scénario n° 2). Autrement dit, les scénarios n° 2 et n° 3 sont définis en termes d'effectifs et non en termes d'individus.

Afin d'appliquer la méthode de capture-recapture avec stratification, nous avons besoin pour chaque scénario de la répartition des effectifs en fonction des variables de stratification (NISS en trois classes, type d'usager et éventuellement tiers (oui/non)) et pour chaque sous-ensemble de blessés. Nous notons k ces sous-ensembles exclusifs :
- k=1 « BAAC seulement »
- k=2 « Registre seulement »
- k=3 « intersection BAAC-Registre »

Nous notons j les strates définies par le NISS, le type d'usager et la présence de tiers (oui/non). Nous notons $p_{1,kj}$ la répartition des strates j dans le sous-ensemble k du scénario n° 1, c'est-à-dire : $p_{1,kj}=n_{1,kj}/n_{1,k}$.

Nous supposons que les caractéristiques d'un sous-ensemble donné (BAAC seulement, Registre seulement, intersection) sont les mêmes d'un scénario à l'autre :

$$p_{2,kj} = p_{1,kj} \text{ et } p_{3,kj} = p_{1,kj}$$

Par exemple, les blessés, que l'on n'a pas identifiés comme communs aux deux sources lors du chaînage standard mais qui le sont en réalité, ne diffèrent pas (en termes de distribution de caractéristiques) de ceux identifiés comme tels par le chaînage standard. En effet, ce qui diffère entre les scénarios, c'est l'inefficacité de la méthode de chaînage, par manque d'information ou par information erronée sur les variables de chaînage. Nous appliquons donc les distributions du scénario 1 du NISS, du type d'usager et du tiers dans chaque sous-ensemble k, aux scénarios 2 et 3, sous-ensemble k :

$$n_{2,kj} = p_{1,kj} \times n_{2,k} \text{ et } n_{3,kj} = p_{1,kj} \times n_{3,k}$$

En fait, le scénario n° 1 est en partie basé sur une estimation de distribution : en effet, par défaut, le NISS n'est pas connu pour les blessés identifiés dans les BAAC seulement (au nombre de 1322). L'étude complémentaire reposant sur la consultation des PV a permis, grâce aux certificats médicaux inclus dans ces PV, d'évaluer le NISS pour 528 d'entre eux (parmi 868 avec un PV consultable). Il a été vérifié que ces 868 blessés étaient représentatifs de l'ensemble des 1322 blessés en ce qui concerne la gravité au sens des BAAC. La distribution du NISS observée sur les 528 blessés de « BAAC seulement » avec PV consultable est étendue à l'ensemble des blessés de « BAAC seulement ».

Traitement des valeurs manquantes

Comme il s'agit d'estimer le nombre total de blessés dans le Rhône, il n'est pas question d'exclure des observations pour cause de valeur manquante.

Il y a un faible nombre d'observations avec des valeurs manquantes sur le type d'usager ou la présence/absence de tiers. Les valeurs manquantes pour le type d'usager sont regroupées avec la catégorie « autres ». Les valeurs manquantes pour la présence/absence de tiers sont dans la catégorie « absence de tiers ». Cela concerne les données individuelles, avant qu'elles soient agrégées.

Il y a un très petit nombre d'observations avec un NISS manquant parmi les observations du Registre de 2001 (0,6 % pour « BAAC intersection Registre » et 1,4 % « pour Registre seulement »). Ce qui est développé dans le point précédent est utilisé ici : pour chacun des trois sous-ensembles (BAAC seulement, Registre seulement, intersection), la distribution observée sur les victimes avec valeurs renseignées est appliquée à la totalité des victimes du sous-ensemble.

Statistiques présentées

Nous présentons les effectifs totaux estimés par capture-recapture, et les incidences corrigées du sous-enregistrement ; elles sont calculées en rapportant le nombre total de blessés estimé par capture-recapture à la population du Rhône de 2001 fournie par l'INSEE. En d'autres termes, l'incidence est estimée en prenant la population du département du Rhône comme population de référence, c'est-à-dire, comme population exposée au risque d'accident de la route dans le département du Rhône. Cela suppose que le risque d'accident de la circulation est un risque de proximité (cf. discussion).

Résultats

Sur le département du Rhône en 2001, les données des forces de l'ordre recensent 4135 blessés (non décédés dans les six jours), et le Registre recense 10 636 blessés vérifiant les critères de définition policière d'un accident de la circulation.

Le scénario n° 1 du chaînage de routine identifie 2813 blessés communs au Registre et aux BAAC. Le résultat de ce chaînage de « routine » en termes de vrais/faux positifs et négatifs est donné dans le tableau ci-dessous. Nous connaissons les effectifs à la marge (colonne de droite) et nous avons estimé les effectifs de faux positifs et faux négatifs. Les autres effectifs en sont déduits.

Tableau 31 : résultat du chaînage de routine entre les données des forces de l'ordre et le Registre, Rhône, 2001

	situation réelle :		
décision :	même individu (matched)	individus différents (unmatched)	total
rapprochement	vrais positifs 2716	faux positifs 97	2813
non-rapprochement	faux négatifs 396	vrais négatifs 926	1322
total	3113	1023	4135

Les critères de qualité du chaînage de « routine » sont estimés : la sensibilité est de 87,0 %, la valeur prédictive positive est de 96,5 %, c'est-à-dire que la probabilité qu'une paire d'enregistrements d'une même victime soit reconnue comme telle est de 87 %, et la probabilité qu'une paire reliée soit réellement une paire d'enregistrements de la même victime est de 96,5 %.
La spécificité est de 91 % et la valeur prédictive négative est de 70 % : la probabilité qu'une paire d'enregistrements de deux victimes distinctes soit reconnue comme telle est de 91 %, et la probabilité qu'une paire non-reliée soit réellement une paire d'enregistrements de deux victimes distinctes est de 70 %.

Le scénario n° 2 (chaînage supplémentaire sur noms et prénoms) correspond à une situation où il y a 2956 blessés communs aux BAAC et au Registre.

Le scénario n° 3 est le scénario « standard » amélioré par l'estimation de 97 faux positifs et 396 faux négatifs. Il correspond à 3112 blessés communs aux BAAC et au Registre. En théorie, par construction de ce scénario, sensibilité, spécificité, valeur prédictive positive et valeur prédictive négative valent 100 %.

Les effectifs des sous-ensembles des trois scénarios sont donnés dans le Tableau 32.

Les résultats des analyses stratifiées de capture-recapture, par niveau de NISS, mais agrégés sur les strates de type d'usager et de tiers sont donnés dans le Tableau 32.

Tableau 32 : résultats de l'analyse capture-recapture stratifiée, agrégés sur type d'usager et tiers (oui/non), pour chaque scénario, Rhône, 2001

	effectif BAAC	effectif Registre	effectif intersection	effectif union des 2	effectif capture-recapture	incidence corrigée /100 000
scénario 1						
NISS 1-3	2 705	7 972	1 774	8 903	12 305	768
NISS 4-8	861	1 791	580	2 071	2 757	172
NISS 9+	570	873	459	984	1 067	67
total	4 135	10 636	2 813	11 958	16 129	1 007
scénario 2						
NISS 1-3	2 694	7 949	1 864	8 780	11 616	725
NISS 4-8	860	1 798	610	2 048	2 621	164
NISS 9+	581	889	483	987	1 057	66
total	4 135	10 636	2 956	11 815	15 294	954
scénario 3						
NISS 1-3	2 683	7 924	1 962	8 645	10 937	683
NISS 4-8	859	1 806	642	2 023	2 487	155
NISS 9+	594	906	508	992	1 048	65
total	4 135	10 636	3 112	11 659	14 471	903

En termes d'effectif de blessés, le scénario n° 1 comptabilise 12 000 blessés (enregistrés) dans le Rhône en 2001, et la méthode de capture-recapture estime leur nombre total à 16 100 (arrondi à la centaine). D'après le scénario 2, il y a 11 800 blessés enregistrés, et 15 300 au total. D'après le scénario 3, il y a 11 700 blessés enregistrés, et 14 500 au total.

Le nombre total de blessés graves (NISS 9+) estimé par les trois scénarios varie peu : entre 1050 et 1070 (arrondi à la dizaine). Le nombre de blessés de gravité modérée (NISS 4-8) varie entre 2490 et 2860 selon le scénario, le nombre de blessés de gravité mineure varie entre 10 940 et 12 310. En termes d'incidence, le nombre de blessés toutes gravités (NISS 1+) est estimé entre 900 et 1 000 pour 100 000 habitants.

Les résultats détaillés (par strate) pour le scénario 3 sont donnés dans le Tableau 33. Les résultats détaillés pour les scénarios n° 1 et n° 2 sont donnés en annexe.

Tableau 33 : résultats détaillés de l'analyse capture-recapture stratifiée, scénario n° 3, Rhône, 2001

type d'usager et tiers	gravité lésionnelle	effectif BAAC	effectif Registre	effectif intersection	effectif union des 2	effectif capture-recapture
piétons						
	NISS 1-3	284	593	207	670	811
	NISS 4-8	141	200	112	229	251
	NISS 9+	129	147	109	167	174
cyclistes						
	NISS 1-3	63	811	52	822	983
	NISS 4-8	21	312	17	317	398
	NISS 9+	19	74	15	78	89
usagers de 2RM						
avec tiers	NISS 1-3	306	689	249	745	846
	NISS 4-8	186	235	129	292	338
	NISS 9+	157	189	138	207	214
sans tiers	NISS 1-3	26	675	15	685	1090
	NISS 4-8	32	259	18	273	455
	NISS 9+	26	109	24	111	118
automobilistes						
avec tiers	NISS 1-3	1558	3652	1127	4084	5049
	NISS 4-8	327	457	252	532	593
	NISS 9+	161	189	128	222	238
sans tiers	NISS 1-3	319	1096	233	1182	1500
	NISS 4-8	130	248	98	280	328
	NISS 9+	87	154	78	163	172
autres						
	NISS 1-3	127	408	79	457	658
	NISS 4-8	21	97	17	102	124
	NISS 9+	14	44	14	44	44

Discussion

La méthode de capture-recapture repose sur quatre conditions.

La première condition de population fermée 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 (NISS 1-3) 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 recensés dans le Registre, et automobilistes, 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'usager 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.

La seconde condition suppose l'indépendance des deux enregistrements. Cette condition n'est pas vérifiée. 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 1995) qu'une corrélation positive se traduit par une sous-estimation du nombre total de blessés. Les effectifs et incidences estimés de blessés graves sont donc vraisemblablement des valeurs planchers.

La troisième condition, celle d'homogénéité de capture, n'est pas vérifiée sur l'ensemble des blessés, mais seulement par sous-groupes. Les facteurs jouant le plus sur la probabilité d'enregistrement dans les BAAC ou dans le Registre ont été pris en compte par stratification : type d'usager, présence de tiers et surtout gravité lésionnelle. De plus, celle-ci a été mesurée grâce à une échelle médicale (le NISS), basée sur la description anatomique des lésions. D'autres facteurs sont associés à la probabilité d'enregistrement dans les BAAC : le type de réseau combiné au caractère urbain/rural et au type de force de l'ordre (cf. chap. 1). La stratification ne permet pas de prendre en compte autant de variables. Il y sera remédié par une modélisation explicite (cf. chap. 4).

La quatrième condition d'identification parfaite des blessés communs aux deux sources est difficile à évaluer. Il n'y a pas de « gold standard » qui nous renseignerait sur la situation réelle de recoupement entre les BAAC et le Registre. Les résultats des trois scénarios nous apportent différents éclairages. Le chaînage supplémentaire avec les noms et prénoms (scénario 2) a permis de trouver des liens supplémentaires mais pas beaucoup plus. Ces liens n'avaient pas été trouvés avec la méthode de routine car ils contiennent deux ou trois erreurs sur les variables de chaînage, ou bien une valeur manquante pour le lieu d'accident (côté Registre) alors que celui-ci est une information majeure pour le chaînage. Le chaînage supplémentaire est basé sur un certain nombre de variables informatives : le mois et l'année de naissance, le nom et le prénom de la victime, et la date et le lieu de l'accident, ce dernier étant très informatif car souvent détaillé : commune, et nom de rue ou numéro de route. Il est cependant certain que des vrais positifs ont été « manqués » car le code Soundex n'est pas optimal pour les noms de la population française. Le scénario 3 est le scénario le plus réaliste, même s'il se base sur des approximations dans les calculs de probabilité.

Les trois scénarios donnent des estimations plutôt proches, dans une fourchette raisonnable. Les résultats sont même très proches pour les effectifs de blessés graves (NISS 9+). Cela conforte mutuellement les résultats.

Il existe (Gallay, Nardone et al. 2002) deux autres conditions à l'approche capture-recapture : pas d'erreur de classement des cas (autrement dit d'identification des individus d'intérêt), et même zone de couverture temporelle et géographique des sources d'enregistrement. Ces deux conditions sont rarement explicitées, et ont été négligées lors de cette étape. Elles sont prises en compte lors de l'étape suivante (cf. chap. 4).

Concernant le traitement des données manquantes, l'hypothèse de distributions identiques (du NISS, du type d'usager et du tiers) dans les trois scénarios pour un sous-ensemble donné (BAAC seulement, Registre seulement, intersection) est basée sur le fait que certaines victimes ne sont pas dans le bon sous-ensemble à cause du manque d'efficacité de la procédure de chaînage, en raison de valeurs erronées ou manquantes sur les variables de chaînage.
Cela suppose en particulier une situation « MCAR » (missing completely at random). Autrement dit, le fait que la variable lieu d'accident soit manquante (variable très importante dans la décision de chaîner ou non, et souvent manquante dans le Registre) ne doit pas être lié aux variables NISS, type d'usager et tiers. Il se pourrait que le lieu manquant soit lié au caractère rural/urbain par exemple ou à l'incapacité de répondre, à cause d'un trop jeune âge ou de blessures très graves. Or, la gravité des blessures et le caractère urbain/rural (via le type de force de l'ordre) définissent les strates du modèle d'enregistrement : le problème ne se pose donc pas à l'intérieur de chaque strate.
Par ailleurs, on note que le lieu d'accident est la seule variable aussi peu renseignée dans le Registre. Il en est sans doute ainsi car sa pertinence n'est pas perçue par le personnel médical (sa principale raison d'être étant de permettre le chaînage avec les données des forces de l'ordre). Cela indiquerait donc que son caractère manquant n'est pas lié aux caractéristiques du blessé ou de l'accident.

Un quatrième scénario basé sur les résultats des études complémentaires « Registre seulement » et « BAAC seulement » avait été envisagé. Nous pensions que ces études nous permettraient d'estimer combien de blessés communs aux deux sources échappaient à la procédure de chaînage.
D'une part, l'étude des blessés identifiés dans le Registre seulement indique que 25 % des répondants auraient eu un procès-verbal rédigé par les forces de l'ordre, et donc devraient être enregistrés dans les BAAC ; d'autre part, l'étude des blessés identifiés dans les BAAC seulement indique que 64 % des « répondants » sont allés consultés dans un service hospitalier couvert par le Registre (certificat médical issu par l'hôpital présent dans le PV).
Si on applique le taux de 25 % à l'ensemble des 8225 blessés identifiés dans le Registre seulement en 2001, on atteint le chiffre de 2056 blessés. Ainsi la procédure de chaînage aurait « manqué » l'identification de 2056 blessés communs aux deux sources. L'effectif des blessés identifiés dans les BAAC seulement étant de 1322, le nombre de 2056 n'est pas réaliste. Trois explications sont envisagées : 1) les répondants ne sont pas représentatifs des non-répondants et ces derniers sont moins susceptibles d'avoir alerté les forces de l'ordre et donc le pourcentage de PV établis est plus faible (cf. chap. 1) ; 2) le taux de 25 % est une sur-estimation dans le sens où les répondants ne peuvent pas toujours savoir si les forces de l'ordre ont rédigé un procès-verbal ou juste une main courante ; 3) un pourcentage non connu de PV n'est pas enregistré sous forme de BAAC. Au maximum, le nombre de blessés « manqués » par le chaînage est de 1322, l'effectif total des blessés identifiés seulement dans les BAAC.
Il serait plus réaliste d'appliquer la proportion de 64 % (de consultation hospitalière confirmée par un certificat médical) à ces 1322 blessés, ce qui conduit à 846 blessés qui seraient communs au Registre et au BAAC (en 2001) et qui auraient été « manqués » par la procédure de chaînage. Mais là encore cet effectif n'est pas très réaliste : la procédure de chaînage supplémentaire avec les noms et prénoms permet d'estimer à 143 le nombre supplémentaire de blessés communs. Ce chaînage utilise de nombreuses variables informatives et autorise un certain nombre d'erreurs dans les variables. Il n'est donc pas très plausible de penser qu'environ 700 autres blessés seraient « manqués ». Cet effectif de 846 blessés est aussi bien plus élevé (deux fois plus) que l'effectif de faux négatifs estimé (396 pour rappel).
Le plus réaliste est que, 1) parmi les blessés identifiés seulement dans les BAAC et ayant consulté un service hospitalier, un certain nombre d'entre eux n'ont pas eu de fiche de notification du Registre remplie (par omission, surcharge de travail des services…), et 2) parmi les blessés identifiés dans le Registre seulement, et ayant été en contact avec les forces de l'ordre, un certain nombre d'entre eux n'ont pas fait l'objet d'un procès-verbal ou celui-ci n'a pas été enregistré sous forme d'un BAAC. Cette idée de quatrième scénario a dû être abandonnée.

Le département du Rhône est le critère d'inclusion du Registre comme lieu d'accident et non comme lieu de résidence, qui est le critère de la plupart des registres médicaux. Il y a deux raisons à cette particularité. La première est qu'il n'est pas réaliste d'envisager une connaissance exhaustive ou

presque, des accidents des habitants du Rhône hors du département. La seconde raison, et la plus importante, c'est que l'insécurité routière est liée à la répartition des types de réseau, au taux d'urbanisation, etc., et donc il importe de se rapporter à une zone géographique. Il est donc plus pertinent de définir le critère d'inclusion géographique avec le département comme lieu de l'accident, plutôt que comme lieu de résidence.

Partant de là, les incidences de blessés dans le Rhône devraient être estimées en rapportant le nombre de blessés dans le Rhône à la population exposée au risque d'accident dans le Rhône. Celle-ci n'est pas connue. A défaut, on utilise la population du Rhône comme population de référence. Se pose alors la question de savoir si la proportion d'accidents hors du département de résidence est faible. Cela rejoint la discussion sur la condition de population fermée, où nous avons vu que la proportion de non rhodaniens ayant un accident dans le Rhône est petite mais non négligeable. On peut penser que cela a seulement un effet à la marge sur la valeur estimée de l'incidence, sauf pour les usagers de poids lourds pour lesquels la proportion d'accident hors du département de résidence est importante.
Le comité national des registres avait soulevé la question de la population de référence ; il avait été accepté, après argumentation, que la population du département du Rhône soit utilisée pour calculer l'incidence des blessés. La population du Rhône est la population moyenne annuelle ; l'exposition mesurée correspond donc à des personnes-années.

Conclusion

Nous rappelons que cette application simple de capture-recapture est une sorte d'étape intermédiaire. Elle nous a permis de bien appréhender le principe de la méthode capture-recapture, mais aussi de mieux cerner la taille du sous-ensemble de blessés communs aux données des forces de l'ordre et au Registre.

Nous allons généraliser l'approche capture-recapture aux données de l'ensemble de la période (1996-2004) pour lesquelles on ne dispose pas d'informations supplémentaires, telles que les noms et prénoms, ou un certificat médical, pour les blessés enregistrés par les BAAC. Il n'est en effet pas envisagé de retourner aux procès-verbaux pour l'ensemble des quelques 10 000 blessés identifiés dans les BAAC seulement sur la période 1996-2004. L'absence des noms et prénoms n'est pas gênante car le scénario 2 basé sur cette information n'est pas le plus réaliste. Le scénario 3 basé sur l'estimation du nombre de faux négatifs et de faux positifs peut s'étendre aux autres années. Il reste à prendre en compte la nouvelle stratégie de chaînage, pour les données postérieures à 2001, suite à la modification de la procédure d'enregistrement des fiches dans le Registre.
Ne disposant pas du NISS pour les blessés identifiés seulement dans les BAAC, il faudra donc le prédire. L'idée est de construire un modèle de prédiction du NISS en fonction de variables disponibles dans les BAAC. Ce modèle sera construit à partir des blessés identifiés dans les deux bases, BAAC et Registre, pour lesquels nous disposons du NISS observé et des variables BAAC (gravité au sens BAAC, caractéristiques du blessé, de l'accident, de l'environnement).

Outre la généralisation aux autres années, nous souhaitons prendre en compte les autres facteurs fortement liés au sous-enregistrement, à savoir le type de réseau et le type de force de l'ordre. Cela sera possible par une modélisation explicite.

Résumé

Contexte :
Si l'on dispose de deux sources d'enregistrement, il est possible d'obtenir une bonne estimation du nombre total de blessés. Dans le département du Rhône, sur l'année 2001, les données des forces de l'ordre dénombrent 4135 blessés, et le Registre 10 636. Grâce aux études complémentaires menées, nous disposons d'informations supplémentaires sur les données de 2001 : gravité lésionnelle des blessés de « BAAC seulement », et noms et prénoms permettant un chaînage supplémentaire.

Méthode :
Afin d'estimer le nombre de blessés non-enregistrés et donc le nombre total de blessés, nous utilisons la méthode de capture-recapture. L'homogénéité de capture n'est vérifiée que par sous-groupes ; nous stratifions donc sur trois variables caractérisant fortement ces sous-groupes : la gravité lésionnelle (mesurée par le NISS), le type d'usager et la présence/absence de tiers. Afin d'évaluer la sensibilité de l'estimateur en fonction du nombre de blessés communs aux deux sources d'enregistrement, nous considérons trois scénarios de chaînage de données.

Résultats :
Selon le scénario le plus plausible, il y a 3112 blessés en commun au Registre et aux données des forces de l'ordre et 11 659 blessés en réunissant les deux sources. La méthode de capture-recapture estime alors le nombre total à 14 471 blessés. En termes d'incidence cela correspond à 903 blessés/100 000 habitants, et à 65/100 000 pour les blessés graves (NISS 9+). Compte tenu d'une dépendance positive entre les données des forces de l'ordre et le Registre, les effectifs et les incidences sont vraisemblablement des sous-estimations.

Conclusion :
La méthode de capture-recapture est applicable aux données des blessés par accidents de la route, en tenant compte de certains écarts aux conditions d'application. L'application sera améliorée au chapitre suivant ; en particulier, elle tiendra compte de l'ensemble des facteurs majeurs de biais de sélection des données des forces de l'ordre (gravité lésionnelle, type d'usager, présence/absence de tiers, type de réseau et type de force de l'ordre). Elle sera étendue à la période 1996-2004.

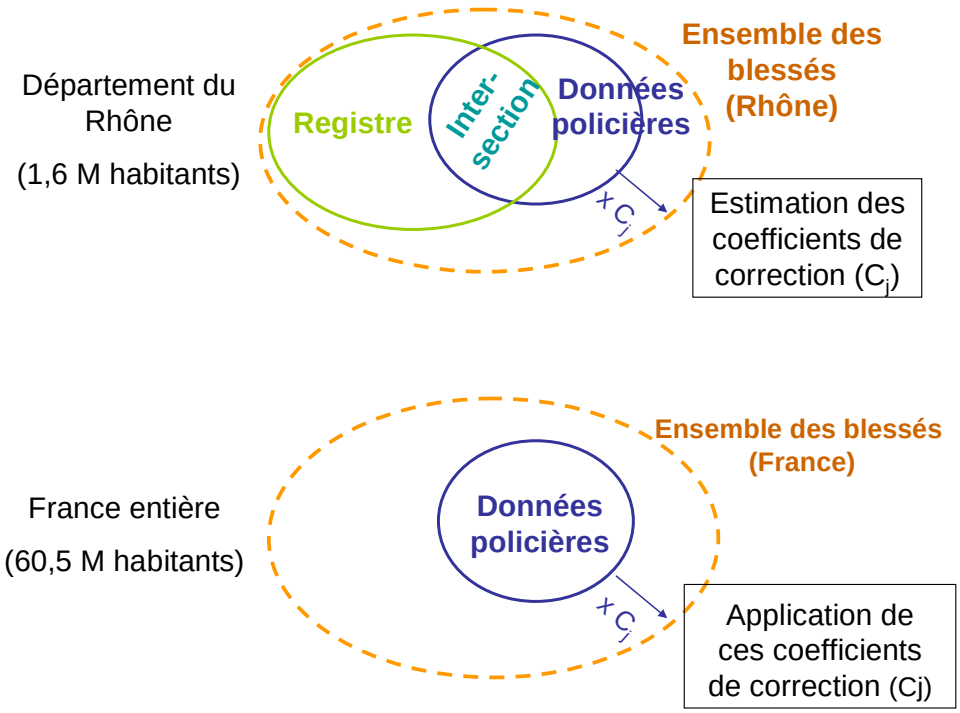
Chapitre 4
Bilan national : nombre de blessés et gravité lésionnelle

Introduction

Ce chapitre fait l'objet d'un article soumis à publication dans une revue anglophone de santé publique (cf. annexe).

Il s'agit maintenant d'estimer le bilan national des blessés de la circulation routière. L'idée est de combiner deux méthodes rencontrées séparément dans la littérature (synthèse bibliographique en annexe). La première est celle de capture-recapture afin d'estimer le bilan exhaustif au niveau départemental, et la seconde est de projeter des résultats locaux (département, région..) à un territoire plus vaste.
Nous construisons au niveau du Rhône un modèle permettant de prédire le bilan exhaustif (du Rhône) à partir des données des forces de l'ordre (du Rhône), et nous appliquons ensuite ce modèle aux données nationales des forces de l'ordre pour obtenir ainsi une estimation du bilan exhaustif national.
L'hypothèse sous-jacente est celle d'homogénéité sur l'ensemble du territoire français des pratiques des forces de l'ordre d'enregistrement des victimes de la circulation routière, par type de force de l'ordre et par catégories définies par les autres facteurs majeurs de biais de sélection. Le principe de la projection est illustré par la Figure 9.

Figure 9 : principe de la projection



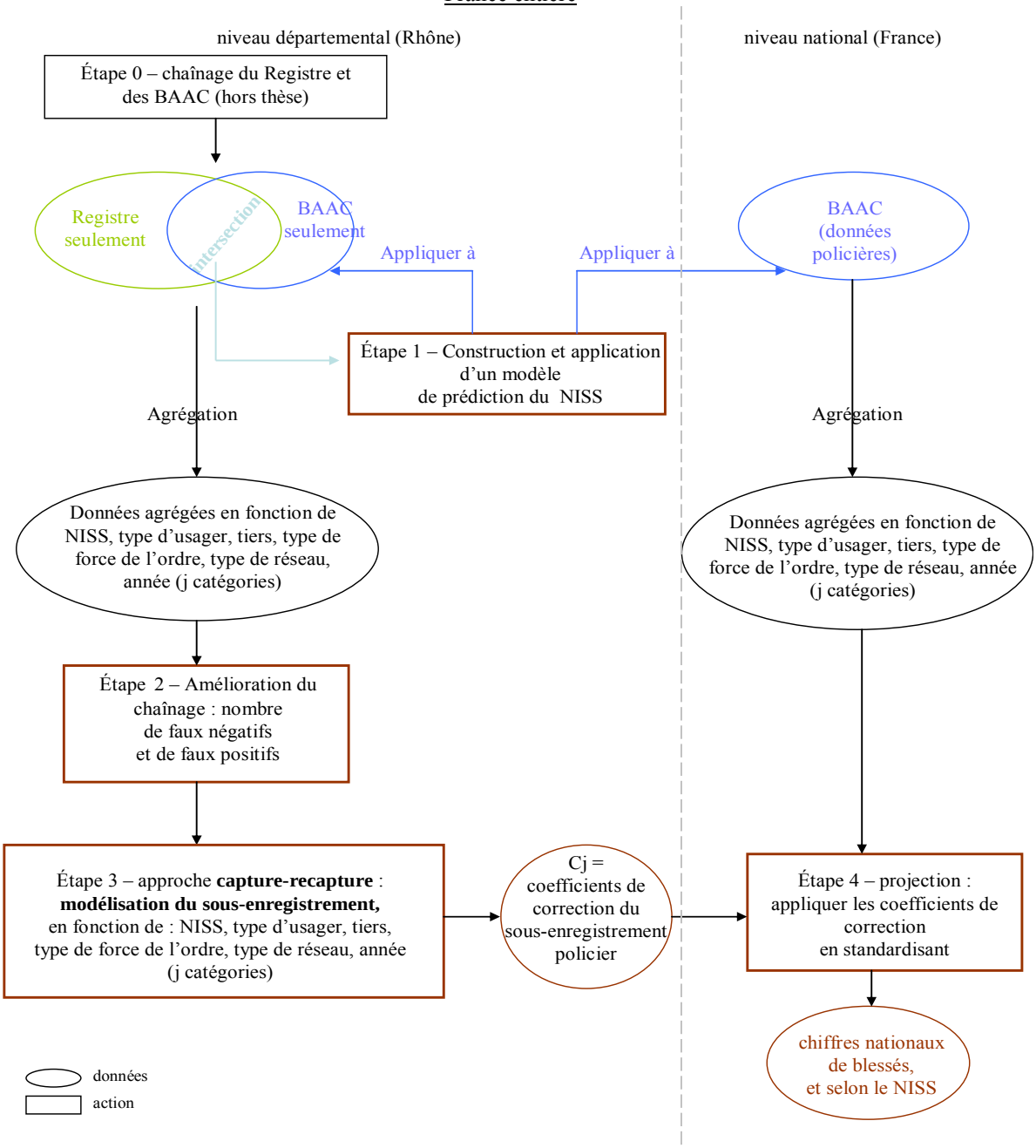
Au chapitre précédent, nous avons appliqué la méthode de capture-recapture aux données du Rhône de 2001, sur lesquelles des études complémentaires avaient été menées. Nous généralisons cette approche à l'ensemble de la période 1996-2004, et nous l'améliorons en tenant compte de l'ensemble des facteurs importants de biais de sélection : gravité lésionnelle, type d'usager, tiers (oui/non), type de force de l'ordre et type de réseau. Pour cela, nous utilisons une modélisation explicite, en l'occurrence

un modèle multinomial logistique. Il exprime la probabilité d'enregistrement dans chaque source. Cette modélisation nous permet d'obtenir directement des coefficients de correction du sous-enregistrement des données des forces de l'ordre. Ceux-ci sont définis en fonction des catégories des facteurs de biais de sélection (facteurs de sous-enregistrement).

La projection au niveau national se fait en appliquant les coefficients de correction aux données nationales des forces de l'ordre, en standardisant sur les facteurs de sous-enregistrement. Il y a une analogie avec la standardisation indirecte. Les indicateurs de référence sont, au lieu des incidences, les coefficients de correction estimés sur le Rhône. Ils sont appliqués, non pas aux personnes-années de la population française étudiée, mais aux nombres de blessés des données nationales des forces de l'ordre. Enfin, nous stratifions non pas sur la structure d'âge et de sexe de la population étudiée, mais sur la structure de la population des blessés en termes de gravité lésionnelle, type d'usager, tiers (oui/non), force de l'ordre et type de réseau.

Avant d'appliquer la méthode de capture-recapture aux données de 1996-2004, quelques étapes sont nécessaires. La première est la prédiction du NISS sur les données des forces de l'ordre, puisque cette information n'y figure pas. Le modèle de prédiction du NISS est construit au niveau du département du Rhône, sur les blessés identifiés dans les deux sources pour lesquelles nous disposons du NISS (estimé dans le cadre du Registre) et des variables des BAAC, notamment du critère de gravité des BAAC.
Une seconde étape nécessaire à une application satisfaisante de la méthode de capture-recapture est l'amélioration du résultat du chaînage entre les données des forces de l'ordre et le Registre : il s'agit de mieux estimer le nombre de blessés communs aux deux sources (données des forces de l'ordre et Registre). À cette fin, nous estimons le nombre de faux négatifs et de faux positifs.
Les différentes étapes du l'estimation du bilan de l'insécurité routière sont schématisées dans la Figure 10.

Figure 10 : ensemble de la procédure d'estimation du bilan de morbidité par accidents de la circulation routière, France entière



Ce chapitre suit la structure IMMRID (Introduction, Matériel, Méthode, Résultats, Interprétation et Discussion) ; la partie Méthode est elle-même décomposée en quatre étapes (cf. Figure 10)
Étape 1 : prédiction du NISS chez les blessés recensés par les forces de l'ordre
Étape 2 : amélioration du chaînage
Étape 3 : modélisation du sous-enregistrement par approche capture-recapture
Étape 4 : application des coefficients de correction aux données nationales des forces de l'ordre

Les présentations des étapes 1 et 3 sont elles-mêmes structurées selon le modèle IMMRID.

Matériel

Nous travaillons sur les données présentées auparavant:
- les données des forces de l'ordre (BAAC), au niveau du Rhône et au niveau national,
- le Registre du Rhône,
- les données communes aux deux sources, au niveau du Rhône.

Selon l'étape de la procédure d'estimation, nous utilisons certaines de ces données et pas d'autres ; cela sera spécifié.

Pour l'ensemble de la procédure, nous excluons les blessés du Registre ne satisfaisant pas aux critères de définition d'un accident de la circulation routière des BAAC : rollers, utilisateurs de skateboard ou de trottinette se blessant seuls ou contre un piéton, et les blessés pour lesquels la localisation de l'accident est clairement spécifiée comme « hors réseau ». Nous n'avions pas appliqué ce critère jusqu'à présent car il n'est pas systématiquement appliqué par les forces de l'ordre et le but était alors de comparer les données des forces de l'ordre et le Registre. Nous l'appliquons ici car il s'agit maintenant d'estimer un bilan de la morbidité routière à l'échelle nationale pour lequel la définition d'accident de la circulation routière utilisée ne doit pas porter à polémique.

La projection porte sur la France métropolitaine. Elle concerne la période 1996-2004, c'est-à-dire depuis que le Registre du Rhône est exploitable et jusqu'au changement en 2005 de la définition des blessés et tués dans les BAAC.

Méthode

États de santé étudiés

Nous estimons les effectifs de blessés pour les critères de santé suivants : les blessés toutes gravités, les blessés graves, les blessés hospitalisés, les blessés avec séquelles majeures. Ces critères sont basés sur des classifications médicales (à l'exception de la notion d'hospitalisation). Nous utilisons le NISS pour mesurer la gravité lésionnelle, et en trois catégories : NISS 1-3 (gravité mineure), NISS 4-8 (gravité modérée), et NISS 9+ (gravité sérieuse et au-delà, sévère et critique). Nous regroupons les trois catégories pour définir les blessés toutes gravités : NISS 1+. Nous utilisons la catégorie NISS 9+ pour définir les blessés graves.

Afin de définir les blessés avec séquelles majeures, nous utilisons l'échelle Injury Impairment Scale (AAAM 1994) (IIS) et nous considérons les personnes avec lésions IIS 3 et plus. La classification IIS attribue pour chaque blessure répertoriée par l'Abbreviated Injury Scale un code de 0 à 6 mesurant les déficiences prévisibles un an après l'accident. Elle prend en compte la mobilité, les aspects esthétique, cognitif, sensoriel, sexuel/reproductif et la douleur. Les valeurs de l'IIS ont été attribuées par consensus entre 35 spécialistes. Cette échelle n'a pas été validée par une comparaison entre les séquelles prédites et les séquelles observées, mais les scores à partir de trois correspondent à des lésions avec des dégâts anatomiques importants : lésions avec perte de substance nerveuse (encéphale, moelle épinière, racines nerveuses), ou perte d'un membre ou destruction de grosses articulations. Il n'y a donc guère d'incertitude sur les séquelles prévues à un an pour ce type de lésions.
Les régions corporelles essentiellement concernées par des lésions IIS 3+ chez les victimes de la circulation routière (Gadegbeku, Ndiaye et al. 2006) sont la tête, la colonne vertébrale et les membres inférieurs. Les blessures IIS 3+ à la tête sont essentiellement des hématomes intracérébraux (IIS 3), des hémorragies intraventriculaires (IIS 3), des hématomes sous-duraux (IIS 4) responsables de séquelles de toutes sortes : motrices, sensorielles, comportementales, mnésiques et/ou intellectuelles. Les blessures IIS 3+ à la colonne vertébrale sont des atteintes au plexus brachial (IIS 3) entraînant une paralysie d'un membre supérieur, des atteintes à la moelle épinière (contusions, plaies), entraînant des paraplégies (IIS 4) et tétraplégies (IIS 5). Les blessures IIS 3+ des membres inférieurs sont

essentiellement des luxations du genou, de la hanche (IIS 3), des fracas articulaires et des amputations (IIS 3 ou 4) donnant lieu à des handicaps orthopédiques.
Nous nous attachons à estimer la fréquence (effectif et incidence) pour l'ensemble des personnes avec lésions IIS 3+, et selon les trois régions corporelles les plus fréquemment concernées.
Nous ne présentons pas d'estimations annuelles mais seulement une estimation moyenne sur la période 1996-2004. En effet, au niveau du département du Rhône les effectifs des victimes avec séquelles majeures sont très petits et leur évolution nous semble trop erratique pour être projetée sur la France entière.

Un blessé hospitalisé est défini comme étant admis comme patient dans un hôpital. Nous nous intéressons au critère hospitalisé pour au moins deux raisons. La première tient au fait qu'à partir de 2005, le critère de gravité dans les BAAC est basé sur l'hospitalisation (oui/non). Les blessés graves correspondent dorénavant aux blessés hospitalisés ; les blessés légers correspondent aux blessés non hospitalisés, ayant fait l'objet de soins médicaux.
La deuxième raison de ce choix est de faciliter les comparaisons internationales, et c'est justement pour cela que ce critère a été choisi comme nouveau critère de gravité dans les données des forces de l'ordre. C'est un critère relativement répandu dans les autres pays occidentaux, et donc utile même s'il a un certain nombre de défauts, dont le fait de dépendre des politiques hospitalières.

En résumé, les états de santé que nous étudions sont :
- les blessés toutes gravités (NISS 1+),
- les blessés graves (NISS 9+),
- les blessés hospitalisés,
- les victimes avec séquelles majeures prévues (au moins une lésion IIS 3+),
- les victimes avec séquelles majeures prévues, lésion IIS3+ à la tête: traumatisés crâniens graves,
- les victimes avec séquelles majeures prévues, lésion IIS3+ à la colonne : traumatisés médullaires,
- les victimes avec séquelles majeures prévues, lésion IIS3+ aux membres inférieurs.

L'ensemble de la procédure de projection est schématisée en Figure 10. Elle est présentée pour l'estimation d'effectifs en fonction du NISS. Toute la procédure est répétée pour les autres critères de santé étudiés : victimes hospitalisées, victimes avec lésions IIS3+, toutes régions corporelles, et victimes avec lésions IIS3+ à la tête, à la colonne ou aux membres inférieurs, séparément. Les détails de la procédure sur ces autres critères (modèles obtenus, paramètres estimés…) sont fournis en annexe.

La classification des victimes selon ces critères est quasi-immédiatement disponible pour les observations du Registre. En ce qui concerne l'information sur l'hospitalisation, elle est fournie par une variable spécifique, ou lorsque celle-ci est manquante, elle est imputée en fonction du type de services hospitaliers visités ou du temps écoulé entre les dates et heures d'entrée et de sortie.

Ces informations n'existent pas pour les blessés enregistrés dans les BAAC. La première étape de la procédure de projection (Figure 10) consiste donc à prédire l'état de santé des blessés BAAC en fonction de ces critères.

Étape 1 : Construction et application d'un modèle de prédiction du NISS

Nous cherchons à prédire le NISS (en trois catégories) pour les blessés recensés seulement par les forces de l'ordre. Il ne s'agit pas de prédire le niveau de NISS à l'échelle individuelle de chaque victime, mais à un niveau agrégé, par groupes définis par diverses caractéristiques du blessé et/ou de l'accident qui sont disponibles dans les BAAC. Le niveau d'agrégation est défini par les variables utilisées dans le modèle de sous-enregistrement (étape 3) et par toute autre variable d'intérêt telle que âge ou sexe.

La principale variable de prédiction du NISS est le critère BAAC blessé léger/ blessé grave. En effet, même si le classement des forces de l'ordre n'est pas satisfaisant, il est très corrélé au NISS (cf. chap. 2).

Matériel

La construction de ce modèle de prédiction repose sur les blessés identifiés comme communs aux BAAC et au Registre, pour lesquels on dispose des deux critères de gravité : celui des BAAC et le NISS. Cela correspond à 22 969 blessés sur la période 1996-2004.

Les observations où le NISS est manquant sont exclues (étant donné qu'il s'agit de la variable à prédire). Elles représentent 1,15 % des observations. Quelques blessés ont des valeurs manquantes pour la variable âge. Pour y remédier, les valeurs âge sont imputées en fonction de type d'usager et sexe (par imputation simple).
22 704 observations sont finalement disponibles pour la construction du modèle de prédiction du niveau de NISS.

Méthode

Modélisation

La variable à prédire est le NISS en trois catégories : NISS 1-3, NISS 4-8 et NISS 9+ ; c'est une variable catégorielle ordinale. Nous la modélisons par un modèle multinomial, de logit cumulé. Il s'écrit de la façon suivante :

$$\text{Logit } Pg = \ln P(NISS \geq g) / P(NISS < g), \text{ pour } g=4, 9 \text{ (valeurs seuils)}$$
$$= \beta_{g0} + \sum_k \beta_{gk} X_{gk} \quad \text{où } X_k \text{ sont les variables prédictives}$$

L'hypothèse des odds (ratios) proportionnels pose que les $ß_{gk}$ soient indépendants de g, la valeur seuil, autrement dit que ce soient des constantes $ß_k$.

Nous utilisons la procédure LOGISTIC du logiciel SAS.

Choix des variables du modèle

Les variables à considérer pour l'inclusion dans le modèle sont les variables qui permettent d'améliorer la prédiction de la gravité des blessures d'une victime une fois que le critère gravité des BAAC est pris en compte : caractéristiques du blessé, de l'accident, de l'environnement au sens large. Il ne peut s'agir que de variables présentes dans les BAAC puisque le but sera d'appliquer le modèle aux données BAAC nationales. Les caractéristiques individuelles considérées sont l'âge, le sexe, le type d'usager, le port d'un dispositif de sécurité (ceinture, casque, autre), et l'alcoolémie. Les caractéristiques de l'accident considérées sont le type d'antagoniste, l'obstacle heurté, la vitesse lors du choc (non disponible, seules le sont les infractions pour excès de vitesse). Les caractéristiques générales considérées sont le type de réseau, en ou hors agglomération, en ou hors intersection, la luminosité (jour / nuit), le type de jour (début de semaine / fin de semaine), les conditions atmosphériques (normales / mauvaises), le type de force de l'ordre déclarant l'accident, et l'année de survenue de l'accident. Le choix des variables considérées repose sur les facteurs de risque établis de gravité lésionnelle (Robertson 1992; Evans 2004).

Certaines variables n'ont finalement pas été retenues car il est apparu que leur taux de remplissage était très variable d'une région à l'autre, et ceci même pour un type de force de l'ordre donné. Par exemple, parmi les données provenant de la force de l'ordre « police », concernant les automobilistes, la proportion de valeurs manquantes varie, pour le port de la ceinture : entre 10 % (Auvergne, Poitou-Charentes) et 34 % (Provence-Alpes Côte d'azur, Aquitaine, Picardie) ; pour le contrôle de l'alcoolémie : entre 4 % (Auvergne, Nord-Pas de Calais) et 47 % (Aquitaine) ; et pour un obstacle fixe heurté, variable non « sensible » : entre 71 % (Limousin) à 86 % (PACA, Bourgogne) alors qu'il existe une catégorie « sans obstacle » pour cette variable. Nous ne retenons donc pas les variables :

port du casque, port de la ceinture, alcoolémie, infractions pour excès de vitesse, et obstacle fixe heurté.

Nous n'avons gardé pour éventuelle inclusion dans le modèle que les variables renseignées à 100 % : type d'usager, type de réseau, luminosité, intersection (oui/non), agglomération, type d'antagoniste, type de force de l'ordre, année. Le type d'antagoniste est le type d'usager de l'antagoniste ; il est construit à partir de la catégorie du véhicule antagoniste, ou du plus lourd si il y en a plusieurs, et il inclut une catégorie « piéton » et une catégorie « sans antagoniste ».

La plupart des variables considérées sont apparues significativement liées à la gravité NISS (en analyses bivariées, ajustées sur le critère gravité des BAAC) : type d'usager, âge, sexe, type d'antagoniste (ou absence), type de réseau, en / hors agglomération, conditions atmosphériques, en / hors intersection. Les variables non significatives en analyse bivariée sont le type de jour (début / fin de semaine), p=0,3 et p=0,5, et les conditions lumineuses (jour / nuit) p=0,13 et p=0,03). En analyse multivariée, deux variables deviennent non significatives : les conditions atmosphériques (p=0,08 et p=0,61), et la variable en / hors intersection (p=0,13 et p=0,44).

Afin de prendre en compte l'évolution des pratiques de classement de la gravité par les forces de l'ordre (cf. chap. 2), nous avons considéré l'inclusion de la variable année sous forme quantitative. Nous avons écarté son inclusion sous forme qualitative, par crainte de sur-ajustement du modèle aux éventuelles particularités des pratiques des forces de l'ordre dans le département du Rhône (dont une identifiée : arrêt entre 1996 et 1997 de l'enregistrement des accidents et blessés par une seule brigade spécialisée pour l'agglomération lyonnaise et transfert de cette activité sur l'ensemble des commissariats de l'agglomération, ce qui s'est traduit par une baisse de 40 % de la proportion des blessés graves, baisse que l'on ne retrouve pas dans le Registre).
L'étude antérieure de comparaison des deux classements de la gravité (NISS et critère des BAAC) ayant montré un fort effet du type de force de l'ordre sur le classement de la gravité, nous avons conduit des analyses séparées par force de l'ordre afin d'explorer les éventuelles interactions de la variable force de l'ordre avec les autres variables. Ces analyses séparées ont montré une très forte variation de l'effet du classement gravité BAAC sur la prédiction du NISS selon la force de l'ordre : l'odds ratio (d'être NISS 9+ ou NISS 4+) associé au classement « blessé grave » vs « blessé léger » varie entre 10,5 et 29,5 selon la force de l'ordre. En conséquence, le modèle global retenu inclut l'interaction gravité BAAC et force de l'ordre (p<0,0001).

En résumé, le modèle de prédiction contient les variables suivantes : interaction entre gravité BAAC et force de l'ordre, type d'usager, sexe, âge, type d'antagoniste, type de réseau et en / hors agglomération. Le type d'usager est groupé en cinq catégories : automobilistes / usagers de 2RM / cyclistes / piétons / autres ; le type d'antagoniste est groupé en six catégories : aucun / voitures et autres / piétons et cyclistes / deux-roues motorisés / véhicules utilitaires / véhicules lourds (poids lourd, car, bus, tram, tracteur, engin). L'âge est en neuf catégories : 0-4 ans / 5-9 / 10-14 / 15-19 / 20-29 / 30-54 / 55-64 / 65-74 / 75 ans + (elles ont été choisies en partie de façon à ce que les taux de gravité soient homogènes dans chacune). Le réseau est en cinq catégories : autoroutes (AR) / routes nationales (RN) / routes départementales (RD) / voies communales / autres.

Sur les données des blessés du Rhône, l'hypothèse des odds proportionnels est rejetée (p<0,05). Cette hypothèse est testée avec la procédure LOGISTIC, dès lors qu'un modèle de logit cumulé est spécifié. Le modèle multinomial des logits cumulés correspond alors à deux modèles logit binomiaux :

$$\text{logit}\, P(\text{NISS } 4+ / X) = \ln \frac{P(\text{NISS} \geq 4/X)}{P(\text{NISS} < 4/X)} = \sum_k \beta_{4k} X_k$$

$$\text{logit}\, P(\text{NISS } 9+ / X) = \ln \frac{P(\text{NISS} \geq 9)}{P(\text{NISS} < 9)} = \sum_k \beta_{9k} X_k$$

Les variables explicatives X_k sont toutes des variables catégorielles, sauf l'année qui est sous forme quantitative. X_0 vaut 1 afin que $ß_0$ corresponde à l'ordonnée à l'origine.

Prédictions

Le modèle de prédiction du NISS est appliqué aux données des forces de l'ordre. Nous rappelons que le but est d'obtenir une meilleure estimation de la gravité de ces blessés que celle fournie par le classement des forces de l'ordre.

Le modèle de prédiction du NISS est dans un premier temps appliqué aux données des forces de l'ordre du Rhône, « BAAC seulement », afin de préparer l'étape 3 de capture-recapture (Figure 10). 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 NISS parmi les variables caractérisant les sous-groupes de probabilité homogène d'enregistrement des blessés.

Le modèle de prédiction du NISS est appliqué aux données nationales des forces de l'ordre, afin de préparer l'étape n° 4 de projection (Figure 10). Lors de la projection proprement dite, les coefficients de correction du sous-enregistrement, qui sont notamment définis selon les trois niveaux du NISS, pourront alors être appliqués aux données nationales des forces de l'ordre.

Les probabilités prédites de NISS 1-3, 4-8 et 9+, pour les différents profils (que l'on note j, définis par les valeurs X_{kj}), sont obtenues par les formules suivantes :

$$P(\text{NISS } 1-3 / X_j) = \frac{\exp(-\sum_k \beta_{4kj} X_{kj})}{1+\exp(-\sum_k \beta_{4kj} X_{kj})}$$

$$P(\text{NISS } 4+ / X_j) = \frac{1}{1+\exp(-\sum_k \beta_{4kj} X_{kj})}$$

$$P(\text{NISS } 9+ / X_j) = \frac{1}{1+\exp(-\sum_k \beta_{9kj} X_{kj})}$$

$$\text{et } P(\text{NISS } 4\text{-}8/X_j) = P(\text{NISS } 4+ / X_j) - P(\text{NISS } 9+ / X_j).$$

Les profils sont définis par la combinaison des variables prédictives 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 NISS 1-3, NISS 4-8, et NISS 9+ pour ce profil.

Résultats

Nous présentons ici les modèles de prédiction construits, et quelques exemples de profils de blessés et les probabilités prédites de gravité NISS associées. Les modèles de prédiction obtenus seront ensuite discutés.

Les paramètres des modèles logistiques sont donnés en Tableau 34 et en Tableau 35.

Tableau 34 : paramètres du modèle de prédiction P(NISS4+), construit sur 22 704 blessés de l'intersection BAAC-Registre, Rhône, 1996-2004

variable	beta	IC à 95 %	
ordonnée à l'origine	-1,17	-1,32	-1,02
force ordre × gravité (p<0,0001)			
gendarmerie, blessé léger	0,00		
gendarmerie, blessé grave	2,42	2,25	2,60
police, blessé léger	-0,77	-0,89	-0,65
police, blessé grave	2,08	1,89	2,27
CRS, blessé léger	-1,13	-1,31	-0,96
CRS, blessé grave	2,10	1,69	2,52
sexe (p<0,0001)			
féminin	0,00		
masculin	0,31	0,24	0,38
âge (p<0,0001)			
0-4 ans	-0,45	-0,74	-0,17
5-9 ans	-0,31	-0,51	-0,11
10-14 ans	-0,09	-0,26	0,07
15-19 ans	-0,14	-0,24	-0,03
20-29 ans	-0,16	-0,24	-0,08
30-54 ans	0,00		
55-64 ans	0,26	0,12	0,40
65-74 ans	0,44	0,28	0,61
75 ans et plus	0,52	0,34	0,71
type d'usager × tiers (p<0,0001)			
piétons, avec tiers	1,28	1,17	1,38
cyclistes, avec tiers	1,06	0,89	1,23
cyclistes, sans tiers	0,70	-0,18	1,57
usagers de 2RM, avec tiers	1,15	1,05	1,25
usagers de 2RM, sans tiers	1,71	1,49	1,93
automobilistes, avec tiers	0,00		
automobilistes, sans tiers	0,74	0,64	0,85
autres, avec tiers	-0,15	-0,36	0,07
autres, sans tiers	0,69	0,37	1,00
type d'antagoniste (p<0,0001)		,	,
piétons ou cyclistes	-0,24	-0,58	0,10
deux-roues motorisés	-0,04	-0,27	0,18
voitures, autres	0,00		
véhicules utilitaires	0,08	-0,07	0,24
véhicules lourds	0,35	0,22	0,48
réseau (p<0,001)			
autoroutes	0,42	0,22	0,62
routes nationales	0,13	0,08	0,38
routes départementales	0,23	0,03	0,24
voies communales	0,00		
autres	-0,01	-0,21	0,19
agglomération (p=0,086)			
oui	0,00		
non	0,10	-0,01	0,22
année (p=0,002)			
(quantitative)	0,02	0,01	0,04

n=22 704, nombre de blessés NISS 4+=8892, coefficient D de Somer=0,588

Tableau 35 : paramètres du modèle de prédiction P(NISS 9+), construit sur 22 704 blessés de l'intersection BAAC-Registre, Rhône, 1996-2004

variable	beta	IC à 95 %	
ordonnée à l'origine	-2,97	-3,17	-2,77
force ordre × gravité (p<0,0001)			
gendarmerie, blessé léger	0,00		
gendarmerie, blessé grave	2,41	2,26	2,55
police, blessé léger	-0,56	-0,72	-0,40
police, blessé grave	2,15	1,97	2,33
CRS, blessé léger	-1,17	-1,42	-0,92
CRS, blessé grave	2,35	2,01	2,69
sexe (p<0,0001)			
féminin	0,00		
masculin	0,31	0,21	0,40
âge (p<0,0001)			
0-4 ans	-0,30	-0,68	0,08
5-9 ans	-0,13	-0,38	0,13
10-14 ans	-0,17	-0,39	0,04
15-19 ans	-0,11	-0,24	0,02
20-29 ans	-0,16	-0,27	-0,05
30-54 ans	0,00		
55-64 ans	0,16	-0,03	0,34
65-74 ans	0,39	0,19	0,59
75 ans et plus	0,47	0,25	0,69
type d'usager × tiers (p<0,0001)			
piétons, avec tiers	1,29	1,15	1,42
cyclistes, avec tiers	0,93	0,70	1,15
cyclistes, sans tiers	0,88	-0,00	1,77
usagers de 2RM, avec tiers	1,30	1,17	1,42
usagers de 2RM, sans tiers	1,61	1,37	1,85
automobilistes, avec tiers	0,00		
automobilistes, sans tiers	0,79	0,65	0,93
autres, avec tiers	-0,03	-0,35	0,30
autres, sans tiers	0,83	0,39	1,27
type d'antagoniste (p<0,0001)			
piétons ou cyclistes	-0,38	-0,85	0,10
deux-roues motorisés	0,13	-0,14	0,39
voitures, autres	0,00		
véhicules utilitaires	0,04	-0,18	0,25
véhicules lourds	0,39	0,23	0,56
réseau (p=0,048)			
autoroutes	0,29	0,01	0,57
routes nationales	0,21	0,03	0,39
routes départementales	0,18	0,05	0,31
voies communales	0,00		
autres	0,07	-0,17	0,32
agglomération (p<0,0001)			
oui	0,00		
non	0,31	0,17	0,45
année (p<0,0001)			
(quantitative)	0,04	0,02	0,06

N=22 704, nombre de blessés NISS 9+ =4191, coefficient D de Somer=0,707

Être classé « blessé grave » par les forces de l'ordre est associé à un paramètre beta nettement positif dans les modèles logistiques de la probabilité de la gravité au sens NISS.
Dans les deux modèles, les paramètres beta associés aux hommes sont positifs, avec les femmes en catégorie de référence. Concernant l'âge, avec les 30-54 ans comme catégorie de référence, les paramètres associés aux personnes au delà de 55 ans sont positifs, avec semble-t-il un gradient croissant en fonction de l'âge.
Pour le type d'usager, les occupants de véhicule motorisé ont des paramètres beta plus élevés dans les accidents sans tiers qu'avec tiers ; c'est l'inverse pour les cyclistes. Parmi les accidents avec tiers, par rapport aux automobilistes (référence), seuls les piétons, cyclistes, usagers de 2RM ont des paramètres positivement liés à la probabilité de gravité. Concernant le type d'antagoniste, avoir comme antagoniste un véhicule lourd est positivement lié à la probabilité de gravité dans les deux modèles (référence= voiture).
Concernant le type de réseau, les autoroutes, routes nationales et routes départementales sont associée à des paramètres bêta positifs (avec les voies communales comme catégorie de référence).
Enfin, l'année sous forme quantitative est associée à un paramètre positif, mais très faible, dans les deux modèles.

Le coefficient D de Somer est égal à 0,59 pour le modèle de prédiction du NISS 4+ et à 0,71 pour le modèle de prédiction du NISS 9+.

Le Tableau 36 présente les probabilités des trois niveaux de NISS pour quelques profils de blessés. Nous présentons des profils où seuls les trois plus grands facteurs de prédiction de la gravité lésionnelle varient : le type d'usager, le classement des forces de l'ordre en blessé « léger » ou « grave », et le type de force de l'ordre établissant le classement.

Tableau 36 : prédiction du niveau NISS en fonction des profils de blessés des BAAC
(profil : sexe masculin, âge 20-29 ans, accident contre voiture, en agglomération, sur voie communale, en 2004)

gravité	force de l'ordre	type d'usager	P(NISS 1-3)	P(NISS 4-8)	P(NISS 9+)
blessé grave	gendarmerie	usager de 2RM	0,06	0,17	0,77
blessé grave	gendarmerie	cycliste	0,07	0,24	0,69
blessé grave	gendarmerie	automobiliste	0,17	0,36	0,47
blessé grave	gendarmerie	piéton	0,05	0,18	0,77
blessé léger	gendarmerie	usager de 2RM	0,42	0,35	0,23
blessé léger	gendarmerie	cycliste	0,44	0,39	0,17
blessé léger	gendarmerie	automobiliste	0,69	0,23	0,07
blessé léger	gendarmerie	piéton	0,39	0,39	0,23
blessé léger	gendarmerie	autre	0,72	0,20	0,07
blessé grave	police	usager de 2RM	0,08	0,20	0,72
blessé grave	police	cycliste	0,09	0,27	0,64
blessé grave	police	automobiliste	0,22	0,37	0,41
blessé grave	police	piéton	0,07	0,21	0,72
blessé grave	police	autre	0,25	0,35	0,40
blessé léger	police	usager de 2RM	0,61	0,25	0,14
blessé léger	police	cycliste	0,63	0,27	0,10
blessé léger	police	automobiliste	0,83	0,13	0,04
blessé léger	police	piéton	0,58	0,28	0,14
blessé léger	police	autre	0,85	0,11	0,04

Discussion

Les valeurs des paramètres des modèles de prédiction sont conformes à ce que l'on sait par ailleurs sur les facteurs de risque de gravité lésionnelle. Les hommes ont en moyenne des accidents plus graves que les femmes, en raison de prise de risques plus élevées (Martin, Lafont et al. 2004); les personnes âgées sont plus gravement blessées, car plus fragiles ; les piétons, cyclistes et usagers de 2RM sont

plus gravement blessés que les automobilistes : ce sont les usagers vulnérables non-protégés par une carrosserie.
Il est aussi connu que, pour les véhicules motorisés, les accidents sans tiers sont plus graves que les accidents avec tiers. L'explication généralement avancée est qu'ils sont en général le résultat de pertes de contrôle du véhicule à vitesse élevée, ou de heurts avec obstacles peu déformables comme des poteaux, des arbres ou des murs (Martin, Derrien et al. 2003). Concernant les cyclistes, il est connu que les accidents sans tiers sont moins graves que les accidents avec tiers : les accidents sans tiers sont en grande partie des chutes de vélo, à vitesse faible ou modérée, alors que les accidents avec tiers sont essentiellement des accidents contre voiture (Laumon 2002).
Il est évident qu'avoir comme antagoniste un véhicule lourd est un facteur de risque de gravité puisque l'énergie dissipée lors du choc est proportionnelle à la masse, et que la gravité des blessures est liée à la quantité d'énergie dissipée lors du choc.
Les autoroutes, routes nationales, et routes départementales sont effectivement des facteurs de risque de gravité par rapport aux routes communales, parce que la vitesse pratiquée y est plus élevée et que l'énergie du choc est proportionnelle au carré de la vitesse. Pour la même raison, les accidents hors agglomération sont plus graves en moyenne que les accidents en agglomération.
Le coefficient positif associé à l'année pourrait signifier une hausse de la gravité des accidents au cours de la période 1996-2004. D'une part, ce n'est pas très plausible car les vitesses moyennes diminuent (ONISR 2005), suite à l'extension des contrôles automatisés de vitesse ; d'autre part, il s'agit de la prédiction de la gravité lésionnelle à partir des variables des BAAC et notamment du classement de la gravité par les forces de l'ordre. Or nous avons vu (cf. chap. 2) que les forces de l'ordre sous-estiment de plus en plus la gravité, à NISS constant (au moins sur la période 1997-2001). Ce paramètre positif correspond donc vraisemblablement à ce phénomène. Le fait d'avoir ce paramètre dans le modèle permet de compenser le phénomène de sous-classement, afin de prédire au mieux le NISS.

Les coefficients D de Somer sont satisfaisants car ils sont proches de 1. Nous rappelons que le coefficient D de Somer est une mesure d'association asymétrique, ici entre les valeurs observées et prédites, et qu'il évolue entre les bornes -1 et 1. L'adéquation du modèle aux données (test de Hosmer-Lemeshow) est détaillée en annexe. Globalement, l'écart relatif entre effectifs prédits et effectifs observés, par décile de la probabilité prédite, P(NISS 9+) ou P(NISS 4+), est relativement petit, indiquant une adéquation tout à fait acceptable.

Nous avons vérifié si les prédictions nous donnaient des valeurs plausibles. Pour cela, nous avons comparé les proportions de blessés légers en fonction des régions. Les proportions de blessés légers les plus fortes se trouvent en Ile de France (62,3 %), en PACA (60,2 %) et les plus basses en Champagne-Ardenne (47,4 %) et en Picardie (46,2 %). Cela correspond bien à des taux de gravité plus élevés en zone rurale qu'en zone urbaine, et cela conforte donc les prédictions.

Nous avons comparé les NISS prédits avec les NISS observés sur un sous-ensemble de blessés n'ayant pas servi à la construction des modèles de prédiction. Les blessés identifiés dans les BAAC seulement (Rhône, 2001, cf. chap. 1) nous fournissent un sous-ensemble de ce type. Les NISS observés sont ceux évalués à partir des certificats médicaux inclus dans les procès-verbaux. Nous comparons donc la distribution des NISS prédits par le modèle de prédiction à celle des NISS évalués à partir des certificats médicaux.

Tableau 37 : comparaison des NISS prédits et observés, parmi les blessés identifiés dans les BAAC seulement, Rhône, 2001

	NISS observé n=528	NISS prédit n= 528
NISS 1-3	69,5 %	64,9 %
NISS 4-8	21,8 %	21,1 %
NISS 9+	8,7 %	14,0 %
ensemble	100,0 %	100,0 %

Le test du chi-deux de comparaison d'une distribution observée à une distribution théorique rejette l'hypothèse nulle ($p<0,01$). La distribution prédite donne une plus grande proportion de blessés graves que la distribution du NISS basée sur les certificats médicaux. La distribution NISS prédite vient du modèle de prédiction construit sur le NISS fourni par le Registre. Dans le Registre, la gravité lésionnelle est consolidée, c'est-à-dire qu'elle est basée sur l'ensemble des diagnostics établis par les différents services que le blessé a consulté. En revanche, dans les procès-verbaux, il y a tout au plus un seul certificat médical ; cela peut éventuellement conduire à une sous-estimation de la gravité. Cela peut donc en partie expliquer la moindre proportion de blessés graves selon le NISS observé que selon le NISS prédit.

Par ailleurs, le modèle de prédiction est construit sur les blessés identifiés comme communs aux données des forces de l'ordre et au Registre. Ces blessés ont vraisemblablement une gravité lésionnelle un peu plus grande que les blessés identifiés seulement dans les BAAC. Cela correspond à un biais de sélection attendu (du Registre). C'est aussi un des résultats de l'étude auprès des blessés identifiés seulement dans les BAAC (cf. chap. 1). Celle-ci a cependant porté sur un sous-groupe de 528 blessés : ceux avec procès-verbal exploitable et avec certificat médical inclus (parmi 1322 blessés), dont la représentativité n'est pas assurée.

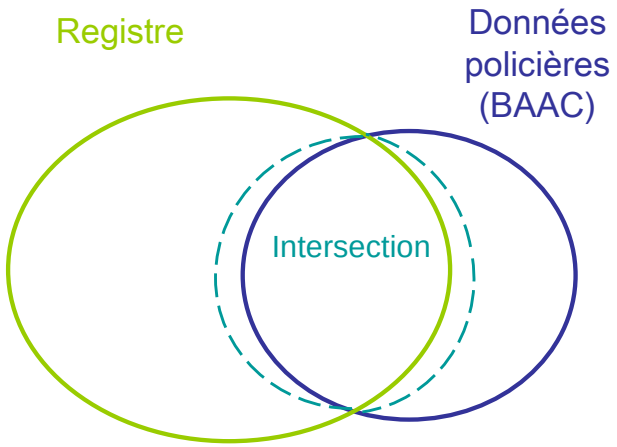
Le modèle de prédiction du NISS est fonction de caractéristiques de l'accident ; il est basé sur l'hypothèse, qu'à configuration égale d'accident, les conséquences moyennes en terme de gravité lésionnelle (sécurité secondaire) sont les mêmes, quelque soit le lieu géographique sur le territoire français. C'est une hypothèse tout à fait plausible. Le point faible réside éventuellement dans la caractérisation de la « configuration de l'accident ». Dans le modèle de prédiction, la configuration de l'accident est mesurée par des caractéristiques de l'accident et des personnes impliquées, à partir des informations des BAAC. Le modèle inclut notamment l'âge et le sexe du blessé, l'interaction type d'usager et tiers, le type d'antagoniste, le type de réseau, et le fait que l'accident ait lieu en agglomération ou en dehors. La gravité lésionnelle d'un blessé dépend aussi du port des dispositifs de sécurité (ceinture, casque), et dans une moindre mesure de l'alcoolémie du blessé ; ces variables sont disponibles dans les BAAC mais nous les avons écarté du modèle de prédiction car leur comparabilité sur le territoire français n'est pas assurée (taux de remplissage lié à la région et au département). Toutefois, le modèle de prédiction inclut aussi le critère de gravité des forces de l'ordre, et nous pouvons penser que les forces de l'ordre classent éventuellement les blessés en « léger » ou « grave » en fonction du port effectif (ou non) des dispositifs de sécurité (gravité attendue), et non en fonction de ce qu'elles notent dans le procès-verbal à ce sujet.
Cela renvoie à une hypothèse déjà mentionnée : le classement de la gravité par les forces de l'ordre est supposé indépendant de la région ou du département, en d'autres termes homogène à travers le territoire français.

Étape 2 : Amélioration du résultat du chaînage

Avant de modéliser le sous-enregistrement, nous avons besoin d'une estimation aussi fiable que possible de la répartition des blessés selon leur source d'enregistrement : « BAAC seulement », « Registre seulement », ou « intersection BAAC-Registre ». Dans ce but, nous corrigeons le résultat du chaînage grâce à l'estimation du nombre de faux positifs et de faux négatifs. 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).
L'amélioration du résultat du chaînage par l'estimation du nombre de faux positifs et de faux négatifs correspond au scénario n° 3 du chapitre 3 de l'approche capture-recapture appliquée aux données de 2001 du Rhône. Le scénario n° 1 était le résultat du chaînage standard effectué régulièrement. Le scénario n° 2 était le résultat du chaînage supplémentaire grâce au recueil des noms et prénoms (par consultation des PV papier) ; il n'était pas envisageable de le reconduire pour les autres années. Le scénario n° 3 est le résultat d'estimations sur la qualité du chaînage : il prend en compte le nombre estimé de faux négatifs et de faux positifs pour obtenir une meilleure estimation du nombre de vrais

positifs et de vrais négatifs. Il est aisément reproductible sur les années autres que 2001. L'estimation des effectifs des faux positifs et des faux négatifs repose sur des calculs de probabilité (Brenner 1994; Brenner and Schmidtmann 1996) et de recherche de valeurs approchées. Il s'agit d'une part de la probabilité de discordance sur les valeurs d'une (ou plusieurs) variable(s) de chaînage alors qu'on compare deux enregistrements de la même victime. Il s'agit d'autre part de la probabilité de concordance (par hasard) sur les valeurs d'une (ou plusieurs) variable(s) de chaînage alors qu'on compare des enregistrements de deux victimes distinctes. Le principe est exactement le même que celui explicité dans le chapitre 3 ; l'application est légèrement différente car il faut prendre en compte un changement dans la décision de rapprochement. En effet depuis 2002, le chaînage avec les données BAAC se fait lors de la saisie des fiches de notification des victimes dans le Registre. Le principe de chaînage est le même (mêmes variables de chaînage) mais la décision de chaîner est basée sur moins de tolérance de discordances dans les variables de chaînage (qui pourraient être dues à des erreurs, ou pas) ; cela vient du fait que la nouvelle stratégie d'enregistrement dans le Registre a pour but d'utiliser l'enregistrement BAAC comme source d'informations pour le Registre. Le détail des estimations des nombres de faux négatifs et de faux positifs est donné en annexe.

Figure 11 : redéfinition de la taille de l'intersection BAAC-Registre



Les effectifs estimés de faux négatifs et de faux positifs permettent de redéfinir les tailles de chaque sous-ensemble : « BAAC seulement », « Registre seulement » et « intersection BAAC-Registre ». L'intersection BAAC-Registre est augmentée de l'effectif faux négatifs - faux positifs et les sous-ensembles « BAAC seulement », « Registre seulement » sont chacun diminués de l'effectif faux positifs - faux négatifs.
Cela s'écrit (avec l'astérisque signifiant nouvel effectif) :

$$n^{*}_{\text{BAAC seul.}} = n_{\text{BAAC seul.}} - n_{\text{faux neg}} + n_{\text{faux pos}}$$

$$n^{*}_{\text{Reg. seul.}} = n_{\text{Reg. seul.}} - n_{\text{faux neg}} + n_{\text{faux pos}}$$

$$n^{*}_{\text{intersection}} = n_{\text{intersection}} + n_{\text{faux neg}} - n_{\text{faux pos}}$$

Un faux négatif est une paire constituée d'un individu de « BAAC seulement » et d'un individu de « Registre seulement ». 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, à savoir : le NISS (en trois niveaux), le type d'usager, le tiers (oui/non), le type de force de l'ordre et le type de réseau (principaux facteurs de biais de sélection).

Nous reprenons la notation du chapitre 3 en notant k les sous-ensembles exclusifs :
- 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, 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. 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 non 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 NISS, type d'usager, tiers, force de l'ordre et réseau) observée avant la réaffectation de ceux qui appartiennent à ce groupe mais n'avaient pas été identifiés comme tels, c'est-à-dire :

$$n^*_{kj} = p_{kj} \times n^*_k$$

Étape 3 : Modélisation du sous-enregistrement par capture-recapture

Nous rappelons que l'approche de capture-recapture (cf. chap. 3) est une formalisation du sous-enregistrement. Nous l'appliquons ici grâce à une modélisation explicite, par un modèle multinomial logistique. 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.

Matériel

L'approche porte sur les données des BAAC du Rhône, du Registre, avec identification des victimes communes, et après avoir amélioré le résultat du chaînage (par l'estimation des nombres de faux négatifs et de faux positifs, lors de l'étape n°2).

Méthode

Conditions

Nous rappelons les conditions d'application de la méthode (cf. chap. 3).

L'approche capture-recapture se base sur quatre conditions clés d'application (Hook and Regal 1995; IWGDMF 1995; IWGDMF 1995; Gallay, Nardone 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, Nardone 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.

Les conditions n° 1 à 4 ont été détaillées dans le chapitre 3. Nous rappelons brièvement à quel point elles sont vérifiées.

La condition n° 1 suppose une population fermée, c'est-à-dire qu'il n'y a pas d'entrée ou sortie entre les moments d'enregistrement dans les BAAC et dans le Registre. Cette condition est en grande partie vérifiée.

La condition n° 2 n'est pas vérifiée. Il y a une dépendance positive entre les deux sources d'enregistrement pour les blessés graves. Dans un tel cas, il a été établi (IWGDMF 1995) que l'estimateur du nombre total de sujets par capture-recapture sous-estime le nombre réel. Il faut donc considérer les résultats comme des valeurs planchers.

La condition n° 3 d'homogénéité de capture n'est vérifiée qu'à l'intérieur de sous-groupes ; ceux-ci sont définis par les facteurs de biais de sélection. Du coté de l'enregistrement policier, les facteurs de biais les plus importants sont la gravité des blessures, le type d'usager, la présence/absence du tiers, le type de réseau, et la force de l'ordre (cf. chap. 1). Les biais de sélection du Registre sont la gravité lésionnelle, la distance à l'hôpital (cf. chap.1) et éventuellement le type d'usager (éventuelle sous-notification des cyclistes (Sigal 2005)). La stratification ne permet pas de prendre en compte autant de variables. Nous utilisons donc la modélisation, et en particulier la modélisation logistique multinomiale (voir ci-dessous) qui permet facilement d'inclure les variables liées aux probabilités d'enregistrement.

La condition n° 4 suppose l'identification parfaite des victimes qui sont communes aux deux sources d'enregistrement. D'une part, le chaînage des victimes du Registre et des BAAC du Rhône est basé sur un certain nombre de variables informatives et discriminantes ; d'autre part, le résultat du chaînage est amélioré par l'estimation du nombre de faux positifs et de faux négatifs (cf. étape 2 de ce chapitre).

Nous détaillons les conditions 5 et 6 qui n'ont pas été prises en compte jusqu'ici.

La condition n° 5 signifie que les deux sources doivent identifier les cas pendant la même période de temps et sur la même zone géographique.
La période de temps est ici 1996-2004. 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 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 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.
Le lieu de l'accident est souvent non renseigné parmi les victimes du Registre. 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 flagrante et il est donc compréhensible que cette variable soit peu renseignée. Le personnel de l'ARVAC 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 (chap. 1), 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, 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. Nous choisissons alors de procéder 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 n° 6 signifie que les cas identifiés par chacune des sources doivent être de vrais 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 des blessés identifiés dans les BAAC seulement (chap. 1) a mis en évidence que certaines personnes classées « blessées » ne l'étaient pas : typiquement il s'agit de personnes envoyées dans un service hospitalier pour observation (femmes enceintes, nourrissons), 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. Elles sont cependant difficilement identifiables (cela impliquerait de retourner aux procès-verbaux papiers pour l'ensemble des blessés identifiés dans les BAAC). 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 1996-2004. 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 parait peu probable que d'autres types d'accident (professionnel ou de loisirs ou de la vie courante, exclusivement) ou d'autres circonstances de blessures (intentionnelles par exemple) soient classées comme accident de la circulation routière par les forces de l'ordre.

Du côté du Registre, les personnes incluses sont assurément blessées puisqu'un des critères d'inclusion du Registre est « 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 (chap.1) 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 1996-2004. Nous tirons au sort les personnes à exclure (année par année) parmi les blessés identifiés dans le Registre seulement.

Nous rappelons que 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 la projection. Pour la même raison, les victimes d'accidents clairement identifiés comme hors réseau routier sont exclues de la projection.

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 ci-dessous reprend une partie de leur article.

1) Sous l'hypothèse d'indépendance des deux sources

Le modèle s'écrit de la façon suivante :
Π_1 = probabilité d'être enregistré dans les BAAC
Π_2 = probabilité d'être enregistré dans le Registre
On note l_1 et l_2 les logit de ces probabilités
$\text{logit}(\Pi_1) = l_1$
$\text{logit}(\Pi_2) = l_2$

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

$$\Pi_0=(1-\Pi_1)\,(1-\Pi_2)=\frac{1}{(1+\exp(l_1))\times(1+\exp(l_2))}$$

Une variable indicatrice q_i est définie, qui 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 ($q_i=3$).
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(être dans BAAC seulement) / P(être observé)
p_2 = P(être dans Registre seulement) / P(être observé)
p_3 = P(être dans l'intersection BAAC - Registre) / P(être observé)

$$p_1=\frac{\Pi_1(1-\Pi_2)}{\Pi_1(1-\Pi_2)+\Pi_2(1-\Pi_1)+\Pi_1\Pi_2}=\frac{\exp(l_1)}{\exp(l_1)+\exp(l_2)+\exp(l_1+l_2)}$$

de la même façon :

$$p_2=\frac{\exp(l_2)}{\exp(l_1)+\exp(l_2)+\exp(l_1+l_2)}$$

$$p_3=\frac{\exp(l_1+l_2)}{\exp(l_1)+\exp(l_2)+\exp(l_1+l_2)}$$

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 :
$\text{logit}(p_k)=\log(p_k/p_1) = \beta_k$, k=2,3 (et $\beta_1=0$)

On ajuste le modèle, et on obtient une estimation des paramètres β_k.

Les paramètres β_k sont reliés à l_1 et l_2 de la façon suivante :

$$\beta_2=\log\left(\frac{p_2}{p_1}\right)=\log\left(\frac{\exp(l_2)/\left(\exp(l_1)+\exp(l_2)+\exp(l_1+l_2)\right)}{\exp(l_1)/\left(\exp(l_1)+\exp(l_2)+\exp(l_1+l_2)\right)}\right)=\log\left(\frac{\exp(l_2)}{\exp(l_1)}\right)=l_2-l_1$$

$$\beta_3 = \log\left(\frac{p_3}{p_1}\right) = \log\left(\frac{\exp(l_1 + l_2)}{\exp(l_1)}\right) = l_1 + l_2 - l_1 = l_2$$

et donc réciproquement :

$$l_2 = \beta_3$$

$$l_1 = \beta_3 - \beta_2$$

En remplaçant les paramètres l_1 et l_2 par leurs estimations, fonctions des estimation des paramètres $ß_k$, on estime Π_0 :

$$\hat{\Pi}_0 = \frac{1}{(1+\exp(\hat{l}_1)) \times (1+\exp(\hat{l}_2))}$$

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

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

On peut alors estimer le nombre total N :

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

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

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

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

La modélisation logit multinomiale nous donne aussi une estimation des probabilités Π_1 et Π_2 (non conditionnelles aux observations), et l'inverse de Π_1 est le coefficient de correction du sous-enregistrement des données des forces de l'ordre, facteur multiplicatif à appliquer aux effectifs des BAAC.

$$\hat{\Pi}_1 = \frac{1}{1+\exp(-\hat{l}_1)} = \frac{1}{1+\exp(\hat{\beta}_2 - \hat{\beta}_3)} \text{ et}$$

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

$\hat{C}_1 = \frac{1}{\hat{\Pi}_1} = 1+\exp(\hat{\beta}_2 - \hat{\beta}_3)$, coefficient de correction du sous-enregistrement des données des forces de l'ordre

2) Sous l'hypothèse de dépendance des deux sources

Nous considérons maintenant la situation 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.

Toujours en notant i l'individu, les probabilités d'enregistrement dans les sources BAAC et Registre sont notées Π_{1i} et Π_{2i} et sont reliées à la caractéristique X (quantitative) de la façon suivante :
$\text{logit}(\Pi_{1i})=l_1 +k_1 X_i$
$\text{logit}(\Pi_{2i})=l_2 +k_2 X_i$
Rien ne contraint k_1 et k_2 à être de valeurs égales.

On prend comme caractéristique la gravité lésionnelle en trois classes : NISS 1-3, NISS 4-8, NISS 9+, que l'on note m (m=1, 2, 3)

On peut alors écrire pour tout individu appartenant à la catégorie m :

$$\text{logit}(\Pi_{1m}) = l_1 + k_{1m}$$
$$\text{logit}(\Pi_{2m}) = l_2 + k_{2m}$$

La probabilité d'un individu i d'être manqué (par les deux sources) est donnée par:
$\Pi_{0i}=(1- \Pi_{1i}) (1- \Pi_{2i})$

Cela s'écrit, pour tout individu i dans la catégorie m :
$\Pi_{0m}=(1- \Pi_{1m}) (1- \Pi_{2m})$

$$\Pi_{0m}= \frac{1}{(1+\exp(l_1 +k_{1m}))} \times \frac{1}{(1+\exp(l_2 +k_{2m}))}$$

On définit la variable indicatrice q_i comme précédemment, et on note les probabilités conditionnelles (aux observations) p_{i1}, p_{i2}, p_{i3}.

Pour tout individu i appartenant à la catégorie m, on a :

$$p_{1m} = \frac{\Pi_{1m}(1-\Pi_{2m})}{\Pi_{1m}(1-\Pi_{2m}) + \Pi_{2m}(1-\Pi_{1m}) + \Pi_{1m}\Pi_{2m}} = \frac{\exp(l_1 + k_{1m})}{\exp(l_1 + k_{1m}) + \exp(l_2 + k_{2m}) + \exp(l_1 + k_{1m} + l_2 + k_{2m})}$$

De la même façon :

$$p_{2m}= \frac{\exp(l_2 + k_{2m})}{\exp(l_1 + k_{1m}) + \exp(l_2 + k_{2m}) + \exp(l_1 + k_{1m} + l_2 + k_{2m})}$$

$$p_{3m}= \frac{\exp(l_1 + k_{1m} + l_2 + k_{2m})}{\exp(l_1 + k_{1m}) + \exp(l_2 + k_{2m}) + \exp(l_1 + k_{1m} + l_2 + k_{2m})}$$

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

$\text{logit}(p_{ki}) = \log\left(\frac{p_{ki}}{p_{1i}}\right) = \beta_k + \theta_k X_i$, avec k=2,3, et k=1 la catégorie de référence ($ß_1$=0, et Θ_1=0)

Ou, si l'on inclut une variable catégorielle, telle que le NISS en trois catégories (m=1, 2, 3), on écrit, pour tout individu i appartenant à la catégorie m :

$$\text{logit}(p_{km}) = \log\left(\frac{p_{km}}{p_{1m}}\right) = \beta_k + \theta_{km}$$

On ajuste le modèle, et on obtient une estimation des paramètres $ß_k$ et $Θ_k$. Ils sont reliés aux paramètres l et k de la façon suivante :

$$\begin{aligned}\beta_2 + \theta_{2m} &= \log\left(\frac{p_{2m}}{p_{1m}}\right)\\ &= \log\left(\frac{\exp(l_2 + k_{2m})/\left(\exp(l_1 + k_{1m}) + \exp(l_2 + k_{2m}) + \exp(l_1 + k_{1m} + l_2 + k_{2m})\right)}{\exp(l_1 + k_{1m})/\left(\exp(l_1 + k_{1m}) + \exp(l_2 + k_{2m}) + \exp(l_1 + k_{1m} + l_2 + k_{2m})\right)}\right)\\ &= \log\left(\frac{\exp(l_2 + k_{2m})}{\exp(l_1 + k_{1m})}\right)\\ &= (l_2 + k_{2m}) - (l_1 + k_{1m})\end{aligned}$$

$$\begin{aligned}\beta_3 + \theta_{3m} &= \log\left(\frac{p_{3m}}{p_{1m}}\right)\\ &= \log\left(\frac{\exp(l_1 + k_{1m} + l_2 + k_{2m})/\left(\exp(l_1 + k_{1m}) + \exp(l_2 + k_{2m}) + \exp(l_1 + k_{1m} + l_2 + k_{2m})\right)}{\exp(l_1 + k_{1m})/\left(\exp(l_1 + k_{1m}) + \exp(l_2 + k_{2m}) + \exp(l_1 + k_{1m} + l_2 + k_{2m})\right)}\right)\\ &= \log\left(\frac{\exp(l_1 + k_{1m} + l_2 + k_{2m})}{\exp(l_1 + k_{1m})}\right)\\ &= l_2 + k_{2m}\end{aligned}$$

Pour la catégorie m de référence, on pose $k_{1m}=k_{2m}= Θ_{1m}= Θ_{2m}=0$, et donc on en déduit :

$$\beta_2 = l_2 - l_1$$
$$\beta_3 = l_2$$

et ensuite :

$$\theta_{2m} = k_{2m} - k_{1m}$$
$$\theta_{3m} = k_{2m}$$

et donc, réciproquement :

$$l_1 = \beta_3 - \beta_2$$
$$l_2 = \beta_3$$
$$k_{1m} = \theta_{3m} - \theta_{2m}$$
$$k_{2m} = \theta_{3m}$$

En remplaçant les paramètres ß et Θ par leurs estimations, on en déduit une estimation des $Π_{0m}$ et ensuite de N et éventuellement de N_0, mais aussi une estimation des probabilités non conditionnelles aux observations : $Π_{1m}$ et $Π_{2m}$. En prenant l'inverse des $Π_{1m}$ on obtient les coefficients de correction du sous-enregistrement des forces de l'ordre.

$$\begin{aligned}\hat{\Pi}_{0m} &= \frac{1}{(1+\exp(l_1 + k_{1m}))\times(1+\exp(l_2 + k_{2m}))}\\ &= \frac{1}{(1+\exp(\hat{\beta}_3 - \hat{\beta}_2 + \hat{\theta}_{3m} - \hat{\theta}_{2m}))\times(1+\exp(\hat{\beta}_3 + \hat{\theta}_{3m}))}\end{aligned}$$

On peut alors estimer le nombre total N :

$$\hat{\Pi}_{0m} = \frac{\hat{N}_{0m}}{\hat{N}_m} \Leftrightarrow$$

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

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

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

Et $\hat{N} = \sum_{m=1}^{3} \hat{N}_m$

On estime aussi les probabilités non conditionnelles aux observations :

$$\hat{\Pi}_{1m} = \frac{1}{1 + \exp(-\hat{l}_1 - \hat{k}_{1m})} = \frac{1}{1 + \exp(\hat{\beta}_2 - \hat{\beta}_3 + \hat{\theta}_{2m} - \hat{\theta}_{3m})}$$ et

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

Et aussi :

$\hat{C}_{1m} = \frac{1}{\hat{\Pi}_{1m}} = 1 + \exp(\hat{\beta}_2 - \hat{\beta}_3 + \hat{\theta}_{2m} - \hat{\theta}_{3m})$, coefficient de correction du sous-enregistrement des données des forces de l'ordre, pour la catégorie m (NB : on pourrait aussi estimer C_{2m}, coefficient de correction du sous-enregistrement du Registre).

On généralise sans problème à plusieurs caractéristiques liées aux probabilités d'enregistrement. Le modèle multinomial généralisé s'écrit alors, avec l'inclusion de variables X_s (qui peuvent être continues ou catégorielles)

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

NB : dans l'ensemble du mémoire, les cœfficients de correction du sous-enregistrement sont indicés par j ; j identifie les catégories définies par la combinaison des variables incluses dans le modèle.

Nous utilisons le logiciel SAS, procédure LOGISTIC et l'option glogit. Les lignes de programme sont données en annexe.

Variables incluses dans le modèle

En premier lieu, les probabilités d'enregistrement dans les BAAC et dans le Registre 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 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 coté des BAAC, les facteurs liés à un fort biais de

sélection sont la gravité lésionnelle, le type d'usager, la présence/absence de tiers, le type de réseau, la force de l'ordre. Du coté du Registre, il s'agit de la gravité lésionnelle, éventuellement du type d'usager, et de la distance à l'hôpital le plus proche (disposant d'un service d'urgences). Cette variable n'est pas introduite car elle n'est pas dans les BAAC, et il n'est pas envisageable de la déterminer pour l'ensemble des données BAAC nationales (ce qui est théoriquement possible à partir des lieux d'accident et des adresses d'hôpitaux disposant d'un service d'urgences).
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.

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 donc traitées par de l'imputation simple. Pour la variable NISS en trois catégories, la valeur manquante (1,8 % des observations) a été remplacée par NISS 1-3, valeur la plus probable. Les valeurs manquantes pour âge, sexe et type d'usager ont été imputées en fonction des deux autres variables ou d'une seule autre (imputation simple dans les deux cas).
Les variables type de réseau et force de l'ordre présentent une proportion non négligeable de valeurs manquantes parmi les observations du Registre : 25,6 % et 22,4 % respectivement. Grosso modo, le type de réseau a été imputé en fonction du type d'usager et de la force de l'ordre en charge du secteur. Réciproquement, la force de l'ordre a été imputée en fonction du réseau et du type d'usager. Les détails de l'imputation sont donnés en annexe (il y a une certaine complexité due au fait que le type de réseau et la force de l'ordre sont parfois tous les deux manquants, que la commune est souvent manquante, et que le réseau autoroutier n'est présent que dans quelques communes).

Résultats

Le tableau suivant donne les paramètres du modèle. 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. Rappelons qu'en toute rigueur, l'odds ratio ne peut pas être interprété comme un risque relatif car l'évènement étudié n'est pas rare.

Tableau 38 : paramètres du modèle de capture-recapture, construit sur 93 298 blessés, BAAC et/ou Registre, Rhône, 1996-2004

variable	source	beta	OR	IC à 95 %	
ordonnée à l'origine	Registre seulement	1,78	5,90	5,30	6,58
ordonnée à l'origine	intersection	2,20	9,03	8,11	10,04
Gravité (p<0,0001)					
NISS 1-3	Registre seulement	1,48	4,37	4,04	4,73
NISS 4-8	Registre seulement	0,82	2,28	2,08	2,50
NISS 9+	Registre seulement		1,00		
NISS 1-3	intersection	-0,35	0,71	0,66	0,76
NISS 4-8	intersection	-0,15	0,86	0,79	0,94
NISS 9+	intersection	0,00	1,00		
type d'usager x tiers (p<0,0001)					
piétons, avec tiers	Registre seulement	-0,04	0,96	0,89	1,04
cyclistes, avec tiers	Registre seulement	0,59	1,81	1,57	2,08
cyclistes, sans tiers	Registre seulement	5,13	168,89	93,25	305,87
usagers de 2RM, avec tiers	Registre seulement	-0,14	0,87	0,81	0,94
usagers de 2RM, sans tiers	Registre seulement	2,51	12,32	10,46	14,52
automobilistes, avec tiers	Registre seulement		1,00		
automobilistes, sans tiers	Registre seulement	0,59	1,80	1,68	1,94
autres, avec tiers	Registre seulement	-0,77	0,46	0,40	0,54
autres, sans tiers	Registre seulement	1,72	5,56	4,50	6,88
piétons, avec tiers	intersection	0,03	1,03	0,95	1,11
cyclistes, avec tiers	intersection	0,15	1,16	1,00	1,35
cyclistes, sans tiers	intersection	1,06	2,89	1,54	5,44
usagers de 2RM, avec tiers	intersection	0,01	1,01	0,94	1,09
usagers de 2RM, sans tiers	intersection	0,67	1,95	1,64	2,33
automobilistes, avec tiers	intersection		1,00		
automobilistes, sans tiers	intersection	0,02	1,02	0,94	1,10
autres, avec tiers	intersection	-0,40	0,67	0,58	0,78
autres, sans tiers	intersection	0,32	1,38	1,10	1,75
type de réseau (p<0,0001)					
autoroutes	Registre seulement	-0,89	0,41	0,35	0,48
routes nationales	Registre seulement	-0,74	0,48	0,43	0,53
routes départementales	Registre seulement	-2,26	0,10	0,10	0,11
voies communales	Registre seulement		1,00		
autres	Registre seulement	-0,60	0,55	0,48	0,64
autoroutes	intersection	-0,35	0,70	0,60	0,83
routes nationales	intersection	-0,49	0,61	0,55	0,69
routes départementales	intersection	-1,91	0,15	0,14	0,16
voies communales	intersection	0,00	1,00		
autres	intersection	-1,67	0,19	0,15	0,23
force de l'ordre (p<0,0001)					
gendarmerie	Registre seulement	0,00	1,00		
CRS	Registre seulement	-1,36	0,26	0,22	0,29
police	Registre seulement	-1,76	0,17	0,16	0,19
gendarmerie	intersection		1,00		
CRS	intersection	-0,52	0,60	0,51	0,69
police	intersection	-0,80	0,45	0,42	0,49
année (p<0,0001)					
quantitative	Registre seulement	0,02	1,02	1,01	1,03
quantitative	intersection	-0,02	0,98	0,97	0,99

Nous présentons les résultats en donnant l'inverse de l'OR si celui-ci est inférieur à 1. Dans « Registre seulement », les cotes sont de 4 contre 1 de trouver des blessés NISS 1-3, par rapport aux blessés NISS 9+, et les cotes sont de 2 contre 1 de trouver des blessés NISS 4-8, par rapport aux blessés NISS 9+. Inversement, dans l'intersection BAAC-Registre, les cotes sont de 1,4 contre 1 de trouver des blessés NISS 9+ par rapport aux blessés NISS 1-3, et de 1,2 contre 1 de trouver des blessés NISS 9+ par rapport aux blessés NISS 4-8.

Concernant le type d'usager, dans « Registre seulement », les cotes sont de 1 contre 1 de trouver des piétons versus des automobilistes avec tiers, de 1,1 contre 1 de trouver des usagers de 2RM avec tiers, de presque 2 contre 1 de trouver des cyclistes blessés avec tiers. La situation est similaire dans l'intersection BAAC-Registre.
Pour les blessés dans un accident sans tiers, dans « Registre seulement », les cotes sont presque de deux contre 1 de trouver des automobilistes sans tiers versus des automobilistes avec tiers, de 12 contre 1 de trouver des usagers de 2RM sans tiers, et de 169 contre 1 de trouver des cyclistes sans tiers.
Dans l'intersection BAAC-Registre, les odds ratios sont de bien moindre ampleur : les cotes sont de 1 contre 1 de trouver des automobilistes sans tiers versus des automobilistes avec tiers, de 2 contre 1 de trouver des usagers de 2RM sans tiers et de 3 contre 1 de trouver des cyclistes sans tiers (versus des automobilistes avec tiers).
Dans « Registre seulement », les cotes sont de 10 contre 1 de trouver des blessés accidentés sur voie communale versus route départementale, de 2 contre 1 de trouver des blessés accidentés sur voie communale versus route nationale et de 2,4 contre 1 de trouver des blessés accidentés sur voie communale versus autoroute. Les ampleurs des odds ratios sont moindres dans l'intersection BAAC-Registre.
Enfin, dans « Registre seulement », les cotes sont légèrement au dessus de 1 contre 1 de trouver des blessés entre une année et la suivante. Dans l'intersection BAAC-Registre, les cotes sont légèrement en dessous de 1 contre 1 de trouver des blessés entre une année et la suivante.

Nous présentons ci-dessous les coefficients de correction du sous-enregistrement des données des forces de l'ordre pour quelques profils de blessés : les blessés dans un accident en 2004, sur route nationale, enregistrés par la force de l'ordre « police », en laissant varier le type d'usager, la présence / absence de tiers et la gravité lésionnelle (NISS).

Tableau 39 : coefficients de correction du sous-enregistrement des données des forces de l'ordre pour quelques profils de blessés (accidents sur routes nationales, en zone police, en 2004)

Type d'usager	tiers	NISS	coefficient de correction
piéton	avec tiers	NISS 1-3	2,5
piéton	avec tiers	NISS 4-8	1,7
piéton	avec tiers	NISS 9+	1,2
cycliste	avec tiers	NISS 1-3	3,6
cycliste	avec tiers	NISS 4-8	2,1
cycliste	avec tiers	NISS 9+	1,4
cycliste	sans tiers	NISS 1-3	96,8
cycliste	sans tiers	NISS 4-8	42,0
cycliste	sans tiers	NISS 9+	16,5
usager de 2RM	avec tiers	NISS 1-3	2,4
usager de 2RM	avec tiers	NISS 4-8	1,6
usager de 2RM	avec tiers	NISS 9+	1,2
usager de 2RM	sans tiers	NISS 1-3	11,3
usager de 2RM	sans tiers	NISS 4-8	5,4
usager de 2RM	sans tiers	NISS 9+	2,7
automobiliste	avec tiers	NISS 1-3	2,6
automobiliste	avec tiers	NISS 4-8	1,7
automobiliste	avec tiers	NISS 9+	1,3
automobiliste	sans tiers	NISS 1-3	3,9
automobiliste	sans tiers	NISS 4-8	2,2
automobiliste	sans tiers	NISS 9+	1,5
autre	avec tiers	NISS 1-3	2,1
autre	avec tiers	NISS 4-8	1,5
autre	avec tiers	NISS 9+	1,2
autre	sans tiers	NISS 1-3	7,6
autre	sans tiers	NISS 4-8	3,8
autre	sans tiers	NISS 9+	2,1

Les coefficients de correction du sous-enregistrement ont une très grande amplitude. Dans l'exemple donné, ils sont compris entre 1,2 (piétons avec tiers, NISS 9+) et 96,8 (cyclistes sans tiers, NISS 1-3). On note l'ampleur décroissante en relation inverse de la gravité (toutes choses égales par ailleurs) et la valeur plus grande pour les accidents sans tiers versus avec tiers (toutes choses égales par ailleurs).

Discussion

Les odds ratios estimés par le modèle de capture-recapture sont tout à fait conformes aux biais de sélection identifiés (chap. 1) et à l'ampleur de ces biais de sélection. Nous ne présentons pas d'intervalle de confiance pour cette étape ; ils sont estimés par bootstrap sur l'ensemble de la procédure, et seront présentés pour les estimations d'effectifs et d'incidences France entière.

Les coefficients de correction du sous-enregistrement paraissent très élevés pour les cyclistes sans tiers. Il faut rappeler que le taux de couverture des cyclistes par les données des forces de l'ordre est bas, voire extrêmement bas pour ceux blessés légèrement. Ces coefficients de correction sont donc tout à fait plausibles.

Nous avons aussi estimé les effectifs de blessés dans le Rhône avec ce nouveau modèle, notamment pour les comparer aux résultats obtenus avec la méthode simple utilisée au chapitre 3. Les résultats sont donnés en annexe. Il ressort que le nombre total de victimes pour l'année 2001 est estimé à 13 700 (versus 14 500 dans le chapitre 3). Ce moindre effectif s'explique en grande partie par le modèle de prédiction du NISS : celui-ci prédit une plus grande proportion de blessés NISS 9+ chez les blessés identifiés seulement dans les BAAC que l'évaluation basée sur les certificats médicaux inclus dans les procès-verbaux (cf. chap. 3), et donc une plus grande proportion de NISS 9+ dans l'ensemble

des blessés, et ceux-ci sont les mieux enregistrés donc avec un coefficient de correction mineur. En revanche, ce modèle de prédiction du NISS signifie un moindre effectif de blessés légers parmi les données des forces de l'ordre, lesquels sont corrigés par de forts coefficients de correction du sous-enregistrement, ce qui se traduit par un effectif total de blessés légers moindre, et donc aussi un effectif des blessés toutes gravités moindre.
Les deux modèles donnent des taux de couverture des BAAC et du Registre légèrement différents. Le modèle plus complet construit ici donne des taux de couverture des BAAC légèrement plus élevés que précédemment. Il donne des taux de couverture du Registre légèrement plus bas pour les blessés NISS 9+, mais légèrement plus haut pour les blessés NISS 4-8.
Rappelons que les deux modèles de capture-recapture diffèrent sur d'autres aspects : le plus important est que le modèle de capture-recapture utilisé ici tient compte de deux facteurs de biais supplémentaires (et importants) de plus que la méthode utilisée dans le chapitre 3 : le type de force de l'ordre et le type de réseau. Une différence supplémentaire est l'exclusion d'une petite proportion de blessés ne satisfaisant pas aux critères des BAAC ou du Registre (par application des conditions 5 et 6 de l'approche capture-recapture).

Étape 4 : Application des coefficients de correction aux données policières nationales

La méthode mise en œuvre est analogue à une standardisation indirecte d'un taux d'incidence (Estève, Benhamou 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, type d'usager, tiers, réseau et force de l'ordre) au lieu des facteurs de confusion âge et sexe.

L'estimation nationale du nombre 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, à la place des strates d'âge et de sexe,
C_j sont les coefficients de correction du sous-enregistrement, de référence (Rhône) au lieu des taux d'incidences de référence, 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.

En écrivant l'étape intermédiaire :

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

nous pouvons utiliser la terminologie de la théorie des sondages (Ardilly 1994). Cela montre qu'il y a d'une part une post-stratification sur les facteurs du sous-enregistrement :

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

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

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

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, réduit ici, à un département. La synthèse bibliographique en annexe fournit plus de détails.

Cette formulation signifie, que pour une catégorie donnée, par exemple les cyclistes blessés graves dans un accident avec tiers sur une route départementale 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 ailleurs en 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.

Les strates sont définies par la combinaison des facteurs de biais de sélection : la gravité lésionnelle (NISS en trois niveaux), le type d'usager, l'implication de tiers (oui/non), le type de réseau et la force de l'ordre. Le NISS a été prédit sur les données des forces de l'ordre France entière lors de l'étape 1. Les autres variables sont disponibles. En ce qui concerne le type de force de l'ordre, nous assimilons la police aux frontières et la préfecture de police de Paris à la force de l'ordre « police ».

Statistiques présentées : effectifs et incidences

Nous présentons les résultats par critère de santé étudié : blessés toutes gravités, blessés hospitalisés, blessés graves, blessés avec séquelles majeures prévues. Nous déclinons chaque critère en fonction du type d'usager, de l'âge et du sexe. Pour comparaison, nous présentons les mêmes critères basés sur les données des forces de l'ordre lorsqu'ils existent.

Les résultats sont donnés en termes d'effectifs et d'incidence pour 100 000 habitants. L'incidence est le ratio entre le nombre de blessés par accidents de la route sur le territoire français et la taille de la population habitant le territoire français. De la même façon qu'au niveau départemental, le nombre de blessés inclut des personnes n'habitant pas en France, et n'inclut pas les français victimes d'un accident à l'étranger. Ce nombre de blessés est cependant rapporté à la population française. L'incidence ainsi mesurée permet des comparaisons internationales, où les nombres de blessés sont comptabilisés de la même façon. Elle permet aussi des comparaisons avec d'autres états de santé dégradée. Les données de population proviennent de l'INSEE ; ce sont des données annuelles, par âge et sexe. L'exposition est ainsi mesurée en personnes-années.

Les résultats selon l'âge et le sexe sont donnés en termes d'incidences. En revanche, pour le type d'usager, nous ne disposons pas des populations correspondantes. Nous présentons donc les résultats seulement en termes d'effectif.

Le taux de létalité est défini comme le nombre de tués (à six jours) rapporté au nombre des blessés et tués. Le taux de gravité, parmi les blessés non décédés (dans les six jours), est défini comme le nombre de blessés graves rapporté au nombre des blessés toutes gravités. Étant affectés par les nouvelles estimations des nombres de blessés, nous donnons leurs estimations, selon les données des forces de l'ordre et selon les résultats de la projection.

Nous calculons et présentons les évolutions d'une année sur l'autre (en pourcentage), ainsi que la moyenne de ces évolutions annuelles.

Intervalles de confiance

Les intervalles de confiance sont obtenus par bootstrap sur l'ensemble de la procédure de projection (soit les quatre étapes). Nous utilisons un bootstrap non-paramétrique avec 2000 itérations. L'intervalle de confiance à 95 % des effectifs est estimé par les percentiles 2,5 % et 97,5 %. Les intervalles de confiance des incidences sont simplement obtenus en rapportant les bornes de l'intervalle de confiance des effectifs à la population correspondante, considérée comme non-aléatoire. Pour un critère de santé donné, les temps de calcul sont de l'ordre de 48 heures sur un PC standard récent.

Validité des modèles et de l'ensemble de la procédure d'estimation

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 résultats avec des estimations extérieures dès qu'elles existent. Cela est possible pour le nombre de traumatisés crâniens graves, le nombre de traumatisés médullaires et le nombre d'hospitalisés.

Résultats

Blessés toutes gravités

Effectifs et incidences

Sur la période d'étude 1996-2004, le nombre annuel moyen de blessés (toutes gravités) est de 150 400 d'après les données des forces de l'ordre ; il est estimé par projection à 514 300. La projection estime 3,4 fois plus de blessés. Les estimations annuelles sont données Tableau 40 et Tableau 41. En 2004, dernière année de la période étudiée, il y a 108 800 blessés d'après les données des forces de l'ordre et 400 200 d'après la projection. Cela correspond à 3,7 fois plus.
En termes d'incidence, en moyenne annuelle, il y a 255 blessés pour 100 000 habitants d'après les données des forces de l'ordre et 871 blessés pour 100 000 habitants d'après la projection. En 2004, l'incidence est respectivement de 180/100000 d'après les données des forces de l'ordre et de 661/100000 d'après la projection.

Tableau 40 : effectif et incidence des blessés (légers et graves), France, source = données des forces de l'ordre

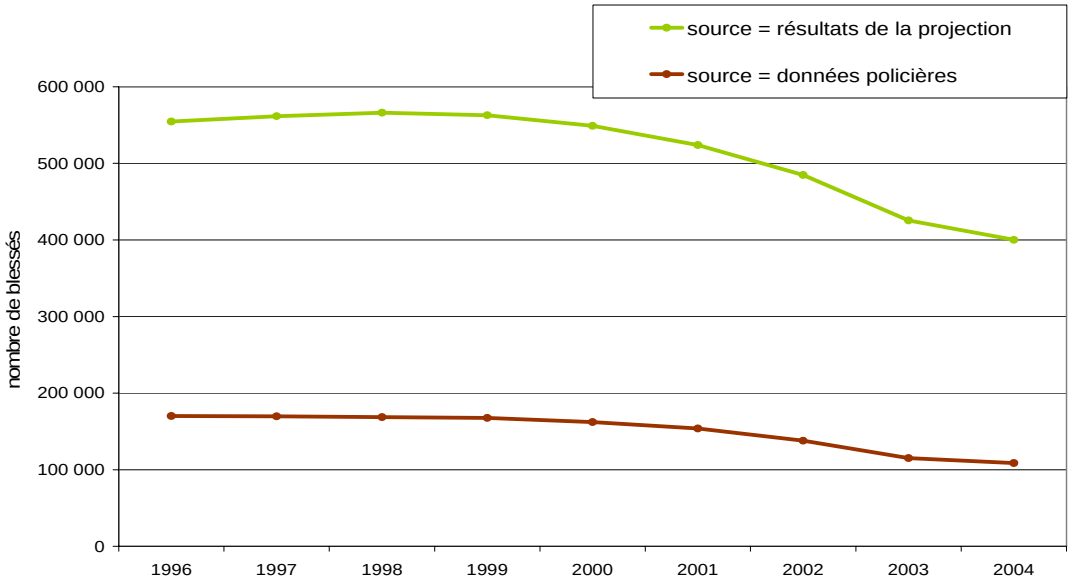
année	effectif	incidence (/100 000 hab.)	évolution
1996	170 117	293,2	
1997	169 577	291,3	-0,6 %
1998	168 535	288,6	-0,9 %
1999	167 572	285,7	-1,0 %
2000	162 117	274,7	-3,8 %
2001	153 945	259,2	-5,6 %
2002	137 839	230,6	-11,0 %
2003	115 170	191,5	-17,0 %
2004	108 727	179,7	-6,2 %
moyenne	150 400	254,9	-5,8 %

Tableau 41 : effectif et incidence des blessés (NISS 1+), France, source = résultats de la projection

année	effectif	IC à 95 %		incidence (/100 000 hab.)	IC à 95 %		évolution
1996	554 688	525 245	576 235	955,9	905,2	993,1	
1997	561 590	533 145	580 784	964,8	915,9	997,8	+0,9 %
1998	566 319	539 100	584 224	969,8	923,2	1000,4	+0,5 %
1999	562 929	537 909	578 256	959,6	917,0	985,8	-1,0 %
2000	548 899	526 193	560 923	930,1	891,7	950,5	-3,1 %
2001	524 035	502 697	534 739	882,3	846,4	900,3	-5,1 %
2002	484 853	462 996	495 834	811,1	774,5	829,5	-8,1 %
2003	425 441	404 023	436 463	707,2	671,6	725,6	-12,8 %
2004	400 236	379 885	410 015	661,3	627,7	677,5	-6,5 %
moyenne	514 332	490 133	528 608	871,3	828,9	894,0	-4,4 %

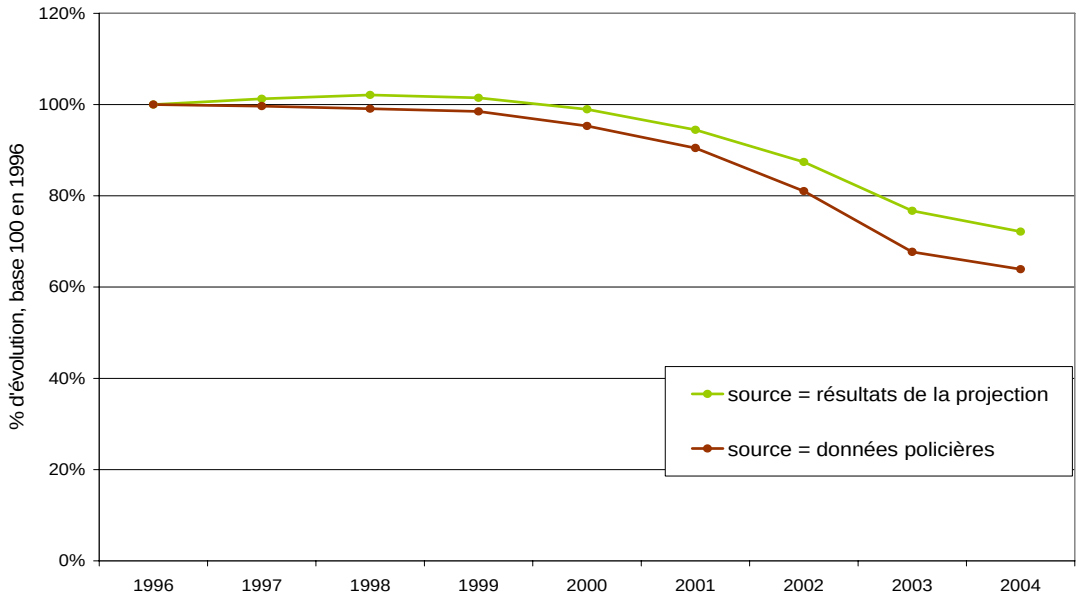
La Figure 12 donne les ordres de grandeur des nombres de blessés selon la source des données : des forces de l'ordre ou issues de la projection.

Figure 12 : nombre de blessés toutes gravités, France, selon la source des données



Le graphique suivant donne les évolutions en base 100 du nombre de blessés toutes gravités selon la source des données : des forces de l'ordre ou résultats de la projection.

Figure 13 : évolution du nombre des blessés toutes gravités, France, selon la source des données

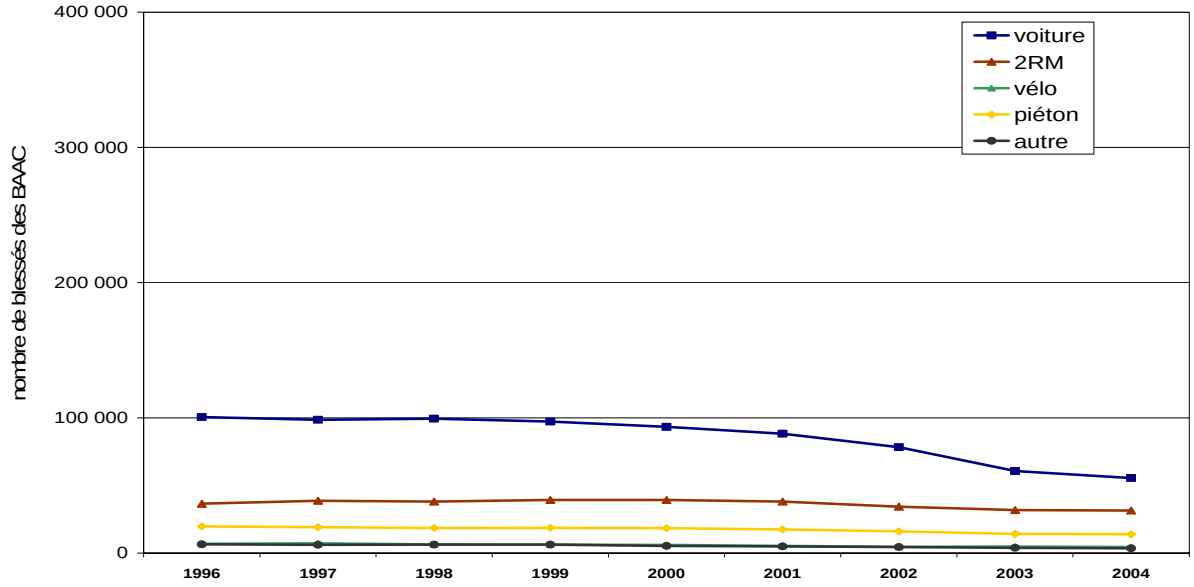


Le nombre total de blessés baisse plus vite selon les données des forces de l'ordre que selon les résultats de la projection.

Blessés toutes gravités, selon le type d'usager

D'après les données des forces de l'ordre, le nombre de blessés toutes gravités arrondi à la centaine, est de 17 300 piétons, 5 700 cyclistes, 36 400 usagers de 2RM, 85 700 automobilistes, et 5 200 autres usagers (effectifs et incidences annuelles en annexe). La figure suivante illustre les ordres de grandeur du nombre total de blessés, par type d'usager, selon les données des forces de l'ordre.

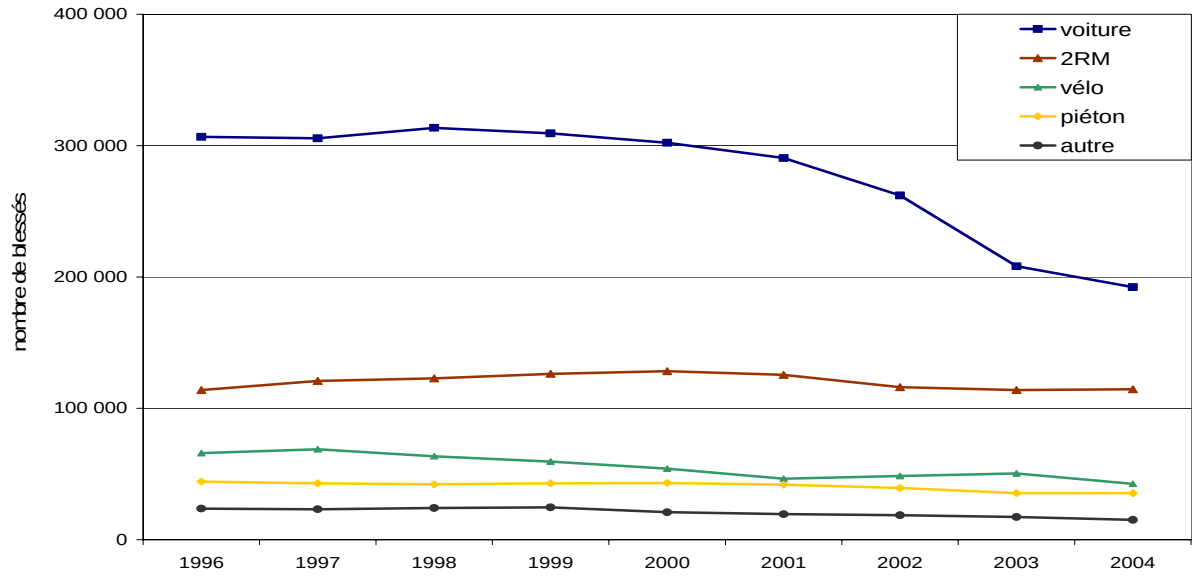
Figure 14 : nombre de blessés toutes gravités selon le type d'usager, France, source = données des forces de l'ordre



Selon les données des forces de l'ordre, le nombre de cyclistes blessés est au même niveau que celui des usagers autres, en dessous du nombre des piétons blessés.

D'après la projection, le nombre annuel moyen de blessés (NISS 1+) est de 40 900 piétons, 55 600 cyclistes, 120 300 usagers de 2RM, 276 700 automobilistes et 20 800 autres usagers (estimations annuelles, effectifs et incidences, en annexe). Le graphique suivant illustre les ordres de grandeur.

Figure 15 : nombre de blessés toutes gravités selon le type d'usager, France, source = résultats de la projection



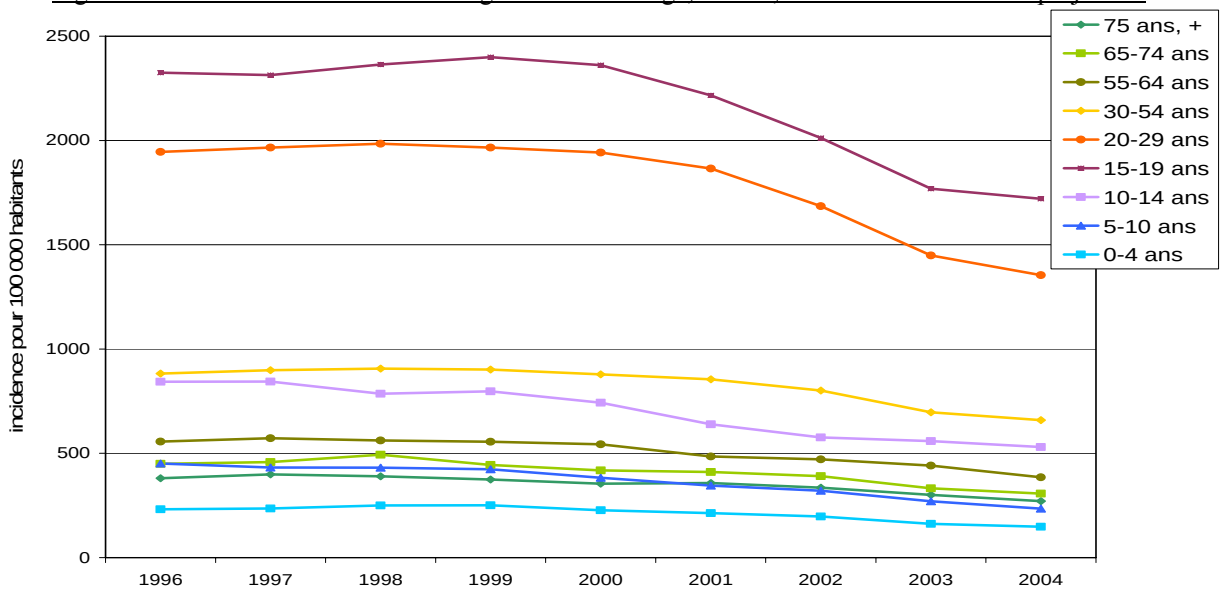
Selon la projection, le nombre de blessés cyclistes est du même ordre de grandeur que le nombre de blessés piétons.
Selon la projection, l'incidence des piétons blessés diminue en moyenne de 3,2 % par an, celle des cyclistes de 5,5 % par an, celle des usagers de 2RM seulement de 0,4 % par an. L'incidence d'automobilistes blessés connaît la plus grande baisse, -5,9 % par an, juste devant la baisse de l'incidence des usagers autres (-5,7 %).

Blessés toutes gravités, selon l'âge

Les résultats selon les données des forces de l'ordre ne sont pas donnés ici (mais en annexe) ; le classement des classes d'âges en fonction de leur incidence de blessés est le même que celui issu de la projection. Les ordres de grandeur diffèrent.

La figure ci-dessous donne les incidences selon la classe d'âge, d'après les résultats de la projection.

Figure 16 : incidence des blessés toutes gravités selon l'âge, France, source = résultats de la projection



Deux classes d'âge se distinguent nettement par une incidence bien plus élevée : les 15-19 ans et les 20-29 ans.
Les évolutions moyennes sur 1996-2004 sont de : -5,1 % chez les 0-4 ans, -7,7 % chez les 5-9 ans, -3,6 % chez les 15-19, - 4,3 % chez les 20-29, -3,5 % chez les 30-54, -4,4 % chez les 55-64, -4,4 % chez les 65-74, et -4,0 % chez les 75 ans et plus. Les deux classes d'âge les plus à risque de blessures connaissent une faible baisse. Ce sont les 0-9 ans (les moins à risque) qui connaissent la baisse la plus forte.

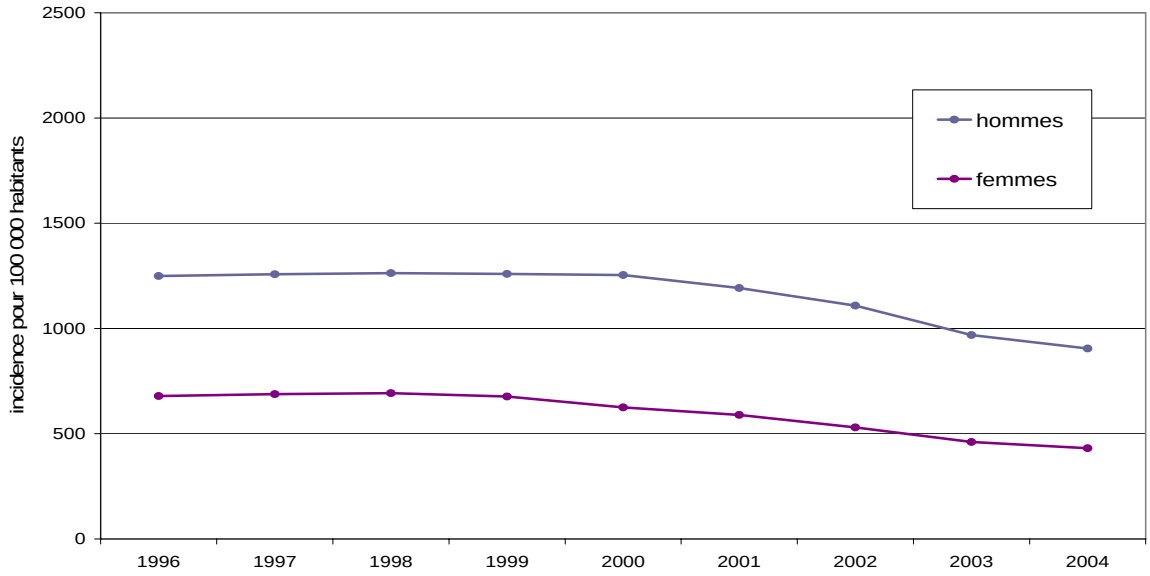
Blessés toutes gravités, selon le sexe

Les incidences selon les données des forces de l'ordre sont données en annexe.

Selon les résultats de la projection, l'incidence des blessés hommes est de 1160/100 000, celle des femmes est de 596/100 000. Le sex-ratio est de 2,0 en moyenne.

La figure ci-dessous illustre l'ampleur des incidences des blessés toutes gravités chez les hommes et chez les femmes.

Figure 17 : incidence des blessés toutes gravités selon le sexe, France, source = résultats de la projection



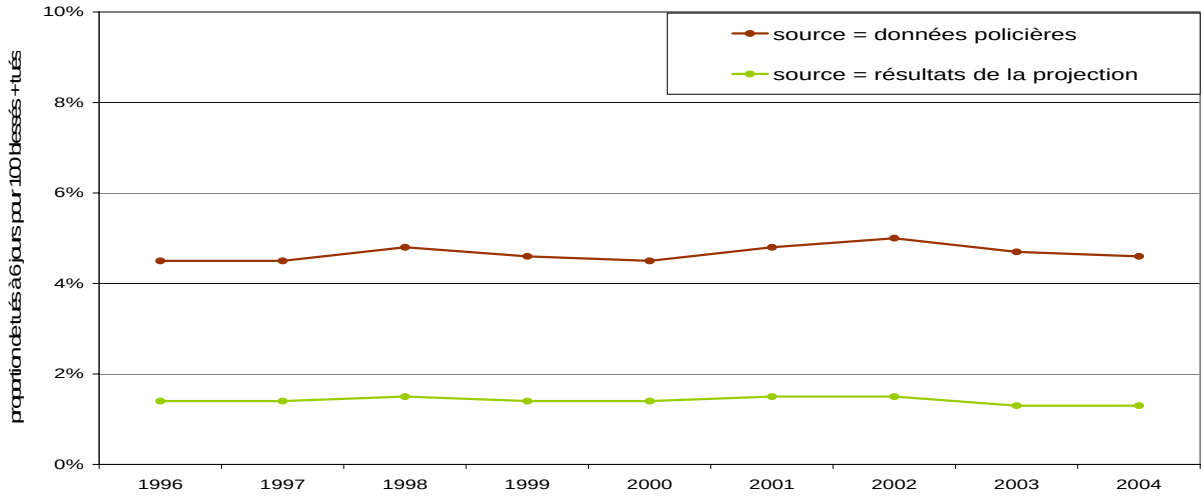
La baisse moyenne est de -3,8 % chez les hommes et de -5,4 % chez les femmes.

Taux de létalité

La létalité est définie ici par le ratio entre le nombre de tués à six jours et la somme des nombres de blessés et tués à six jours. Avec les estimations par projection, le numérateur ne change pas mais le dénominateur est largement augmenté, et donc la létalité est réévaluée à la baisse. Selon les données des forces de l'ordre le taux annuel de létalité est de 4,7 % en moyenne sur la période 1996-2004. D'après les résultats de la projection, le taux de létalité est à 1,4 % en moyenne.

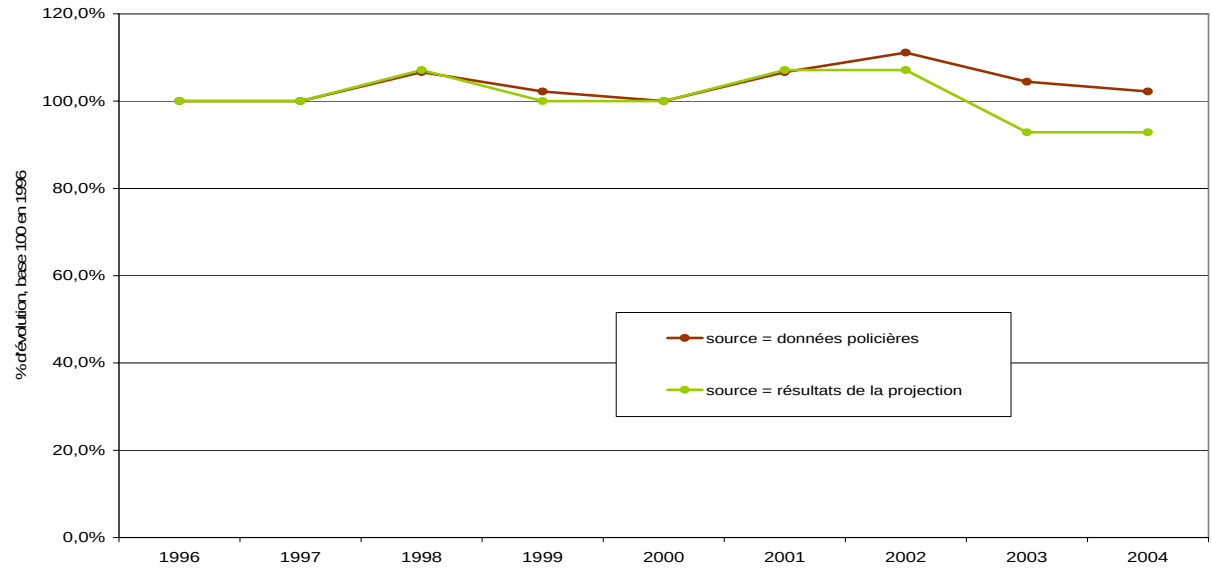
La figure ci-dessous illustre les valeurs différentes du taux de létalité selon la source des données.

Figure 18 : taux de létalité, France, selon la source des données



L'évolution du taux de létalité selon la source est donnée dans le graphique ci-dessous.

Figure 19 : évolution en base 100 du taux de létalité, France, selon la source des données

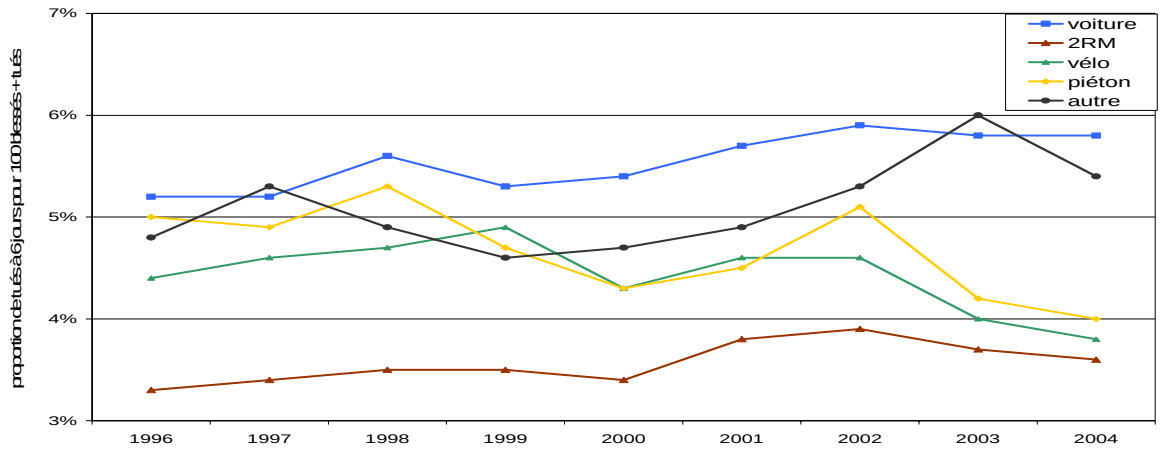


Il n'y a pas d'évolution franche ; les deux courbes sont assez fluctuantes (en lien direct avec la fluctuation du nombre de tués). Si nous calculons l'évolution moyenne sur la période, la létalité selon les données des forces de l'ordre augmente de +0,38 % et de +0,35 % selon la projection.

Létalité selon le type d'usager

NB : Les échelles des deux graphiques ci-dessous sont différentes, le but étant de rendre comparables l'ampleur des variations.
Le graphique ci-dessous donne le taux de létalité des types d'usager selon les données des forces de l'ordre.

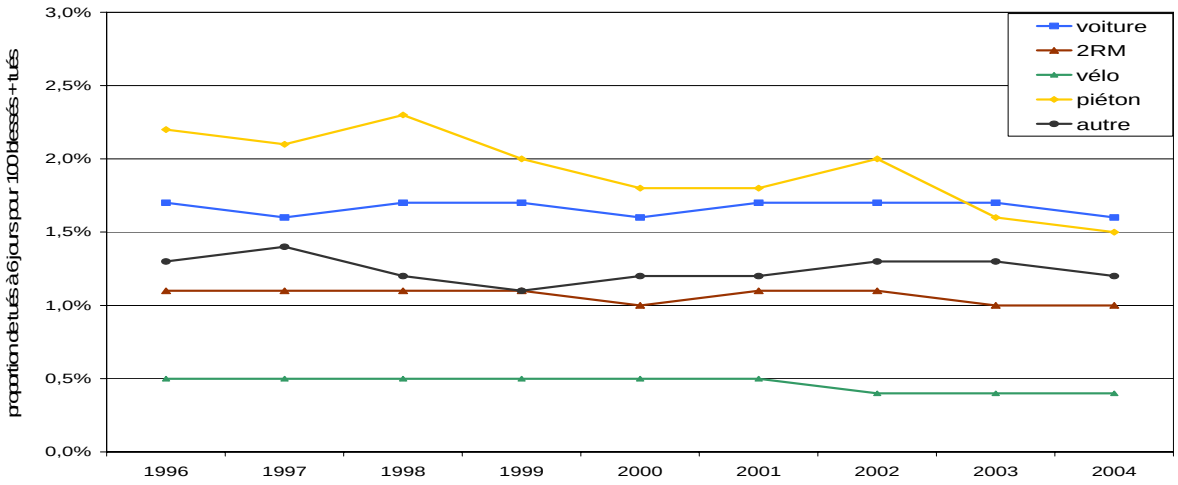
Figure 20 : létalité, selon le type d'usager, France, source = données des forces de l'ordre



D'après les données des forces de l'ordre, les automobilistes ont la plus forte létalité, et les usagers de 2RM ont la plus faible. Piétons et cyclistes ont sensiblement la même létalité.
La létalité des piétons et celle des cyclistes semblent diminuer entre 1996 et 2004, contrairement à celles des automobilistes, des usagers de 2RM et des usagers « autre » qui, elles, semblent augmenter.

Le graphique suivant donne la létalité par type d'usager, selon les résultats de la projection.

Figure 21 : létalité, selon le type d'usager, France, source = résultats de la projection



Le classement des types d'usager selon leur létalité est très différent selon la source des données.
D'après la projection, les piétons ont la létalité la plus haute ; cela provient de ce que la létalité des autres usagers est bien plus modifiée à la baisse (car leur dénominateur s'est bien plus amplifié par la correction du sous-enregistrement). Les cyclistes ont la plus basse létalité (cela s'explique par une grande proportion d'accidents sans tiers où la létalité est très basse).
En termes d'évolution, la létalité des piétons semble diminuer. La létalité des autres usagers semble plutôt stable.

Le taux de létalité selon l'âge, et selon le sexe sont donnés en annexe ; le classement des classes d'âge en fonction de leur létalité est le même selon la source des données (forces de l'ordre ou résultats de la projection) ; il en est de même pour la létalité selon le sexe.

Blessés hospitalisés

Effectifs et incidences

Nous rappelons que les blessés hospitalisés ne sont pas décomptés par les forces de l'ordre avant 2005. Les effectifs de blessés hospitalisés « d'après les données des forces de l'ordre » ci-dessous sont des effectifs prédits, comme proportions des effectifs totaux de blessés dénombrés par les forces de l'ordre (étape 1 de l'ensemble du processus d'estimation). Cette prédiction conduit à une moyenne annuelle de 61 400 blessés hospitalisés. L'estimation par projection conduit à 137 000 blessés hospitalisés en moyenne annuelle. Cela correspond à 2,2 fois plus. Les estimations annuelles sont présentées ci-dessous, Tableau 42 et Tableau 43.

Tableau 42 : effectif et incidence prédits des blessés hospitalisés, France, source = données des forces de l'ordre

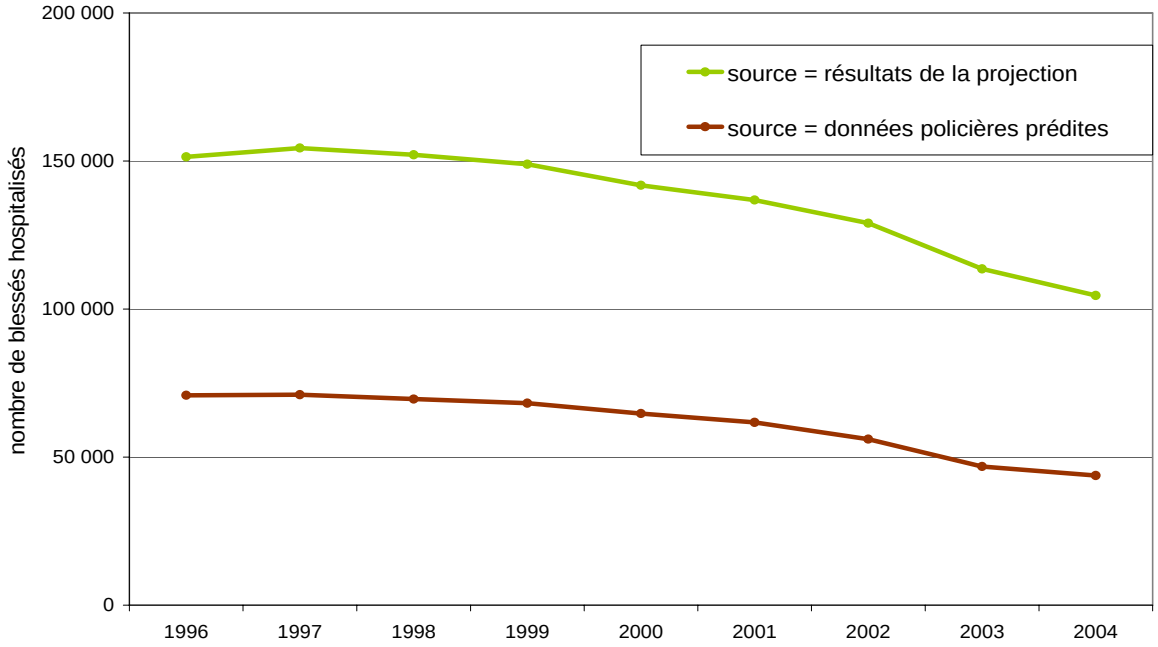
année	effectif	IC à 95 %		incidence (/100 000 hab.)	IC à 95 %		évolution
1996	70 878	69 032	72 745	122,1	119,0	125,4	
1997	71 055	69 504	72 662	122,1	119,4	124,8	-0,1 %
1998	69 613	68 290	71 046	119,2	116,9	121,7	-2,3 %
1999	68 216	67 030	69 453	116,3	114,3	118,4	-2,4 %
2000	64 723	63 548	65 900	109,7	107,7	111,7	-5,7 %
2001	61 719	60 509	62 941	103,9	101,9	106,0	-5,3 %
2002	56 066	54 776	57 255	93,8	91,6	95,8	-9,7 %
2003	46 804	45 542	47 966	77,8	75,7	79,7	-17,0 %
2004	43 779	42 404	45 072	72,3	70,1	74,5	-7,0 %
moyenne	61 428	60 071	62 782	104,1	101,8	106,4	-6,2 %

Tableau 43 : effectif et incidence des blessés hospitalisés, France, source = résultats de la projection

année	effectif	IC à 95 %		incidence (/100 000 hab.)	IC à 95 %		évolution
1996	151 427	143 633	157 129	261,0	247,5	270,8	
1997	154 411	146 880	159 605	265,3	252,3	274,2	+1,7 %
1998	152 129	145 236	156 566	260,5	248,7	268,1	-1,8 %
1999	148 944	142 550	152 582	253,9	243,0	260,1	-2,5 %
2000	141 801	135 968	144 937	240,3	230,4	245,6	-5,4 %
2001	136 859	131 124	139 881	230,4	220,8	235,5	-4,1 %
2002	128 981	122 858	132 364	215,8	205,5	221,4	-6,4 %
2003	113 557	107 538	117 031	188,8	178,8	194,5	-12,5 %
2004	104 568	98 803	107 967	172,8	163,3	178,4	-8,5 %
moyenne	136 964	130 510	140 896	232,1	221,1	238,7	-4,9 %

Le graphique ci-dessous illustre les ordres de grandeur du nombre de blessés hospitalisés, selon la source des données.

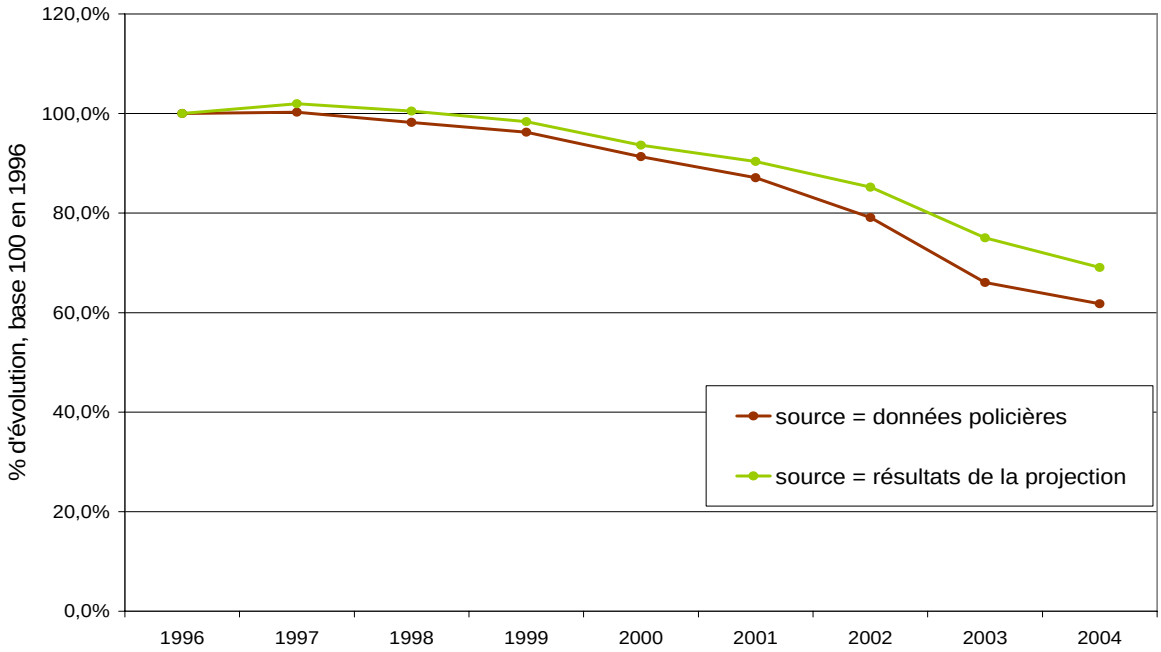
Figure 22 : nombre de blessés hospitalisés, France, selon la source des données



Il montre l'ampleur de la différence, entre les sources de données, des effectifs de blessés hospitalisés

L'évolution du nombre de blessés hospitalisés selon la source des données est donnée par le graphique ci-dessous.

Figure 23 : évolution en base 100 du nombre de blessés hospitalisés, France, selon la source des données

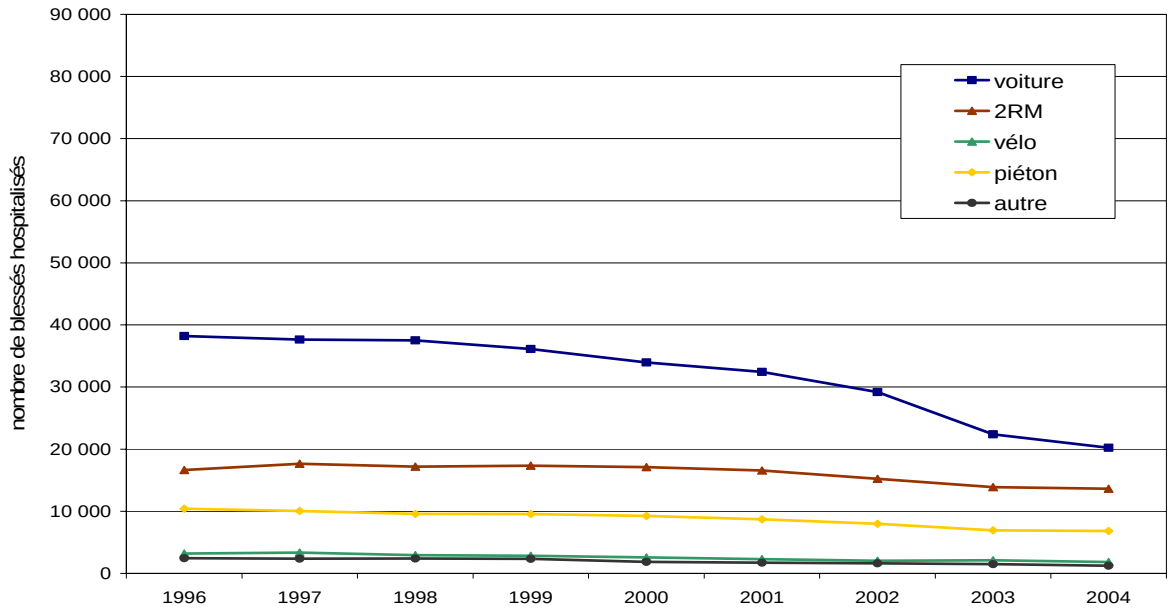


Ce graphique illustre la baisse moindre du nombre de blessés hospitalisés, selon les résultats de la projection.

Blessés hospitalisés, selon le type d'usager

Les ordres de grandeur des nombres de blessés hospitalisés selon le type d'usager, basés sur les données des forces de l'ordre, sont donnés par le graphique ci-dessous.

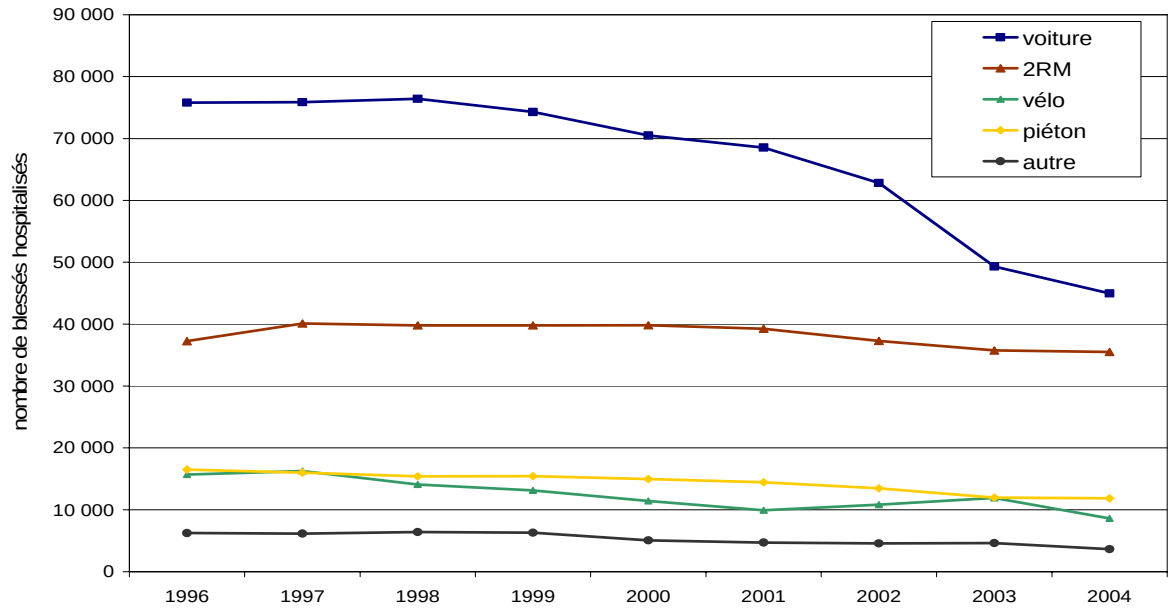
Figure 24 : nombre de blessés hospitalisés prédits selon le type d'usager, France, source = données des forces de l'ordre



Selon les données des forces de l'ordre, les cyclistes hospitalisés sont au même nombre que les usagers autres hospitalisés, et les moins nombreux. Viennent ensuite, par ordre croissant, les piétons, puis les usagers de 2RM et enfin les automobilistes.

La figure ci-dessous donne les ordres de grandeur des hospitalisés selon les résultats de la projection.

Figure 25 : nombre de blessés hospitalisés selon le type d'usager, France, source = résultats de la projection



Selon la projection, les effectifs de blessés hospitalisés sont bien plus élevés, quel que soit le type d'usager. L'ordre des types d'usager selon la source des données n'est pas changé, mais les cyclistes hospitalisés se trouvent en fait au même niveau que les piétons hospitalisés (en effectifs). L'écart entre les types d'usager n'est pas le même non plus : les piétons, les cyclistes et les usagers « autres » ont des nombres d'hospitalisés bien plus faibles que les automobilistes et les usagers de 2RM. Ceux-ci sont d'ailleurs de plus en plus proches.
La baisse du nombre d'automobilistes hospitalisés est forte : -6,5 %, alors que celle des usagers de 2RM est faible : -1,1 % (évolution annuelle moyenne). Elle est de -4,5 %% pour les piétons, -6,9 % pour les cyclistes, et -6,5 % pour les autres usagers.
Les effectifs et les incidences annuels sont donnés en annexe.

Blessés graves

Effectifs et incidences

Les données des forces de l'ordre indiquent un nombre de blessés graves de 28 000 en moyenne par an (ceux nécessitant plus de six jours d'hospitalisation). Par la méthode de projection, le nombre moyen de blessés graves (au sens NISS 9+) est estimé à 60 800 par an. Cela représente 2,2 fois plus.
En termes d'incidence, cela correspond à 47/100 000 blessés graves par an d'après les données des forces de l'ordre et à 103/100 000 d'après la projection. Les estimations annuelles sont présentées dans les deux tableaux ci-dessous.

Tableau 44 : effectif et incidence des blessés graves (nécessitant plus de 6 jours d'hospitalisation), France, source = données des forces de l'ordre

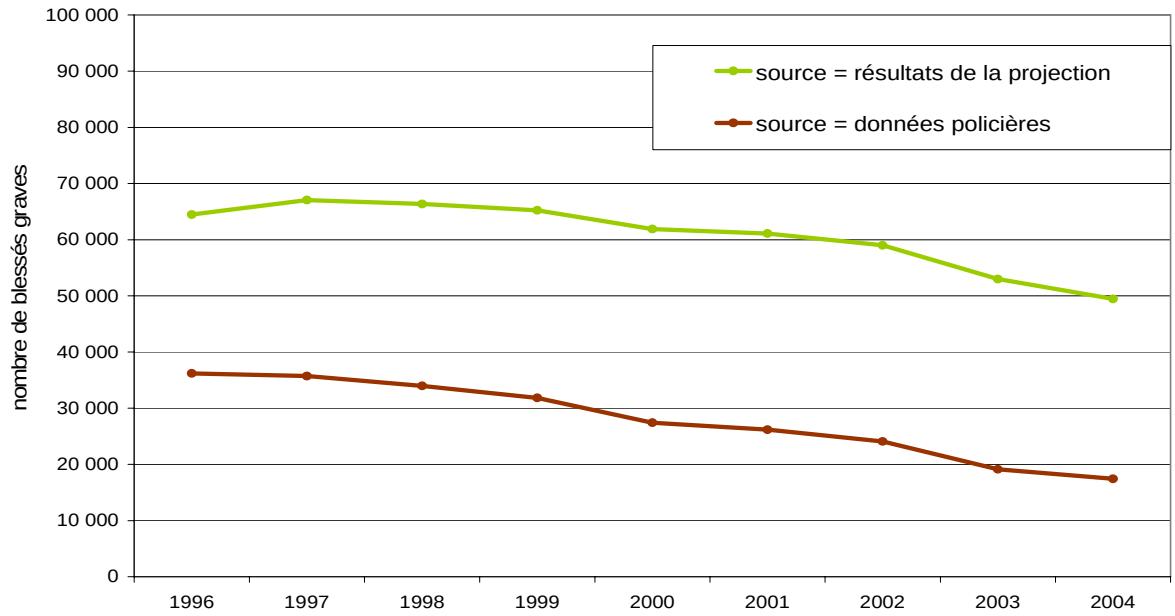
année	effectif	incidence (/100 000 hab.)	évolution
1996	36 204	62,4	
1997	35 716	61,4	-1,6 %
1998	33 977	58,2	-5,2 %
1999	31 851	54,3	-6,7 %
2000	27 407	46,4	-14,5 %
2001	26 192	44,1	-5,0 %
2002	24 091	40,3	-8,6 %
2003	19 110	31,8	-21,1 %
2004	17 435	28,8	-9,4 %
moyenne	27 998	47,3	-9,0 %

Tableau 45 : effectif et incidence des blessés graves (NISS 9+), France, source = résultats de la projection

année	effectif	IC à 95 %		incidence (/100 000 hab.)	IC à 95 %		évolution
1996	64 485	60 332	67 644	111,1	104,0	116,6	
1997	67 036	63 030	69 798	115,2	108,3	119,9	+3,6 %
1998	66 359	62 753	68 714	113,6	107,5	117,7	-1,3 %
1999	65 244	62 002	67 178	111,2	105,7	114,5	-2,1 %
2000	61 905	58 865	63 586	104,9	99,8	107,8	-5,7 %
2001	61 095	58 098	62 764	102,9	97,8	105,7	-1,9 %
2002	59 017	55 747	60 886	98,7	93,3	101,9	-4,0 %
2003	52 978	49 465	54 998	88,1	82,2	91,4	-10,8 %
2004	49 465	46 122	51 604	81,7	76,2	85,3	-7,2 %
moyenne	60 843	57 379	63 019	103,0	97,2	106,7	-3,7 %

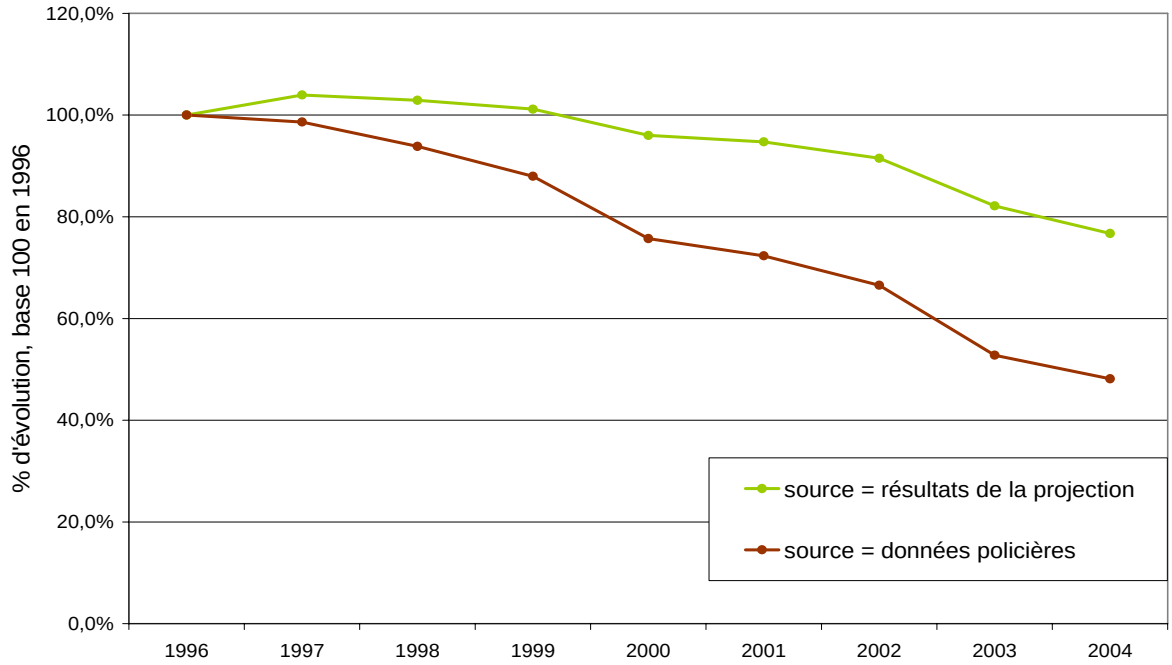
Le graphique ci-dessous montre les ordres de grandeur des effectifs de blessés graves selon la source des données.

Figure 26 : nombre de blessés graves, France, selon la source des données



Le graphique ci-dessous donne les évolutions en base 100 selon la source des données.

Figure 27 : évolution du nombre de blessés graves, France, selon la source des données

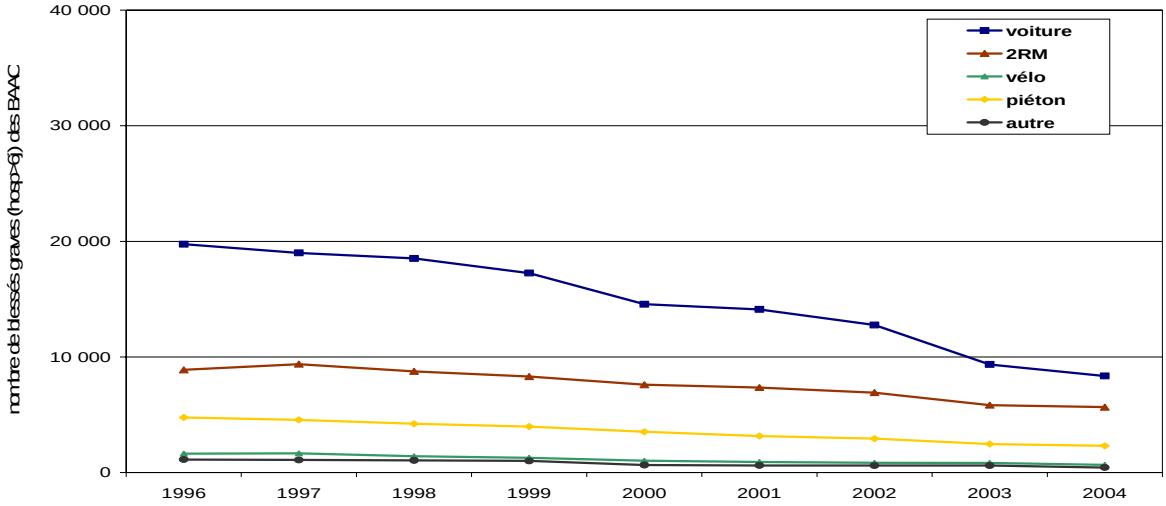


On note, si le nombre de blessés graves est bien en baisse, que la baisse est bien moindre selon les résultats de la projection que selon les données des forces de l'ordre.

Blessés graves, selon le type d'usager

Les blessés graves selon les données des forces de l'ordre (et selon leur définition) sont, en moyenne annuelle, arrondis à la centaine, au nombre de 1 100 cyclistes, 3 600 piétons, 7 600 usagers de 2RM, 14 900 automobilistes, et 800 autres usagers (effectifs et incidences annuels en annexe). Le graphique ci-dessous illustre leur ampleur et leur classement selon le type d'usager, d'après les données des forces de l'ordre.

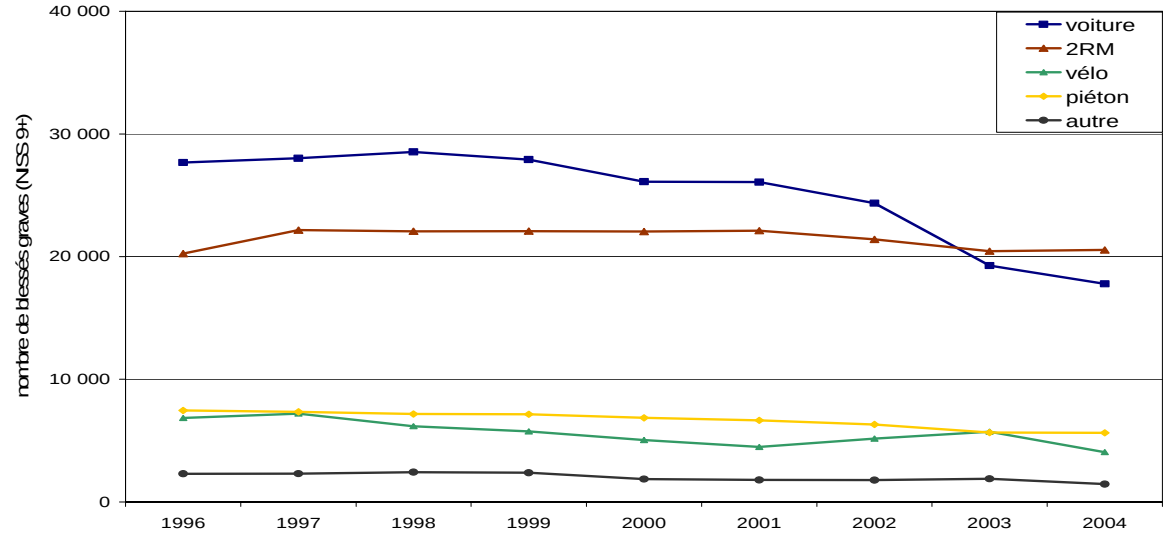
Figure 28 : nombre de blessés graves (nécessitant plus de 6 jours d'hospitalisation), selon le type d'usager, France, source = données des forces de l'ordre



Selon les données des forces de l'ordre, les cyclistes et les usagers autres sont au même niveau, en nombre de blessés graves.
Les incidences de blessés graves baissent de -4,8 % pour les piétons, -6,1 % pour les cyclistes, -1,9 % pour les usagers de 2RM, -6,8 % pour les automobilistes et -7,2 % pour les autres usagers.

D'après la projection, le nombre de blessés graves (au sens NISS 9+) est en moyenne annuelle de 5 600 cyclistes, 6 700 piétons, 21 500 usagers de 2RM, 25 100 automobilistes, et 2 000 autres usagers (estimations annuelles, effectifs et incidences en annexe).

Figure 29 : nombre de blessés graves (NISS 9+), selon le type d'usager, France, source = résultats de la projection



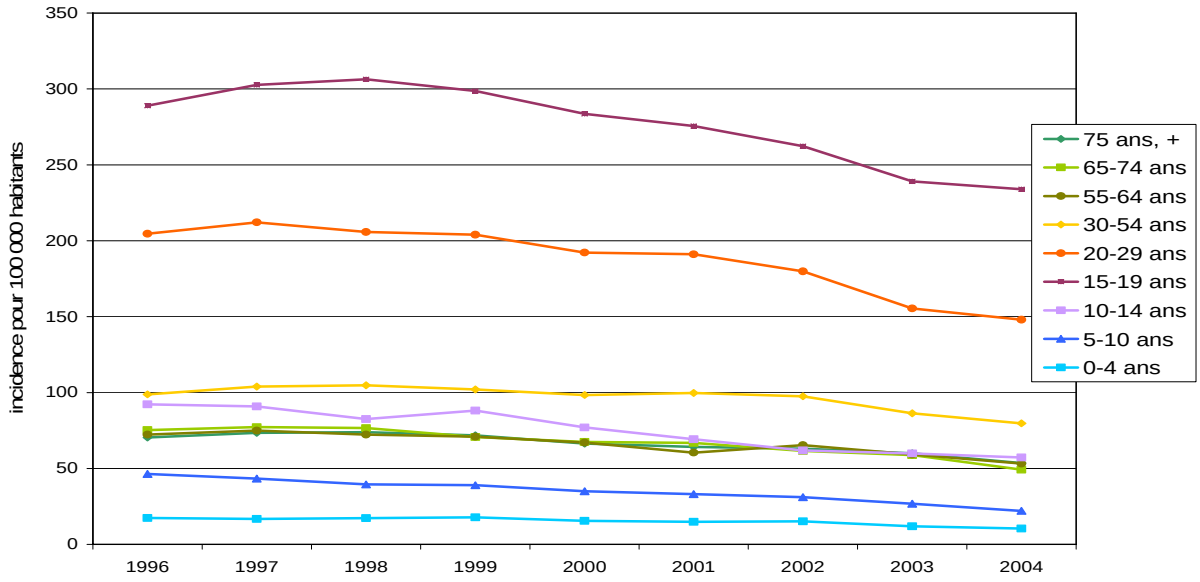
D'après la projection, les nombres de blessés graves cyclistes et de blessés graves piétons sont du même ordre de grandeur, autour de 6 000 - 7 000. Le nombre de blessés graves usagers de 2RM et automobilistes sont aussi du même ordre de grandeur, entre 20 000 et 30 000.
La courbe des blessés graves usagers de 2RM croise en 2003 celle des blessés graves automobilistes : en 2004 le nombre de blessés graves usagers de 2RM est supérieur à celui des blessés graves automobilistes. L'incidence de blessés graves baisse de -5,6 % chez les automobilistes et seulement de -0,3 % chez les usagers de 2RM (en moyenne annuelle). Elle baisse de -3,9 % pour les piétons, de -5,8 % pour les cyclistes, et de -5,4 % pour les autres usagers. Les baisses sont de bien moindre ampleur que celles affichées par les données des forces de l'ordre.

Blessés graves selon l'âge

Les nombres de blessés graves par classes d'âge selon les données des forces de l'ordre sont donnés en annexe ; le classement des classes d'âge selon l'incidence de blessés graves est le même quelle que soit la source des données. Les ordres de grandeur diffèrent.

Le graphique ci-dessous donne l'ampleur des incidences de blessés graves selon l'âge, d'après les résultats de la projection.

Figure 30 : incidence de blessés graves (NISS 9+) selon l'âge, France, source = résultats de la projection



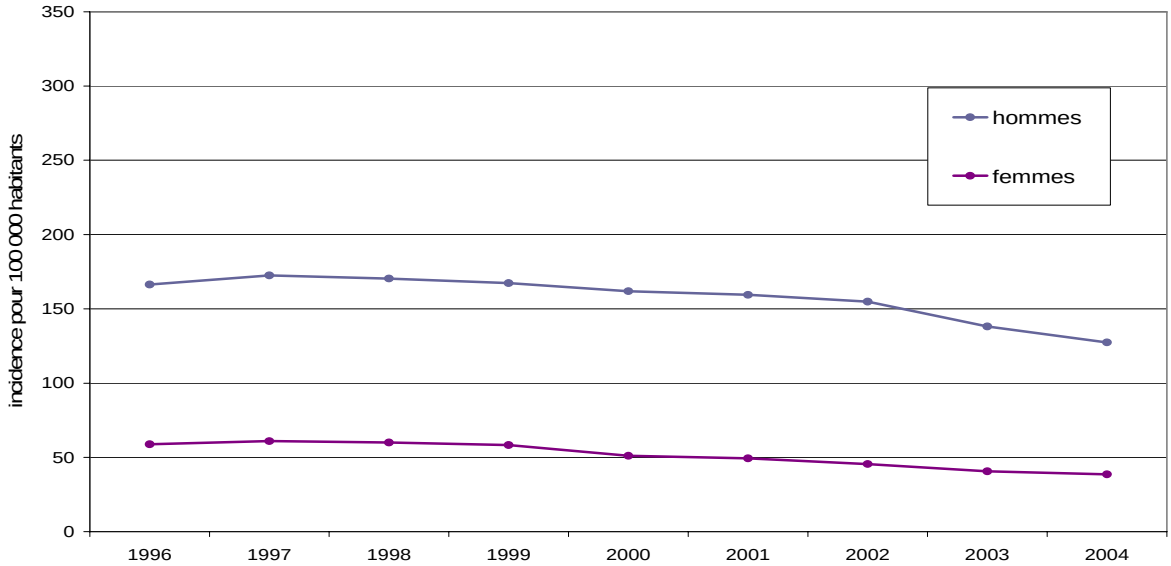
Les 20-29 ans et surtout les 15-19 ans ont les plus fortes incidences de blessés graves, respectivement de 188/100 000 et 277/100 000 en moyenne sur la période.
Les évolutions moyennes annuelles sont de -6,0 % chez les 0-4 ans, -8,4 % chez les 5-9 ans, -5,0 % chez les 10-14 ans, -2,2 % chez les 15-19 ans, -3,6 % chez les 20-29 ans, -1,5 % chez les 30-54 ans, -2,9 % chez les 55-64 ans, -4,6 % chez les 65-74 ans, et -3,9 % chez les 75 ans et plus. Les baisses sont assez faibles chez les plus touchés (les 15-19 et 20-29 ans) ; elles sont les plus fortes chez les moins touchés : les 0-4 ans et 5-10 ans. Les baisses sont moindres que celles basées sur les données des forces de l'ordre. Les estimations annuelles sont données en annexe.

Blessés graves selon le sexe

Les nombres de blessés graves par sexe selon les données des forces de l'ordre sont détaillés en annexe. Le classement hommes/femmes selon l'incidence de blessés graves est le même quelle que soit la source des données. Les ordres de grandeur diffèrent.

Le graphique ci-dessous donne l'ampleur de l'incidence de blessés graves chez les hommes et chez les femmes. Le ratio homme/femme est de 3,1 en moyenne sur la période.

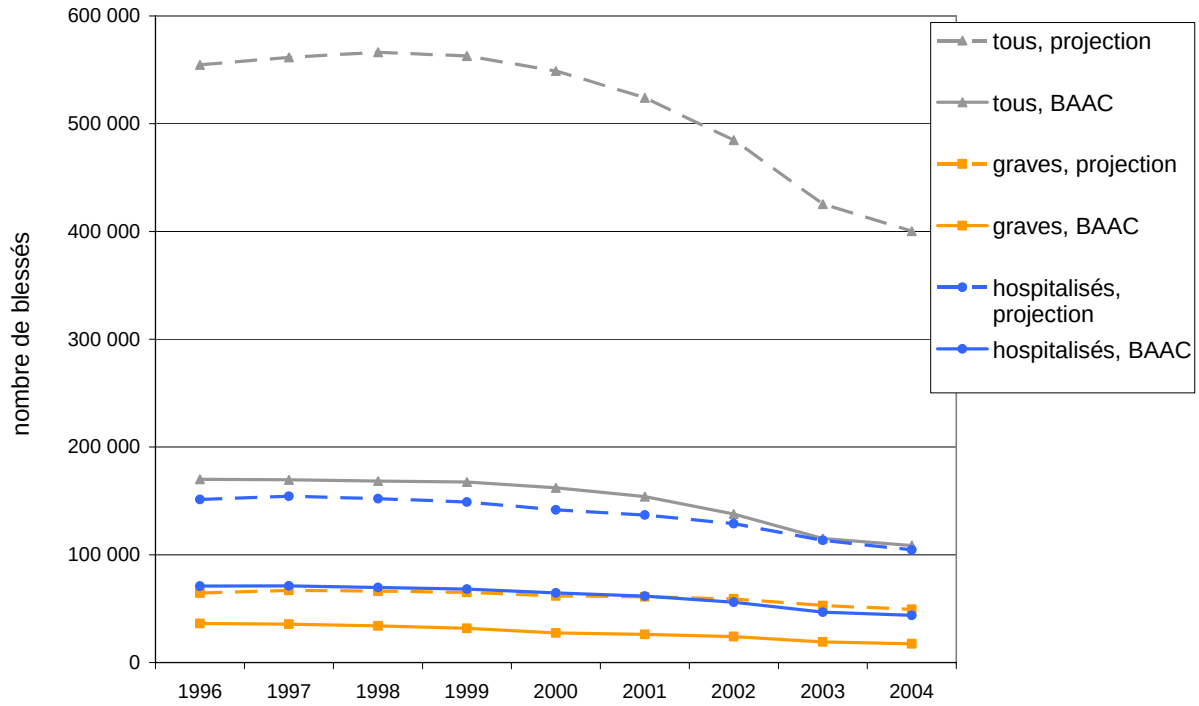
Figure 31 : incidence de blessés graves (NISS 9+) selon le sexe, France, source = résultats de la projection



L'incidence baisse en moyenne de -3,2 % chez les hommes et de -5,0 % chez les femmes par an sur 1996-2004. Ces baisses sont moindres que celles basées sur les données des forces de l'ordre.

Le graphique ci-dessous résume les résultats pour les blessés toutes gravités, les blessés hospitalisés et les blessés graves.

Figure 32 : nombre de blessés (tous, blessés graves et blessés hospitalisés), France, selon la source des données



Ce graphique montre clairement les différences d'ordre de grandeur selon la source des données : des forces de l'ordre ou résultats de la projection.
Quel que soit le critère étudié : blessés toutes gravités, blessés graves, ou hospitalisés, les évolutions sont à la baisse sur la période 1996-2004 et la baisse des incidences estimées par projection est moindre que la baisse des incidences basées sur les données des forces de l'ordre. L'écart est particulièrement grand pour les évolutions d'incidence de blessés graves.

Taux de gravité (chez les blessés non décédés)

Rappelons qu'il s'agit du ratio entre le nombre de blessés graves et le nombre total de blessés. Il est donc sensible aux nouvelles estimations du nombre de blessés graves et du nombre de blessés légers.
D'après les données des forces de l'ordre, ce taux de gravité est en moyenne de 18,4 %. Les taux annuels sont donnés dans le tableau ci-dessous.

Tableau 46 : taux de gravité (basé sur les hospitalisés plus de 6 jours), France, source = données des forces de l'ordre

année	Taux de gravité
1996	21,3 %
1997	21,1 %
1998	20,2 %
1999	19,0 %
2000	16,9 %
2001	17,0 %
2002	17,5 %
2003	16,6 %
2004	16,0 %
moyenne	18,4 %

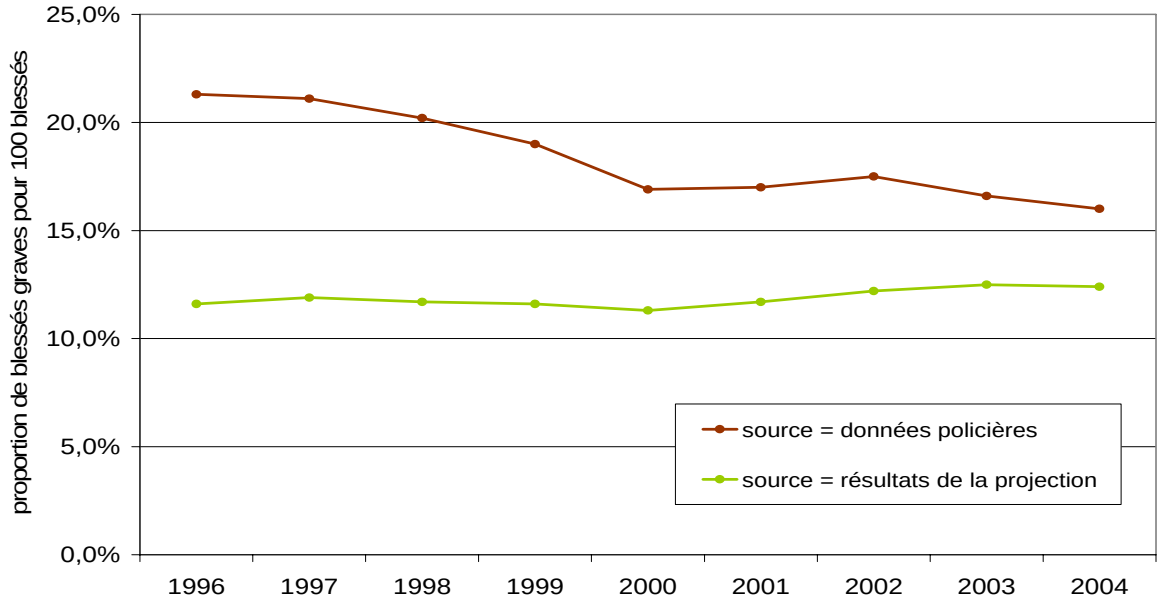
Le taux de gravité basé sur les effectifs obtenus par projection est en moyenne de 12 blessés graves parmi 100 blessés. Il est plus bas que celui basé sur les données des forces de l'ordre. Cela vient du fait que le nombre de blessés légers est corrigé par un facteur multiplicatif plus grand que le nombre de blessés graves.
Les taux annuels de gravité sont donnés dans le tableau ci-dessous.

Tableau 47 : taux de gravité (basé sur le NISS 9+), France, source = résultats de la projection

année	Taux de gravité	IC à 95 %	
1996	11,6 %	10,9 %	12,3 %
1997	11,9 %	11,3 %	12,6 %
1998	11,7 %	11,1 %	12,3 %
1999	11,6 %	11,1 %	12,1 %
2000	11,3 %	10,8 %	11,7 %
2001	11,7 %	11,2 %	12,1 %
2002	12,2 %	11,6 %	12,7 %
2003	12,5 %	11,8 %	13,1 %
2004	12,4 %	11,7 %	13,1 %
moyenne	11,8 %	11,3 %	12,3 %

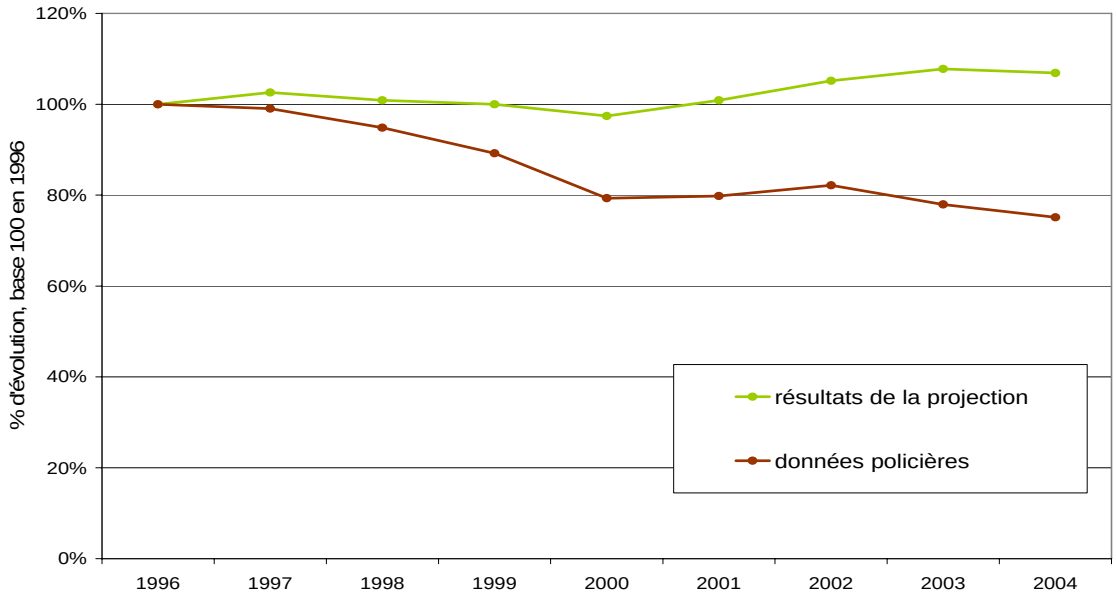
La figure ci-dessous illustre les ordres de grandeur du taux de gravité, selon la source des données, et leur évolution.

Figure 33 : taux de gravité, France, selon la source des données



Les évolutions en base 100 du taux de gravité selon la source des données sont données dans le graphique ci-dessous.

Figure 34 : évolution en base 100 du taux de gravité, France, selon la source des données

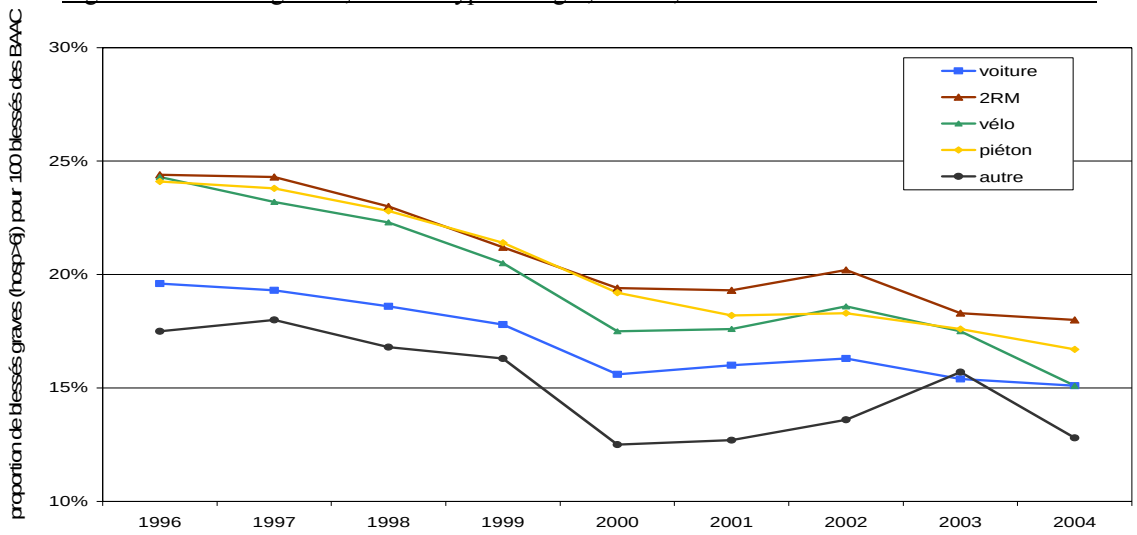


La gravité basée sur les données des forces de l'ordre (et leur définition) diminue, alors que la gravité basée sur les résultats de la projection est stable de 1996 à 2001 puis augmente légèrement.

Taux de gravité selon le type d'usager

NB : les échelles des deux graphiques ci-dessous ne sont pas les mêmes ; le but est d'avoir les mêmes ampleurs de variations.
La graphique ci-dessous donne les taux de gravité par type d'usager selon les données des forces de l'ordre.

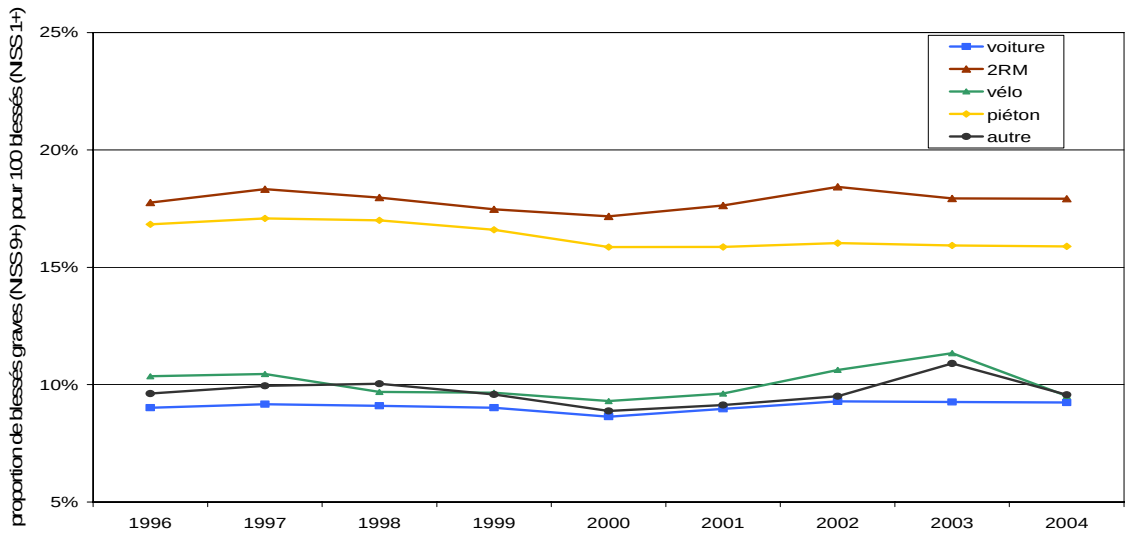
Figure 35 : taux de gravité, selon le type d'usager, France, source = données des forces de l'ordre



Selon les données des forces de l'ordre, la gravité est la plus élevée pour les usagers de 2RM, piétons et cyclistes. L'évolution de la gravité est à la baisse pour tous les types d'usagers.

Le graphique ci-dessous donne les taux de gravité par type d'usager selon les résultats des la projection.

Figure 36 : taux de gravité, selon le type d'usager, France, source = résultats de la projection



Selon la projection, les usagers de 2RM et les piétons ressortent avec un taux de gravité nettement plus élevé que les autres usagers. Cyclistes, automobilistes et autres usagers ont un taux de gravité bien moindre et équivalent.
Le taux de gravité des cyclistes est nettement rabaissé car le dénominateur (ensemble des blessés) est fortement augmenté par la correction du sous-enregistrement (bien plus que pour les autres usagers).
Les taux de gravité basés sur les résultats de la projection semblent stables sur la période 1996-2004 pour tous les types d'usagers.

Blessés avec séquelles majeures prévues

Effectifs et incidences

Nous rappelons que nous ne présentons pas d'estimation annuelle pour les effectifs de victimes avec séquelles majeures, car les évolutions observées sur le département du Rhône nous paraissent trop erratiques pour être projetées sur la France entière.
Les effectifs de blessés avec séquelles majeures sont mis en regard, dans toute cette partie, avec les effectifs de tués : il s'agit de comparer les ordres de grandeur. Les effectifs de tués sont fournis par les données des forces de l'ordre (nous rappelons que les tués sont relativement bien enregistrés).
Le nombre de victimes avec séquelles majeures est estimé à 7479 personnes par an en moyenne. L'incidence des victimes avec séquelles majeures est estimé à 12,6/100 000 par an. Le tableau ci-dessous met en regard les effectifs et incidences des tués à six jours avec ceux des blessés avec séquelles majeures, totaux, et par région corporelle atteinte.

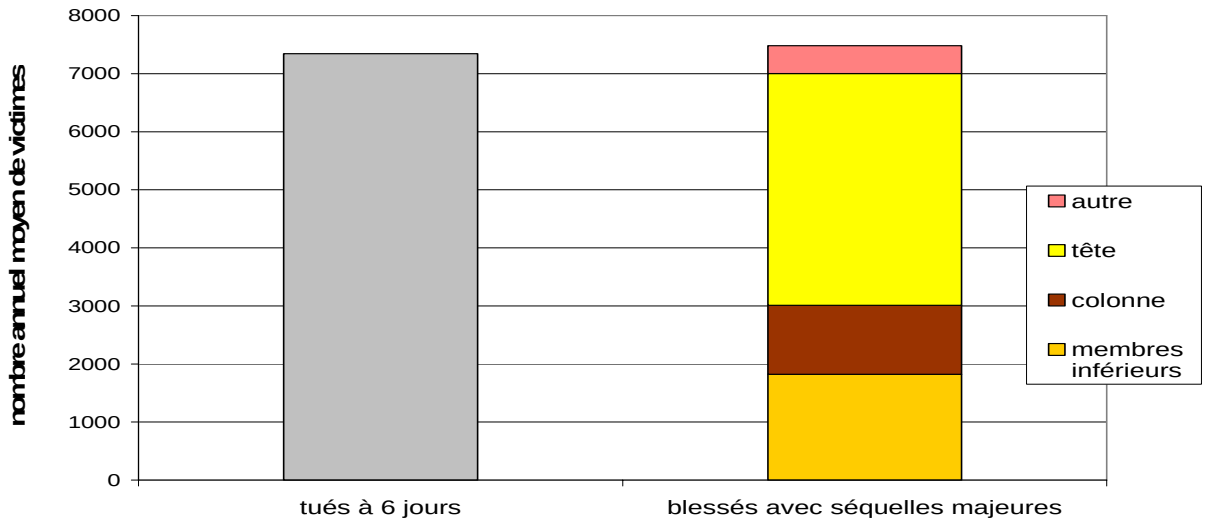
Tableau 48 : effectif et incidence, des tués et des blessés avec séquelles majeures, France, moyennes annuelles sur 1996-2004

source	critère de santé	effectif	IC à 95 %		incidence	IC à 95 %	
données policières	tués (à 6 jours)	7344	-	-	12,4	-	-
résultats de la projection	blessés avec séquelles majeures	7479	6554	8158	12,6	11,1	13,8
	région atteinte :						
	tête	3993	3390	4487	6,8	5,7	7,6
	colonne (moelle épinière)	1145	853	1438	1,9	1,4	2,4
	membres inférieurs	1821	1411	2173	3,1	2,4	3,7

Les effectifs de victimes avec séquelles majeures sont du même ordre de grandeur que les effectifs de tués ; ils sont même très proches.

Le graphique ci-dessous illustre la comparaison des tués à 6 jours et des blessés avec séquelles majeures.

Figure 37 : effectif de tués à six jours, et de blessés avec séquelles majeures (résultats de la projection), France, moyenne annuelle sur 1996-2004,



La répartition des blessés avec séquelles majeures selon la région corporelle atteinte indique une très forte proportion d'atteintes à la tête (53 %), suivie par les atteintes aux membres inférieurs (24 %), puis à la colonne (16 %).

Blessés avec séquelles majeures, selon le type d'usager

Les tableaux ci-dessous donnent pour comparaison les effectifs de tués à six jours selon le type d'usager, et les effectifs de blessés avec séquelles majeures, selon le type d'usager.

Tableau 49 : effectif et incidence de tués selon le type d'usager, moyennes annuelles sur 1996-2004, France, source = données des forces de l'ordre

	effectif	incidence
tués (à 6 jours)	7344	12,4
piétons	811	1,4
cyclistes	256	0,4
usagers de 2RM	1291	2,2
automobilistes	4722	8,0
autres	264	0,4

Tableau 50 : effectif de blessés avec séquelles majeures, selon le type d'usager, moyennes annuelles sur 1996-2004, France, source = résultats de la projection

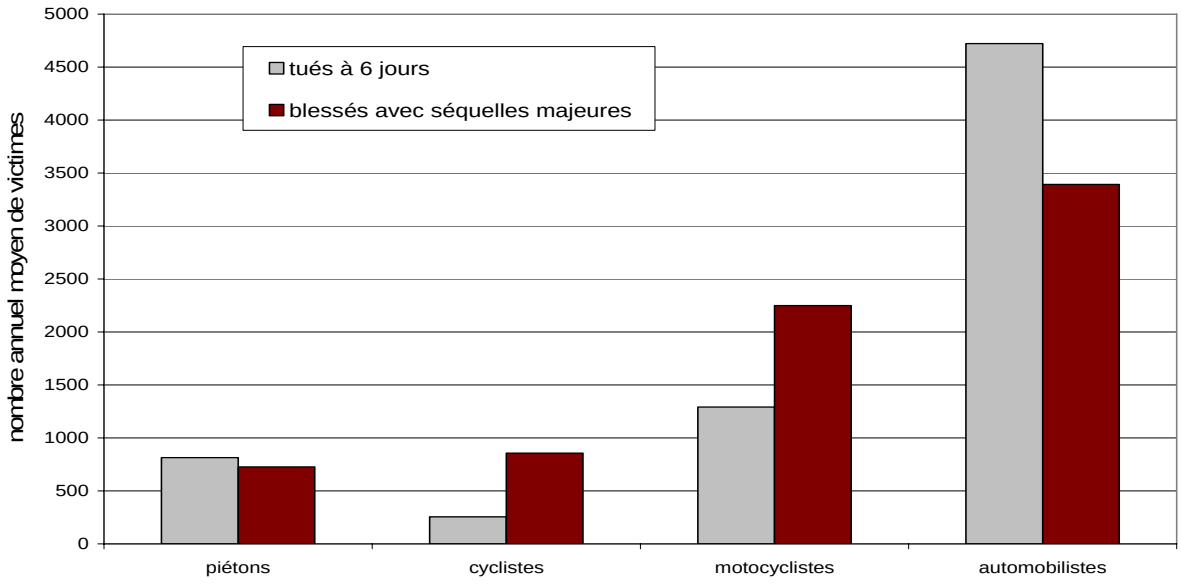
	effectif	IC à 95%		incidence	IC à 95%	
blessés avec séquelles majeures	7479	6554	8158	12,6	11,1	13,8
piétons	726	578	875	1,2	1,0	1,5
cyclistes	857	504	1207	1,4	0,9	2,0
usagers de 2RM	2249	1838	2568	3,8	3,1	4,3
automobilistes	3393	2869	3811	5,7	4,9	6,4
autres*	255	115	391	0,4	0,2	0,7

* atteintes IIS3+ à la colonne vertébrale non estimables (car non observées dans le Rhône)

Les répartitions des types d'usager sont différentes selon que l'on s'intéresse aux tués à six jours ou aux blessés avec séquelles majeures. Alors que le nombre de cyclistes tués est bien en deçà du nombre de piétons tués, le nombre de cyclistes avec séquelles majeures est du même ordre de grandeur que celui des piétons. Les usagers de 2RM tués sont trois fois moins nombreux que les automobilistes tués (en moyenne chaque année) mais en termes de blessés avec séquelles majeures, les usagers de 2RM ne sont plus que 1,5 fois moins nombreux que les automobilistes.

L'histogramme ci-après donne par type d'usager la comparaison entre nombre de tués et nombre de blessés avec séquelles majeures.

Figure 38 : effectifs de tués et de blessés avec séquelles majeures, selon le type d'usager, France, moyenne annuelle sur 1996-2004

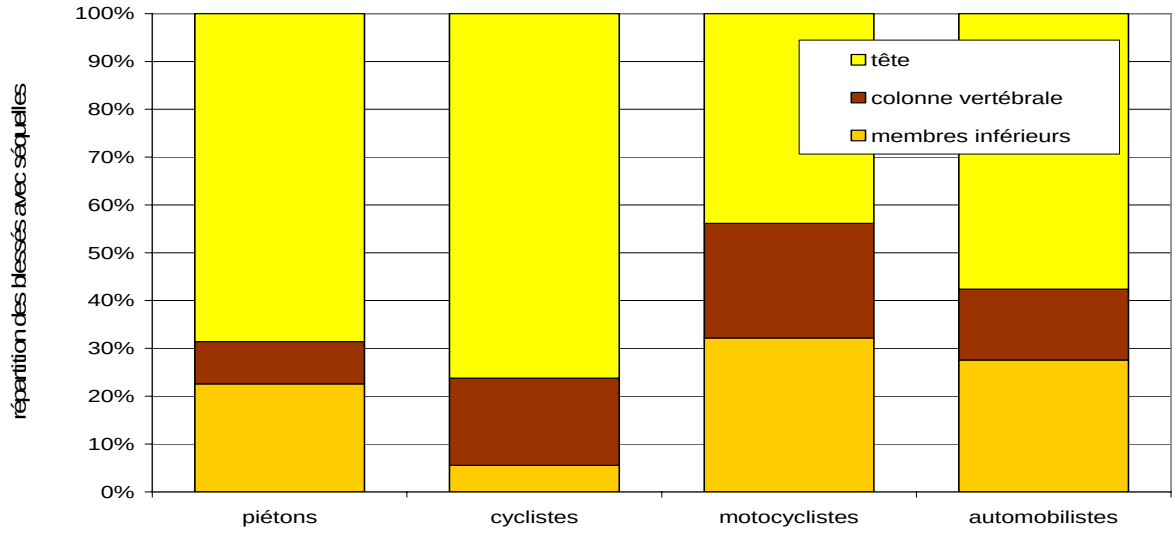


Les cyclistes et les usagers de 2RM se retrouvent bien plus souvent lourdement handicapés que tués suite à un accident de la circulation routière (respectivement 3,3 et 1,7 fois plus souvent). Les piétons ont un léger sur-risque d'être lourdement handicapés plutôt que tués (ratio=1,1). Pour les automobilistes, c'est le contraire : ils sont moins souvent lourdement handicapés que tués (ratio=0,7).

Blessés avec séquelles majeures, selon le type d'usager et la région corporelle

Le graphique ci-dessous donne la répartition corporelle selon le type d'usager.

Figure 39 : répartition des blessés avec séquelles majeures, selon la région corporelle atteinte, et le type d'usager, France, 1996-2004, source = résultats de la projection

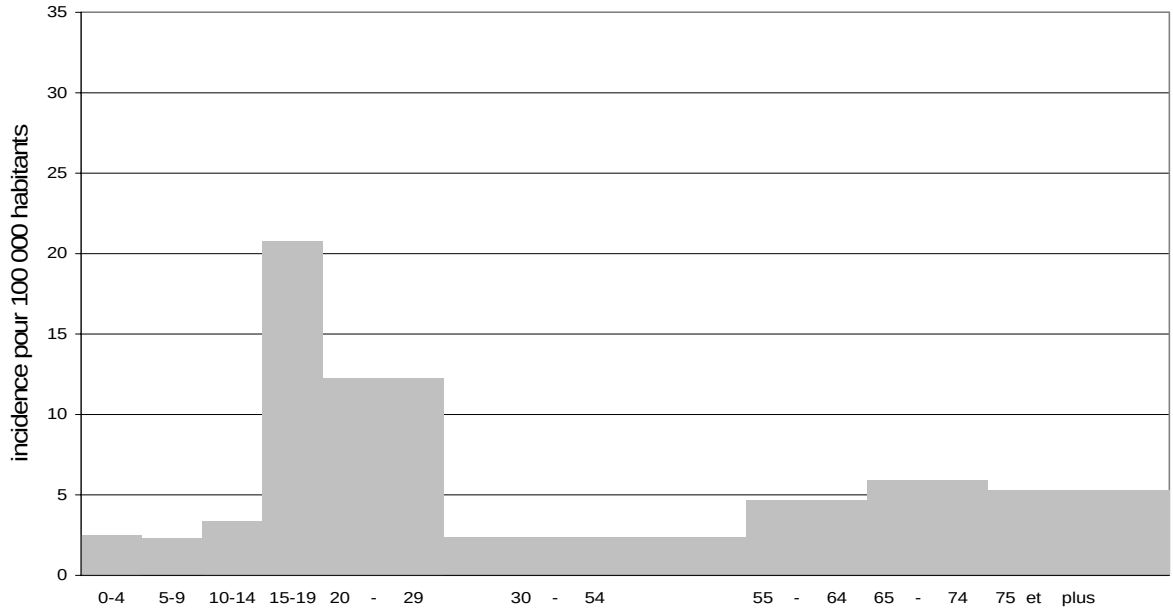


Parmi ceux qui garderont des séquelles, les usagers de 2RM sont particulièrement touchés à la colonne et aux jambes, comparativement aux autres usagers ; les cyclistes et les piétons sont particulièrement touchés à la tête. Toujours parmi ceux qui garderont des séquelles, les piétons sont relativement peu touchés à la colonne et les cyclistes peu touchés aux membres inférieurs.

Blessés avec séquelles majeures, selon l'âge

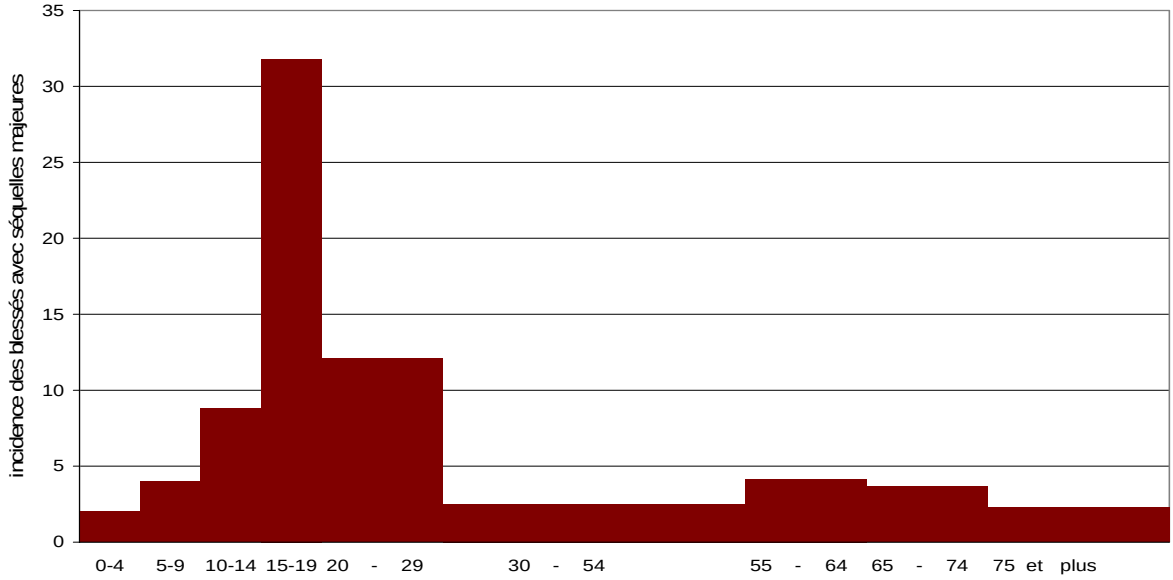
Pour base de comparaison, nous présentons l'incidence des tués par classe d'âge.

Figure 40 : incidence de tués à six jours, selon l'âge, France, moyenne annuelle sur 1996-2004, source = données des forces de l'ordre



La figure ci-dessous donne l'incidence des blessés avec séquelles majeures, selon les classes d'âge.

Figure 41 : incidence de blessés avec séquelles majeures, selon l'âge, France, moyenne annuelle sur 1996-2004, source = résultats de la projection

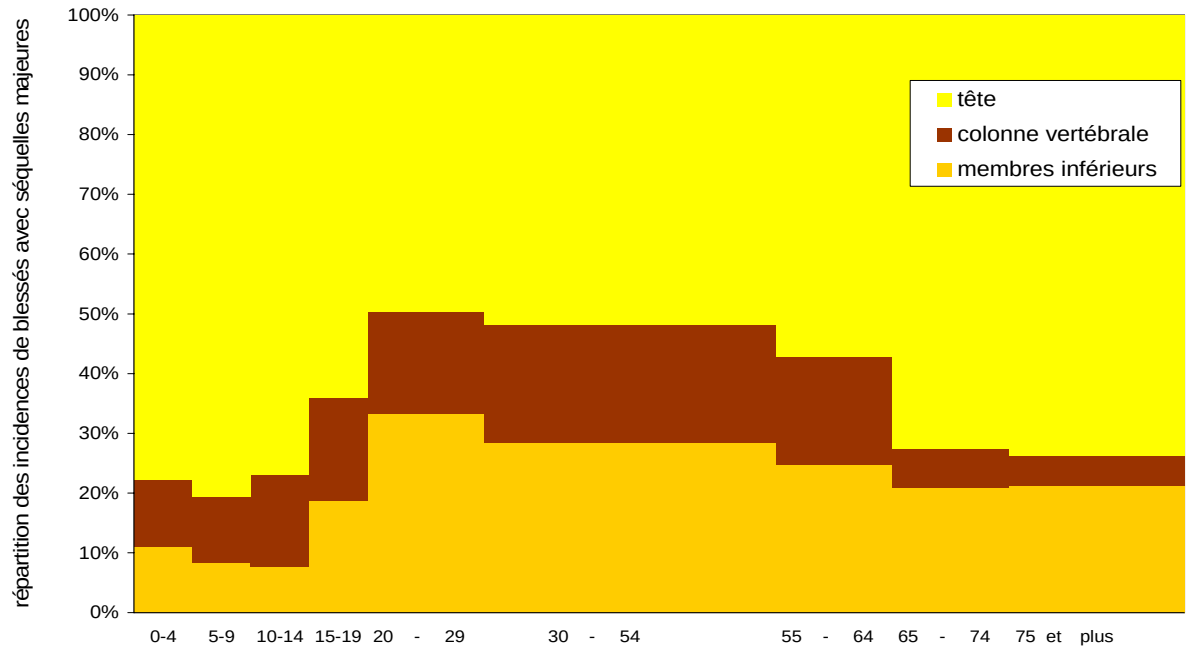


L'incidence de blessés avec séquelles majeures est particulièrement élevée chez les 15-19 ans ; elle atteint 31,8/100 000. Pour les classes d'âge 5-9 ans, 10-14 et 15-19, l'incidence de blessés avec séquelles majeures est plus élevée que celle des tués à 6 jours. Pour les personnes âgées de 65 ans et plus, c'est l'inverse : le risque d'être tué dans un accident de la route est plus élevé que celui d'être lourdement handicapé.

Blessés avec séquelles majeures selon l'âge et la région corporelle

Le graphique ci-dessous donne la répartition entre les trois régions corporelles les plus fréquemment touchées, par classe d'âge.

Figure 42 : répartition des blessés avec séquelles majeures, selon la région corporelle et l'âge, France, 1996-2004, source = résultats de la projection

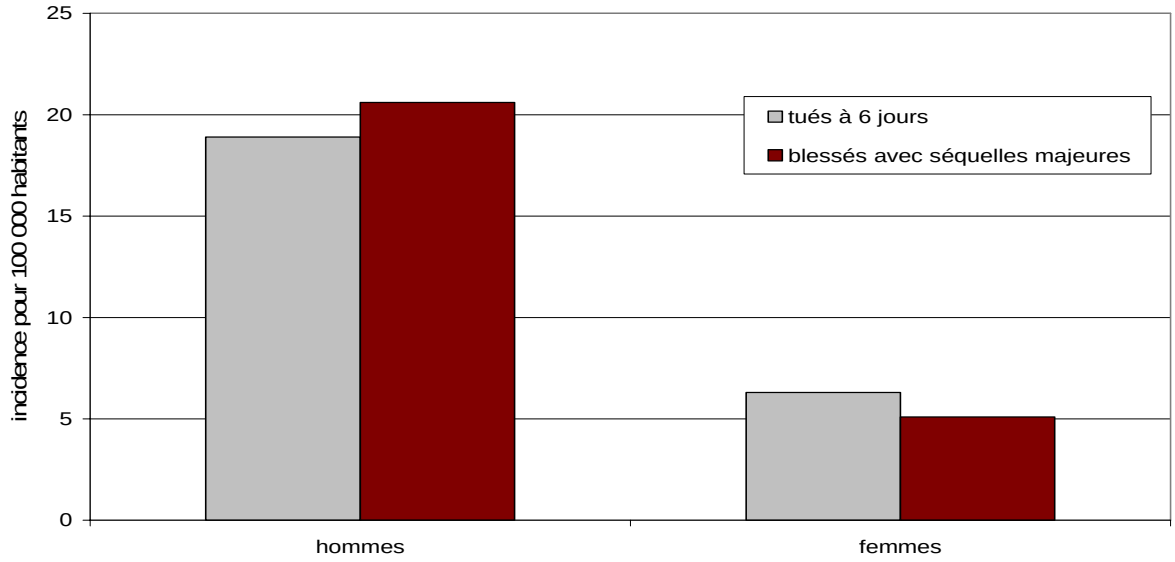


Parmi ceux qui garderont des séquelles, ceux qui sont le plus concernés par des traumatismes crâniens sont les enfants (0-14 ans) et les âgés (65 ans et plus) ; les plus concernés par des traumatismes médullaires sont les 15-64 ans; les moins concernés sont les plus de 65 ans.

Blessés avec séquelles majeures, selon le sexe

Le graphique ci-dessous donne la comparaison des tués à six jours et des blessés avec séquelles majeures par sexe.

Figure 43 : incidence de blessés avec séquelles majeures, selon le sexe,
France, moyenne annuelle sur 1996-2004, source = résultats de la projection



Cela montre d'une part l'ampleur des atteintes IIS3+ chez les hommes comparativement aux femmes (sex ratio=4,0), d'autre part que les hommes ont légèrement plus de risque d'être lourdement handicapés que tués, alors que c'est le contraire chez les femmes.

Blessés avec séquelles majeures, selon le sexe et la région corporelle

Le graphique ci-dessous donne les répartitions corporelles des atteintes IIS3+ selon le sexe.

Figure 44 : répartition des blessés avec séquelles majeures, selon la région corporelle, et le sexe,
France, 1996-2004, source = résultats de la projection

La répartition des régions corporelles atteintes par des lésions IIS3+ n'est pas la même selon le sexe : le risque d'atteinte à la colonne vertébrale est assez élevé chez les hommes, relativement aux autres régions corporelles.

Discussion

Forces et faiblesses

Application de la méthode de capture-recapture

La méthode de capture-recapture est basée sur quatre conditions clés (Hook and Regal 1995; IWGDMF 1995; Gallay, Nardone et al. 2002) et deux conditions implicites (Gallay, Nardone et al. 2002). Une première condition est celle d' « homogénéité de capture », autrement dit de probabilité homogène entre les individus d'être enregistrés par une source donnée. Si cette homogénéité n'est vérifiée qu'à l'intérieur de sous-groupes, il faut stratifier sur la caractéristique définissant ces sous-groupes ou l'inclure comme variable d'ajustement dans un modèle. La probabilité d'être enregistré dans le fichier BAAC dépend essentiellement de la gravité lésionnelle, du type d'usager, de la présence / absence de tiers dans l'accident, du type de force de l'ordre, du type de réseau, de la zone urbaine / rurale (facteurs de biais de sélection). La probabilité d'être enregistré dans le Registre dépend de la gravité lésionnelle, de la distance à l'hôpital et éventuellement du type d'usager. Toutes ces caractéristiques à l'exception de rural/urbain et de la distance à l'hôpital ont été incluses dans le modèle. Cette condition de l'approche capture-recapture est donc largement prise en compte.

Une deuxième condition est celle d'indépendance des deux sources. La dépendance entre les deux sources implique que l'estimateur du nombre total de sujets est biaisé. Il a été démontré (IWGDMF 1995) qu'il y a sur-estimation si la dépendance est négative, et sous-estimation si la dépendance est positive. Ici, il y a une dépendance positive pour les accidents graves. Les estimateurs des effectifs de blessés dans le Rhône sont donc biaisés à la baisse (sous-estimation) ; autrement dit, les estimateurs des coefficients de correction sont biaisés à la baisse et donc les estimateurs des effectifs et incidences au niveau national le sont aussi. Les résultats obtenus sont donc vraisemblablement des sous-estimations.

Une troisième condition est celle de l'identification parfaite des individus communs aux deux sources. La procédure de chaînage utilisée en routine est basée sur un nombre certain de variables : la date (et l'heure) de l'accident, le lieu de l'accident (la commune et éventuellement le nom de la rue ou de la route), le type d'usager, le sexe, l'année et le mois de naissance de la victime. Ces variables sont informatives et la procédure de chaînage est donc plutôt de bonne qualité. De plus, le résultat du chaînage a été amélioré par l'estimation et la prise en compte du nombre de faux négatifs et de faux positifs. Autrement dit, les sous-ensembles de vrais positifs et de vrais négatifs ont été mieux estimés. Nous nous sommes donc approchés le plus possible de la condition d'identification parfaite des sujets communs aux deux sources.

La quatrième condition de la méthode de capture-recapture est celle d'une population fermée, c'est-à-dire l'absence d'entrée ou de sortie entre les enregistrements dans les deux sources. Une entrée entre les deux enregistrements signifierait une victime accidentée en dehors du Rhône venant consulter un service hospitalier du Rhône. Elle ne sera normalement pas incluse dans le Registre, car un des critères d'inclusion est justement « accidenté dans le Rhône » ; cela a en outre été contrôlé a posteriori. Une sortie entre les deux enregistrements signifierait qu'une personne enregistrée par les forces de l'ordre comme blessée dans le Rhône irait consulter un service hospitalier loin du département du Rhône. Cela est peu probable pour les blessés graves, et ne concerne que les blessés légers qui pourraient préférer consulter près de leur lieu d'habitation, hors du Rhône. Or la majorité des accidents sont des accidents de proximité. Il pourrait manquer de cette façon environ 8 % des blessés légers (NISS 1-3) (cf. chap. 3).

Une des conditions souvent implicite de l'approche capture-recapture est celle de même période de temps et de même zone géographique des deux sources d'enregistrement. La période de temps commune est une date d'accident sur la période 1996-2004. La zone géographique commune est le

département du Rhône comme lieu d'accident. Certains blessés du Registre s'étant avérés accidentés en dehors du Rhône, nous avons exclu du Registre la proportion de blessés correspondant à cette situation.

La deuxième condition implicite de la méthode de capture-recapture est l'absence d'erreur de classement des cas. Nous avons exploré si les deux critères « blessé » et « accident de la circulation routière » étaient vérifiés. Nous avons exclu des données des forces de l'ordre les personnes qui étaient probablement non-blessées (mais hospitalisées pour observation par exemple). Nous avons aussi exclu du Registre une faible proportion de personnes dont l'accident était d'une nature autre que routière.
Nous rappelons que les victimes d'accident de roller/skateboard/trottinette seul ou contre un piéton ont déjà été exclues lors de l'application de la définition officielle de l'accident de la circulation routière. De façon analogue, les victimes d'accidents clairement identifiés comme hors réseau (jardin autour de la maison, cour d'école…) ont été exclues des données du Registre.

Les conditions de la méthode de capture-recapture sont donc en grande partie vérifiées ; la condition d'indépendance n'est pas vérifiée, et cela nous conduit à considérer les résultats d'effectifs et d'incidences comme des valeurs plancher.

Prédiction du critère de santé chez les blessés enregistrés dans les données des forces de l'ordre

Concernant l'étape de prédiction du NISS chez les blessés enregistrés dans les BAAC, nous avons vu (étape 1 de ce chapitre) qu'il y a une éventuelle sur-estimation de la proportion de blessés graves (NISS 9+). Nous rappelons que le modèle de prédiction est construit sur les blessés identifiés comme communs au Registre et aux données des forces de l'ordre, pour lesquels nous disposons du NISS consolidé (dans le Registre), et des variables des BAAC (notamment le critère de gravité des BAAC) les seules disponibles comme variables potentiellement prédictives. Ces blessés sont sans doute un peu plus gravement blessés en moyenne que les blessés identifiés dans les BAAC seulement (biais de sélection du Registre sur la gravité). Cela a été confirmé sur un sous-ensemble des blessés accidentés en 2001 (cf. chap. 3). Ce sous-ensemble ne peut cependant pas être utilisé pour la prédiction car sa représentativité n'est pas assurée. Il comporte une sur-représentation de procès-verbaux issus des CRS ; or la force de l'ordre est liée au type de réseau et au caractère urbain/rural de l'environnement.
Les conséquences d'une éventuelle sur-estimation de la proportion de blessés graves sur le bilan national seraient une légère sur-estimation du nombre de blessés graves, mais surtout une sous-estimation non-négligeable du nombre de blessés légers (NISS 1-3) à cause de l'effet amplificateur des coefficients de correction, bien plus élevés pour les blessés légers que pour les blessés modérés et graves.

Il faut aussi rappeler que la prédiction du NISS chez les blessés enregistrés dans les données des forces de l'ordre est basée sur l'hypothèse qu'à configuration égale d'accident, les conséquences moyennes en terme de gravité lésionnelle sont les mêmes (sécurité secondaire), et cela, quel que soit le lieu (département, région) où l'accident s'est produit. Étant donné que cette prédiction se base en grande partie sur le critère de gravité des forces de l'ordre, l'hypothèse sous-jacente est celle d'homogénéité sur le territoire national des pratiques des forces de l'ordre de classement de la gravité, par type de force de l'ordre.

Nous rappelons que des modèles de prédiction ont aussi été construits pour les autres critères de santé : hospitalisés, blessés avec séquelles majeures, quelle que la soit la région corporelle atteinte, et par région atteinte (tête, colonne vertébrale, membres inférieurs). L'étude de l'adéquation des modèles aux données est présentée en annexe. En résumé, l'adéquation est plutôt bonne.

Éléments de validation

Les résultats obtenus ont été comparés à des estimations extérieures, dès qu'elles existaient. Une première comparaison porte sur le nombre de blessés hospitalisés : nous avons prédit le nombre de blessés hospitalisés parmi les BAAC (étape 1 du processus de projection), autrement dit, parmi le nombre total de blessés recensés par les forces de l'ordre, combien ont vraisemblablement été hospitalisés (compte tenu de leurs caractéristiques). Ainsi en 2004, nous estimons ce nombre à 43 779 (Tableau 42), et comme les autres critères de santé étudiés, il s'inscrit dans une évolution à la baisse. Par ailleurs, depuis 2005, les forces de l'ordre décomptent les blessés hospitalisés ; ainsi en 2005 le bilan annuel de l'ONISR (ONISR 2005) recense 39 811 blessés hospitalisés. Compte tenu du contexte d'évolution à la baisse, notre estimation est tout à fait cohérente avec ce nombre observé. Cela apporte un élément de validation à la première étape du processus de projection, celle de prédiction, parmi les blessés recensés par les forces de l'ordre, de leur état de santé, ici du statut d'hospitalisé. Les prédictions selon les autres critères de santé étudiés (NISS 1-3, 4-8, 9+, victimes avec séquelle majeure) sont basées sur la même méthode.

Une deuxième comparaison porte sur le nombre incident de victimes avec traumatisme médullaire. D'après la caisse nationale de l'assurance maladie, il y a eu 2343 nouveaux cas de paraplégies (IIS 4 et 5) en moyenne annuelle sur 1996-2004 (il s'agit des nouvelles prises en charge en affection de longue durée (CNAM 2007)). Par ailleurs, des chercheurs (Saillant, Pascal-Moussellard et al. 2005) estiment, en appliquant les taux d'incidence d'autres pays industrialisés, États-Unis essentiellement, qu'il y a environ 2000 cas annuels de traumatismes médullaires. Il est aussi estimé que 60-70 % (ibid.) de ces traumatismes sont dus à des accidents de la voie publique, qui sont essentiellement des accidents de la circulation (les autres étant des accidents de piétons seuls, comptabilisés comme accidents de la vie courante et ayant peu de risques d'aboutir à une lésion médullaire). Cela donne environ 1200-1400 traumatisés médullaires suite à un accident de la circulation routière chaque année. Par ailleurs, notre méthode de projection estime ce nombre à 1145 (IC à 95 % = 853-1438), moyenne annuelle sur 1996-2004 (Tableau 48). Ces effectifs sont très proches.

Une troisième comparaison concerne le nombre incident de blessés avec un traumatisme crânien grave. Sur la région Aquitaine, en 1996, il a été recensé 642 cas de traumatisés crâniens avec une lésion AIS 4 ou 5 à la tête, toutes causes, résidents et non-résidents de la région (Masson, Thicoipe et al. 1998; Masson, Thicoipe et al. 2001). Si cet effectif est restreint aux accidents de la circulation (environ 50 %, ibid), et que les décès quasi-immédiats sont exclus (au moins 18 %), l'effectif est alors de 265 (détail en annexe). Par ailleurs, nous appliquons la procédure de projection en nous restreignant à la seule région Aquitaine, c'est-à-dire nous appliquons les coefficients de correction (estimés sur le Rhône) aux données BAAC de la région Aquitaine (seule l'étape 4 est différente). Nous obtenons un effectif de 244 traumatisés crâniens graves (moyenne annuelle sur 1996-2004, IC à 95% = 203-278). Les deux effectifs estimés sont très proches.

Hypothèse sous-jacente à la projection

La projection est basée sur l'hypothèse d'homogénéité sur le territoire français des pratiques des forces de l'ordre. Celle-ci se fonde sur la structure centralisée et l'organisation centralisée des forces de l'ordre. Rappelons aussi que cette hypothèse porte sur chaque type de force de l'ordre séparément, c'est-à-dire que nous supposons l'homogénéité des pratiques par type de force de l'ordre.
Il semblerait cependant que les pratiques des forces de l'ordre puissent dépendre des consignes de procureurs (nous avons reçu quelques commentaires informels dans ce sens). Nous avons-nous-mêmes observé des différences dans les pratiques des forces de l'ordre. En effet, le taux de valeurs manquantes de certaines variables, notamment les variables « sensibles » comme le port de la ceinture ou le contrôle de l'alcoolémie varie d'une région à l'autre, ou d'un département à l'autre, avec des taux élevés notamment en régions PACA et Aquitaine. Nous avons exclu ce type de variables de notre

procédure d'estimation afin que les variations régionales des pratiques des forces de l'ordre en termes de tolérance vis-à-vis de certains comportements ne viennent pas gêner notre modélisation (rappelons que ces variables n'auraient été utiles que dans l'étape de prédiction du NISS). Ce que nous ne pouvons pas prendre en compte c'est une éventuelle variation du taux de sous-enregistrement, et une éventuelle variation dans le sous-classement de la gravité des blessés, entre les départements ou les parquets (dépendants des procureurs).
Selon un rapport sur les variations spatio-temporelles des accidents (Orselli 2003), il y a « des différences très importantes dans les pratiques de recensement des blessés légers par les forces de l'ordre ». Un graphique (diagramme 6, p.119, chap. 4) compare les départements les taux d'enregistrement, en distribuant les taux de blessés légers par million d'habitants (moyenne sur 1995-2001) en fonction de la densité de population du département ; celle-ci étant très lié au taux de gravité. À densité égale, les taux de gravité variables reflèteraient des taux d'enregistrement variables. Le département du Rhône se trouve au milieu du nuage de points ; il constitue ainsi un département moyen (en terme de taux d'enregistrement), et nous pouvons ainsi penser que pour la France prise dans son ensemble, les estimations sont bonnes, mais que ce ne serait pas nécessairement le cas pour des estimations au niveau des départements. L'utilisation du seul département du Rhône comme base de toutes nos estimations est sans doute le point le plus faible de notre procédure d'estimation. L'extension du Registre à d'autres départements permettrait d'y remédier.

Des coefficients de correction du sous-enregistrement des données des forces de l'ordre ont été estimés dans d'autres pays, également par comparaison à des données hospitalières. En Norvège, des coefficients de correction ont été estimés ; ils sont définis en fonction de la gravité (en deux catégories), du type d'usager, de la présence de tiers (oui/non) et du type d'antagoniste (Hvoslef 1994). En Grande-Bretagne, ces facteurs ont été définis et estimés en fonction de la gravité (trois catégories) et du type d'usager (James 1991). Actuellement, le projet Européen SAFETYNET inclut un sous-groupe, WP1, Task 1.5 intitulé « estimation of the real number of road casualties » (auquel nous participons). Ce sous-groupe estime des coefficients de correction du sous-enregistrement policier par comparaison à des données hospitalières dans huit pays (Autriche, Espagne, France, Grande-Bretagne, Grèce, Hongrie, Pays-Bas, République Tchèque). Ces coefficients de correction sont définis en fonction de la gravité et du type d'usager (Broughton, Yannis et al. 2005). Le but est de fournir des estimations nationales corrigées du sous-enregistrement. Dans au moins quatre des pays participant à ce projet, l'estimation de coefficients de correction est faite à une échelle régionale avec l'idée d'appliquer ces estimations aux données nationales des forces de l'ordre, en standardisant sur gravité et type d'usager. Le principe de la projection est donc le même que le nôtre. En revanche, aucune de ces études n'utilise l'approche capture-recapture ; cela revient à supposer que l'ensemble formé par la réunion des enregistrements policiers et hospitaliers est exhaustif, qu'aucun blessé (même léger) n'échappe aux deux sources d'enregistrement.

La méthode de capture-recapture a été utilisée en traumatologie routière, mais les études se sont restreintes à des sous-populations : enfants (Roberts and Scragg 1994; Jarvis, Lowe et al. 2000), adolescents et jeunes adultes (Morrison and Stone 2000), piétons et cyclistes (Dhillon, Lightstone et al. 2001), conducteurs de poids lourds (Meuleners, Lee et al. 2006), grandes villes (Razzak and Luby 1998; Tercero and Andersson 2004) ou l'île de la Réunion (Aptel, Salmi et al. 1999).

Notre approche est la combinaison de ces deux méthodes : la méthode de capture-recapture sur l'ensemble des blessés de la route (i.e. non restreinte à une sous-population), qui permet l'estimation de coefficients de correction tenant compte des principaux biais de sélection, puis la projection au niveau national en stratifiant sur ces facteurs de biais (analogie avec la standardisation indirecte).

Les données d'accidentologie présentent une structure hiérarchique : accident-véhicule(s)-blessé(s), et de ce fait, il existe une certaine corrélation entre les victimes d'un même véhicule et d'un même accident. Cependant il a été montré (Lenguerrand, Martin et al. 2006), que, sauf pour des études bien particulières, l'effet de cette corrélation sur l'estimation ponctuelle est négligeable, et faible sur

l'estimation de la variance. Les intervalles de confiance étant estimés par la méthode du bootstrap, nous avons donc décidé dans ce travail de ne pas tenir compte de cette structure hiérarchique.

Les intervalles de confiance sont relativement étroits. Selon Tilling et Sterne (Tilling and Sterne 1999), Zwane et Van Der Heijden (Zwane and van der Heijden 2003), ils sont trop étroits car ils sont basés sur les données observées (même si ré-échantillonnées par le bootstrap) alors que l'approche capture-recapture fait justement l'hypothèse de sujets non observés (non recensés). Ces sujets ont des caractéristiques différentes des sujets observés (à cause des biais de sélection).
Des chercheurs ont proposé d'utiliser du bootstrap paramétrique (Buckland and Garthwaite 1991; Zwane and van der Heijden 2003) c'est-à-dire de tirer au sort les observations à partir d'une distribution, en l'occurrence multinomiale. Cependant cela ne couvre que l'étape de capture-recapture parmi l'ensemble de la procédure d'estimation.

Interprétation des résultats

Comparées aux chiffres basés sur les données des forces de l'ordre, les incidences estimées par projection sont bien plus élevées. Pour les blessés toutes gravités, le ratio entre l'incidence estimée par projection et celle basée sur les données des forces de l'ordre est d'environ 3,7 ; il est de 2,2 pour les blessés graves et de 2,2 pour les blessés hospitalisés. Dire qu'il y a 400 000 blessés (toutes gravités) en 2004 au lieu de 108 000, et 49 500 blessés graves au lieu de 17 435, change l'ampleur de ce problème de santé publique.

En termes d'évolution, quel que soit le critère de santé étudié (blessés toutes gravités, blessés graves, ou hospitalisés) les baisses basées sur les résultats de la projection sont moindres que celles issues des forces de l'ordre, et particulièrement pour les blessés graves : une baisse d'un peu moins du tiers d'après nos estimations versus une baisse de moitié d'après les données des forces de l'ordre.

Cette différence d'ampleur de la baisse provient de la modélisation d'une évolution des pratiques des forces de l'ordre. Sur le Rhône, nous observons une légère baisse du taux d'enregistrement dans les données des forces de l'ordre, ainsi qu'une tendance, à gravité donnée, à classer de plus en plus souvent les victimes comme blessés légers. Ces évolutions observées sur le Rhône ont été incluses dans les modèles, et ainsi généralisées à l'ensemble du territoire français. En effet, le processus de projection se base sur l'hypothèse d'homogénéité sur l'ensemble de la France métropolitaine des pratiques des forces de l'ordre d'enregistrement et de classement des victimes de la route. Dit autrement, cela signifie que les pratiques des forces de l'ordre dans le département du Rhône n'y sont pas spécifiques.

Notons que même si ces baisses sont de moindre ampleur que celles basées sur les données des forces de l'ordre, ces baisses en termes d'effectifs et d'incidence pour les blessés toutes gravités, pour les blessés graves et pour les hospitalisés vont dans le même sens que la large baisse du nombre de tués, notamment depuis 2002. C'est en effet en juillet 2002 que le président de la République a annoncé un certain nombre de mesures de sécurité routière, notamment l'installation en nombre de caméras de contrôle automatique des vitesses. Cette mesure a porté ses fruits : les vitesses ont diminué (ONISR 2005). La baisse la plus forte concerne la fréquence des dépassements de vitesse autorisée. Les vitesses moyennes ont aussi diminué mais de façon moindre. La baisse n'est pas uniforme chez tous les types d'usagers ; les automobilistes sont ceux qui ont le plus baissé leur vitesse. La baisse n'est pas non plus uniforme selon le type de réseau : elle concerne surtout les réseaux en agglomérations, et quasiment pas les autoroutes de dégagement (limitées à 110km/h).

Par ailleurs, les données des forces de l'ordre indiquent un taux de létalité à peu près stable et un taux de gravité en baisse. Les résultats de la projection indiquent que les deux taux sont plutôt stables (avec éventuellement, pour le taux de gravité, une légère augmentation depuis 2001). La fréquence accrue du sous-classement de la gravité par les forces de l'ordre (à gravité NISS donnée) explique la baisse de la

gravité basée sur les données des forces de l'ordre. Le taux de gravité basé sur les résultats de la projection est stable puis en légère hausse depuis 2001-2002. Cette légère hausse paraît curieuse, d'autant plus que les vitesses diminuent (rappelons que l'énergie dissipée lors d'un choc est proportionnelle au carré de la vitesse). Il est possible que notre procédure d'estimation ne corrige pas complètement le sous-enregistrement des blessés légers. C'est aussi ce que laissait penser la comparaison des NISS prédits par notre modèle avec les NISS évalués d'après les certificats médicaux issus des procès-verbaux des blessés de « BAAC seulement » de 2001 (étape 1 chap. 4).

Les résultats obtenus par notre procédure d'estimation ne nous permettent toujours pas de savoir si la large baisse des tués observée depuis 2002 s'est traduite par une baisse proportionnelle des blessés très graves (en particulier ceux avec séquelles majeures) ou si, au contraire, il y aurait eu un phénomène de « glissement » des tués vers les blessés très graves. Les effectifs de blessés très graves sur le département du Rhône sont faibles et leur évolution paraît erratique ; il faudrait disposer de données équivalentes au Registre sur d'autres départements, pour évaluer correctement leur évolution.

Les résultats par type d'usager montrent des situations nouvelles par rapport aux données des forces de l'ordre : les blessés cyclistes (toutes gravités) sont au même nombre que les blessés piétons, et les blessés graves usagers de 2RM sont au même nombre que les blessés graves automobilistes. Ces situations ne sont pas visibles avec les données des forces de l'ordre, à cause des biais de sélection : le taux d'enregistrement est faible pour les usagers de 2RM se blessant seuls, et il est très faible pour les cyclistes, qu'il y ait un tiers ou non.
Nous n'avons pas cherché à obtenir la quantité d'exposition, par type d'usagers, telle que kilomètres parcourus, ou mieux les temps passés, qui permettent de comparer les usagers non motorisés aux usagers motorisés (Lenguerrand in press). Cela donnerait une information sur le risque associé à chaque mode de transport. C'est une information intéressante mais nous nous intéressons d'abord à l'ampleur de l'insécurité routière. Les résultats en nombre de blessés sont la résultante de l'exposition et du risque associé à chaque mode de transport. Ils indiquent les enjeux en termes de santé publique.

La baisse du nombre de blessés (toutes gravités, et blessés graves) est bien moindre pour les usagers de 2RM que pour les autres types d'usagers. Cette évolution bien spécifique peut s'expliquer par plusieurs choses. D'une part, les usagers de 2RM ont très peu baissé leur vitesse, que ce soit en termes d'écarts à la vitesse autorisée ou de vitesses moyennes (ONISR 2005). D'autre part, on peut penser que les piétons et les cyclistes ont bénéficié de la baisse des vitesses des automobilistes : gravité moindre et fréquence moindre des accidents corporels.
Si les usagers de 2RM ont peu baissé leur vitesse cela vient en partie du fait qu'ils sont peu identifiables par les caméras de contrôle : seuls les deux-roues de plus de 50 cm^3 sont immatriculés, et à l'arrière, et rares sont les caméras qui lisent les plaques d'immatriculation arrières.
Un autre élément d'explication est l'augmentation de l'usage des deux-roues motorisés. Le parc des cyclomoteurs s'est réduit mais le parc des deux-roues de plus de 50 cm^3 a fortement augmenté (ONISR 2005) ; ces véhicules circulent à des vitesses plus élevées que les cyclomoteurs, ce qui entraîne une gravité plus élevée en cas d'accident, mais aussi une fréquence d'accidents plus élevée.
Un résultat de ces évolutions est qu'il y a en 2004 autant de blessés graves chez les usagers de 2RM que chez les automobilistes. Il s'agit bien d'effectifs et non de proportions. Cela signifie que les usagers de 2RM représentent le même enjeu que les automobilistes en termes de blessés graves, autrement dit qu'ils doivent être traités avec la même priorité que les automobilistes, voire plus, vu que l'usage du deux-roues motorisé se développe fortement.

Les résultats par âge et sexe sont similaires en termes de classement des classes d'âge ou des groupes hommes/femmes selon l'incidence, ou selon le taux de gravité ou de létalité, à ceux basés sur les données des forces de l'ordre. Cela vient du fait que l'âge et le sexe ne sont pas des facteurs majeurs de biais de sélection ou de biais de classement de la gravité. Cependant, le type d'usager étant lié à l'âge et au sexe, les ratios d'incidence entre classes d'âge sont modifiés (cela n'a pas été présenté). Les classes d'âge avec les plus fortes incidences sont les jeunes de 15-19 et 20-29 ans ; ils constituent aussi une grande proportion des usagers de 2RM et des cyclistes. Ceux-ci étant largement sous-représentés

dans les données des forces de l'ordre, leur incidence est plus amplifiée par la projection que celle d'autres classes d'âge.
Les résultats par âge et sexe, toujours en termes de classement des classes d'âge ou des groupes hommes/femmes selon l'incidence ou selon le taux de gravité, mais aussi en termes d'ampleur entre incidences des sous-groupes, sont les mêmes que ceux fournis directement par le Registre, concernant les blessés du Rhône (Laumon 2002).

Un résultat essentiel est que le nombre de victimes avec séquelles majeures est égal au nombre de tués à six jours (moyenne annuelle sur 1996-2004). Autrement dit, le risque de se retrouver lourdement handicapé à la suite d'un accident de la route est égal au risque d'être tué. Cela n'est vrai qu'en moyenne. Pour les usagers de 2RM et les cyclistes, le risque d'être lourdement handicapé est supérieur au risque d'être tué.
Quel que soit le type d'usager c'est en premier lieu une atteinte à la tête qui conduit à un handicap majeur. Il y a cependant des différences entre types d'usagers. Comparativement aux autres blessés avec séquelles, les usagers de 2RM sont plus souvent touchés à la colonne vertébrale et aux membres inférieurs (on peut voir ici l'effet protecteur du casque vis-à-vis des séquelles à la tête). Pour les cyclistes, c'est essentiellement une atteinte à la tête qui conduit à un handicap lourd. Les répartitions des atteintes selon les types d'usagers sont les mêmes que celles obtenues sur les données du Registre concernant les seuls blessés du Rhône (Gadegbeku, Ndiaye et al. 2006).

En termes d'association, c'est-à-dire de répartition des nombres de blessés ou de rapports des taux d'incidence, selon les classes d'âge, les sexes, et les types d'usagers, nous retrouvons les résultats tirés directement du Registre (ibid). C'est une conclusion importante en soi : nous pouvons généraliser les associations mises en évidence sur les données du Registre à la France entière.

Résumé

Contexte :
Nous avons vu au chapitre précédent qu'il est possible d'appliquer la méthode de capture-recapture aux données des blessés de la route, recensés par les forces de l'ordre et par le Registre. Nous avons pu obtenir une estimation du bilan exhaustif au niveau rhodanien. Il s'agit maintenant d'estimer le bilan exhaustif de la morbidité routière au niveau national.

Méthode :
Nous appliquons la méthode de capture-recapture au niveau du Rhône, en tenant compte des facteurs majeurs de biais de sélection : gravité lésionnelle, type d'usager, tiers (oui/non), type de réseau et type de force de l'ordre. Nous estimons ainsi des coefficients de correction du sous-enregistrement des données des forces de l'ordre. Nous appliquons ensuite ces coefficients de correction aux données nationales des forces de l'ordre, en standardisant sur les facteurs de biais du sous-enregistrement. Il y a analogie avec la méthode de standardisation indirecte. Nous obtenons ainsi le premier bilan national des victimes de la circulation routière.

Résultats :
En 2004, les données des forces de l'ordre dénombrent 108 727 blessés (non-décédés) dont 17 435 blessés graves (nécessitant plus de 6 jours d'hospitalisation). Selon la projection, il y a eu 400 200 blessés toutes gravités, dont 49 500 blessés graves (NISS 9+), 104 600 blessés hospitalisés. Le nombre de blessés graves (NISS 9+) se décompose en 5 600 cyclistes, 6 700 piétons, 21 500 usagers de 2RM, 25 100 automobilistes, et 2 000 autres usagers (moyenne annuelle sur 1996-2004). Le nombre de blessés avec séquelles majeures est de 7500 en moyenne annuelle sur la période.

Conclusion :
Les effectifs estimés sont bien plus élevés que les effectifs basés sur les données des forces de l'ordre (3,7 fois plus pour les blessés toutes gravités et 2,2 fois pour les blessés graves). Cela change l'ampleur de ce problème de santé publique. Il y a autant de blessés avec séquelles majeures que de tués, chaque année. Les enjeux sont aussi déplacés : les cyclistes sont au même niveau que les piétons en termes de nombre de blessés, et les usagers de 2RM sont au même niveau que les automobilistes en termes de nombre de blessés graves (mais pas en termes de tués). Enfin, alors que les données des forces de l'ordre indiquent une baisse du taux de gravité, les résultats de la projection indiquent que ce taux est plutôt stable.

CONCLUSION

Contrairement à d'autres types de traumatismes, ou à la plupart des pathologies, les traumatismes dus à la circulation routière bénéficient d'un enregistrement national. Cet enregistrement, effectué par les forces de l'ordre, se veut un recensement.
Nous avons vu (chap. 1) que ce n'est pas le cas : l'exhaustivité est loin d'être atteinte pour les blessés non décédés. Cela vient en partie d'un sous-enregistrement par les forces de l'ordre elles-mêmes. Cela vient également de la non-alerte des forces de l'ordre par les blessés et/ou impliqués eux-mêmes.
Le sous-enregistrement, englobant ces deux phénomènes, est non seulement d'ampleur conséquente mais il dépend de diverses caractéristiques. Les facteurs majeurs de biais de sélection des données des forces de l'ordre (mis en évidence chap. 1) sont la gravité lésionnelle, le type d'usager, la présence / absence de tiers impliqué dans l'accident, le type de réseau, et le type de force de l'ordre en charge du secteur. Les blessés les moins enregistrés sont les blessés légers, les cyclistes, les blessés dans un accident sans tiers, et particulièrement les usagers de 2RM et les cyclistes se blessant seuls. Une enquête (cf. chap.1) auprès des blessés non identifiés dans les données des forces de l'ordre a mis en évidence que le taux d'alerte des forces de l'ordre par les blessés était très dépendant de leur mode de transport, et notamment que les cyclistes les alertaient très rarement. La gravité des blessures ne semble pas jouer sur le taux d'alerte des forces de l'ordre, mais elle joue sur le taux de procès-verbaux établis.
Les données des forces de l'ordre ne contiennent pas d'information médicale à proprement parler, mais un critère de la gravité des blessés en deux catégories « blessés légers » et « blessés graves », basé sur une durée d'hospitalisation (plus ou moins de six jours). Ce critère est positivement corrélé au NISS, score de gravité basée sur une classification médicale des blessures. La classification effectuée par les forces de l'ordre ne peut être qu'une approximation. Elle est liée à des caractéristiques de l'accident ou du blessé, mais pas n'importe lesquelles : certaines de ces caractéristiques sont effectivement prédictives de la gravité réelle (chap. 2).

L'identification et surtout la quantification des points faibles des données des forces de l'ordre ont été obtenues par comparaison avec les données du Registre des victimes d'accidents de la circulation routière dans le Rhône. Le Registre est utilisé comme référence. D'une part, il est bien plus proche de l'exhaustivité que ne le sont les données des forces de l'ordre (chap. 3) ; les biais de sélection du Registre, à savoir la gravité lésionnelle, la distance à l'hôpital, et éventuellement le type d'usager, sont de faible ampleur (chap.1). D'autre part, le Registre contient des informations médicales, notamment le bilan lésionnel complet de chaque victime, codé selon une classification traumatologique internationale (l'AIS).

Le Registre a l'inconvénient de ne couvrir que le département du Rhône. Les données des forces de l'ordre ont l'immense avantage de couvrir l'ensemble du territoire français.

La coexistence des deux sources d'enregistrement sur le département du Rhône a permis d'estimer le nombre de blessés dans le Rhône échappant aux deux recensements, par l'approche de capture-recapture, et ainsi d'estimer le bilan exhaustif de l'insécurité routière dans le Rhône (chap. 3).

Afin d'estimer le bilan national de la morbidité routière, nous avons construit dans un premier temps un modèle de sous-enregistrement au niveau du Rhône. Ce modèle est une modélisation explicite de l'approche capture-recapture, qui prend en compte les facteurs de biais de sélection les plus importants, à savoir la gravité lésionnelle, le type d'usager, la présence de tiers (oui/non), le type de réseau et le type de force de l'ordre. Il fournit des coefficients de correction du sous-enregistrement des données des forces de l'ordre. Dans un deuxième temps, nous avons appliqué ce modèle au territoire national et obtenu ainsi une estimation du bilan national de l'insécurité routière. Autrement dit, nous avons appliqué les coefficients de correction du sous-enregistrement aux données nationales des forces de l'ordre, en stratifiant sur les facteurs de biais de sélection.

L'hypothèse sous-jacente à cette projection est une homogénéité sur le territoire national des pratiques des forces de l'ordre d'enregistrement et de classement des blessés, par type de force de l'ordre et par catégorie définie par les facteurs de sous-enregistrement.

En résumé, nous avons combiné l'utilisation de deux approches existant en épidémiologie : l'approche capture-recapture et une approche analogue à la standardisation indirecte (projection d'indicateurs de référence, estimés à une échelle locale/régionale, vers une échelle plus grande, en standardisant sur les facteurs pertinents).

Nous proposons ainsi la première estimation du bilan de l'insécurité routière en France, corrigé du sous-enregistrement, et, qui plus est, défini selon des classifications médicales.

Ce bilan consiste en effectifs et incidences de blessés toutes gravités, de blessés hospitalisés, de blessés graves et de blessés avec séquelles majeures, avec des estimations annuelles sur la période 1996-2004 (sauf exception). Elles sont déclinées selon le type d'usager, l'âge et le sexe des blessés.

Le problème de santé publique des victimes par accidents de la circulation routière est d'envergure bien plus grande que ne l'indiquent les données des forces de l'ordre ; c'est ainsi que pour 2004 :

- au lieu de 109 000 blessés toutes gravités, notre estimation est de 400 000,
- au lieu de 18 000 blessés graves, notre estimation est de 50 000.

La part des différents types d'usagers dans le bilan de l'insécurité routière est également modifiée :

- les enjeux cyclistes et piétons sont du même ordre : même nombre de blessés toutes gravités et de blessés graves (autour de 40 000 et 6 000 respectivement),
- l'enjeu deux-roues motorisés est majeur : le nombre de blessés graves usagers de 2RM a rejoint le nombre de blessés graves automobilistes (autour de 20 000 en 2004),
- l'enjeu des blessés avec séquelles majeures s'avère aussi important que celui des tués,
- enfin, les cyclistes et les usagers de 2RM présentent un plus grand risque d'être lourdement handicapés que tués.

Les résultats obtenus ont été comparés à des estimations extérieures, dès qu'elles existaient. De cette façon nous avons validé les estimations concernant le nombre de traumatisés médullaires France entière, le nombre de traumatisés crâniens graves en Aquitaine, et le nombre d'hospitalisés dans les données nationales des forces de l'ordre.

L'étude présentée ici peut être poursuivie sur divers aspects.

1) Il faudrait obtenir des intervalles de confiance qui tiennent compte de l'existence des blessés non recensés. La méthode du bootstrap paramétrique semble une solution.

2) Le nombre de blessés et de tués est un premier élément de mesure du problème de santé publique qu'est l'insécurité routière. Il faut aussi tenir compte de l'âge jeune des victimes d'accident de la circulation routière. Il serait donc fort utile d'utiliser la mesure globale fournie par les année de vies ajustées sur l'incapacité (AVAI (WHO 2004; Granados, Lefranc et al. 2005)). Cette notion englobe les deux notions d'années potentielles de vie perdues, et d'années de vies vécues avec incapacité.

3) Les usagers de 2RM sont un enjeu majeur, mais ils sont composés de sous-groupes assez hétérogènes, en termes de puissance de véhicule, et donc de vitesse et d'accès aux divers types de réseau. Il serait donc utile de pouvoir disposer d'une estimation séparée de leur bilan de morbidité, en distinguant les usagers de cyclomoteurs (moins de 50 cm3) et les usagers de motocyclettes (plus de 50 cm3). Cette distinction est très mal renseignée dans le Registre, mais elle est fournie par les données des forces de l'ordre. Une prédiction au niveau du Registre et/ou au niveau du bilan rhodanien paraît possible.

4) Ce travail de thèse a porté sur la période 1996-2004, période pendant laquelle les définitions des forces de l'ordre des blessés légers et graves étaient basées sur une durée de 6 jours d'hospitalisation. Les définitions ont changé en 2005. Les catégories de « blessés graves » et « blessés légers » sont dorénavant définies en fonction de l'hospitalisation (oui/non). Ces changements de définition nous conduisent à anticiper une nouvelle procédure d'estimation ; le principe et les modèles seront les mêmes mais les paramètres auront des valeurs différentes. Cependant, quelques années supplémentaires de recueil de données sont nécessaires afin de pouvoir disposer d'effectifs suffisants au niveau rhodanien.

5) Nous avons fait l'hypothèse d'homogénéité des pratiques d'enregistrement et de classement des blessés par les forces de l'ordre, sur l'ensemble du territoire français. L'extension du Registre à d'autres départements serait extrêmement bénéfique pour étudier les éventuelles variations départementales de ces pratiques des forces de l'ordre et le cas échéant, d'améliorer les modèles.

6) Le problème du sous-enregistrement des blessés par accidents de la route n'est pas limité à la France. Tous les pays avec un recensement policier y sont confrontés. La méthode présentée ici est transposable. Il suffit qu'un deuxième recensement existe, de préférence médical pour avoir une bonne mesure de la gravité lésionnelle, et sous réserve de vérifier certaines conditions, notamment celles de la méthode de capture-recapture.

Bibliographie

AAAM (1990). Association for the Advancement of Automotive Medicine, The Abbreviated Injury Scale (1990 revision), Des Plaines, Illinois.

AAAM (1994). Association for the Advancement of Automotive Medicine. Injury Impairment Scale, Des Plaines, Illinois.

Abeni, D. D., G. Brancato, et al. (1994). "Capture-recapture to estimate the size of the population with human immunodeficiency virus type 1 infection." Epidemiology **5**(4): 410-4.

Agran, P. F., D. N. Castillo, et al. (1990). "Limitations of data compiled from police reports on pediatric pedestrian and bicycle motor vehicle events." Accident Analysis and Prevention **22**(4): 361-70.

Alho, J. (1990). "Logistic regression in capture-recapture models." Biometrics **46**: 623-635.

Alsop, J. and J. Langley (2001). "Under-reporting of motor-vehicle traffic crash victims in New-Zealand." Accident Analysis and Prevention **33**(3): 353-359.

Aptel, I., L. R. Salmi, et al. (1999). "Road accident statistics: discrepancies between police and hospital data in a French island." Accident Analysis and Prevention **31**(1-2): 101-108.

Ardilly, P. (1994). Les techniques de sondage. Paris, Editions Technip.

Armstrong, M. (2000). Survol des questions touchant l'utilisation d'identificateurs personnels, Statistiques Canada.

Austin, K. (1995). "The identification of mistakes in road accident records: part 2, casualty variables." Accident Analysis and Prevention **27**(2): 277-282.

Baker, S., B. O'Neill, et al. (1974). "The Injury Severity Score: a method for describing patients with multiple injuries and evaluating emergency care." The Journal of Trauma: Injury, Infection, and Critical Care **14**(3): 187-196.

Barancik, J. and D. Fife (1985). "Discrepancies in vehicular crash injury reporting: northeastern Ohio trauma study IV." Accident Analysis and Prevention **17**(2): 147-154.

Bergeron, E., A. Lavoie, et al. (2005). "Comorbidity and age are both independent predictors of length of hospitalization in trauma patients." Canadian Journal of Surgery **48**(5): 361-366.

Bernillon, P., L. Lievre, et al. (2000). "Record-linkage between two anonymous databases for a capture-recapture estimation of underreporting of AIDS cases: France 1990-1993." International Journal of Epidemiology **29**(1): 168-174.

Bloor, M., A. Leyland, et al. (1991). "Estimating hidden populations: a new method of calculating the prevalence of drug-injecting and non-injecting female street prostitution." Bristish Journal of Addiction **86**(11): 1477-1483.

Brenneman, F., B. Boulanger, et al. (1998). "Measuring Injury Severity: Time for a Change?" The Journal of Trauma, Injury, Infection and Critical Care **44**(4): 580-582.

Brenner, H. (1994). "Application of capture-recapture methods for disease monitoring: potential effects of imperfect record linkage." Methods of Information in Medicine **33**(5): 502-6.

Brenner, H. and I. Schmidtmann (1996). "Determinants of homonym and synonym rates of record linkage in disease registration." Methods of Information in Medicine **35**(1): 19-24.

Brenner, H., C. Stegmaier, et al. (1994). "Estimating completeness of cancer registration in Saarland/Germany with capture-recapture methods." European Journal of Cancer **30A**(11): 1659-63.

Brenner, H., C. Stegmaier, et al. (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**(4): 426-30.

Broughton, J., G. Yannis, et al. (2005). Deliverable 1.6 Estimation of the real number of road accident casualties. SafetyNet project, intermediate progress report on Task 1.5. SafetyNet project, intermediate progress report on Task 1.5.

Buckland, S. and P. Garthwaite (1991). "Quantifying Precision of Mark-Recapture Estimates Using the Bootstrap and Related Methods." Biometrics **47**(1): 255-268.

Bull, J. P. and B. J. Roberts (1973). "Road accident statistics--A comparison of police and hospital information." Accident Analysis and Prevention **5**(1): 45-53.

Chao, A., P. K. Tsay, et al. (2001). "The applications of capture-recapture models to epidemiological data." Statistics in Medecine **20**(20): 3123-57.

Charnay, P. and B. Laumon (2002). Le registre des victimes d'accidents de la circulation routière du Rhône. Bron, INRETS**:** 119.

Chiron, M., B. Gadegbeku, et al. (2004). Les conséquences sanitaires des accidents de la route. La sécurité routière en France, bilan de l'année 2003. ONISR. Paris, La documentation Française**:** 195-200.

Clark, D. E. (2004). "Practical introduction to record linkage for injury research." Injury Prevention **10**(3): 186-191.

CNAM. (2007). "Affection Longue Durée - incidence." Retrieved 3 juillet 2007, from http://www.ameli.fr/l-assurance-maladie/statistiques-et-publications/donnees-statistiques/affection-de-longue-duree-ald/index.php.

Cormack, R. M. (1989). "Loglinear models for capture-recapture." Biometrics **45**: 395-413.

Crocetti, E., G. Miccinesi, et al. (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**(5): 417-23.

Cryer, P. C., S. Westrup, et al. (2001). "Investigation of bias after data linkage of hospital admission data to police road traffic crash reports." Injury Prevention **7**(3): 234-241.

Dhillon, P. K., A. S. Lightstone, et al. (2001). "Assessment of hospital and police ascertainment of automobile versus childhood pedestrian and bicyclist collisions." Accident Analysis and Prevention **33**(4): 529-537.

Dove, A. F., J. C. Pearson, et al. (1986). "Data collection from road traffic accidents." Archives of Emergency Medicine **3**(3): 193-198.

Elvik, R. and A. B. Mysen (1999). "Incomplete accident reporting; meta-analysis of studies made in 13 countries." Transportation Research Record **1665**: 133-140.

Estève, J., E. Benhamou, et al. (1994). Descriptive epidemiology. Lyon, International agency for research on cancer.

Evans, L. (2004). Traffic safety. Bloomfield Hills, Science Serving Society.

Farmer, C. M. (2003). "Reliability of police-reported information for determining crash and injury severity." Traffic Injury Prevention **4**(1): 38-44.

Fienberg, S. E. (1992). "Bibliography on capture-recapture modelling with applications to census undercount adjustments." Survey methodology **18**: 143-154.

Fisher, N., S. W. Turner, et al. (1994). "Estimating numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis." Bristish Medical Journal **308**(6920): 27-30.

Frankema, S., E. Steyerberg, et al. (2005). "Comparison of Current Injury Scales for Survival Chance Estimation: An Evaluation Comparing the Predictive Performance of the ISS, NISS, and AP Scores in a Dutch Local Trauma Registration." The Journal of Trauma, Injury, Infection and Critical Care **58**(3): 596-604.

Frischer, M., M. Bloor, et al. (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**(4): 997-1000.
Frischer, M., A. Leyland, et al. (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**(3): 170-81.
Gadegbeku, B., A. Ndiaye, et al. (2006). "Séquelles majeures en traumatologie routière, registre du Rhône, 1996-2003." Bulletin Epidemiologique Hebdomadaire **36**: 267-272.
Gallay, A., A. Nardone, et al. (2002). "La méthode capture-recapture appliquée à l'épidémiologie : principes, limites et applications." Revue d'Epidémiologie et de Santé Publique **50**(2): 219-32.
Garton, M. J., M. I. Abdalla, et al. (1996). "Estimating the point accuracy of population registers using capture-recapture methods in Scotland." Journal of Epidemiology and Community Health **50**(1): 99-103.
Gemmell, I., T. Millar, et al. (2004). "Capture-recapture estimates of problem drug use and the use of simulation based confidence intervals in a stratified analysis." Journal of Epidemiology and Community Health **58**(9): 758-765.
Goldberg, M., A. Chevalier, et al. (2007). "Recommandations de déontologie et bonnes pratiques en épidémiologie (version France-2007)." Retrieved 4 décembre 2007.
Granados, D., A. Lefranc, et al. (2005). "Les « Années de vie ajustées sur l'incapacité » : un outil d'aide à la définition des priorités de santé publique ?" Revue d'Epidémiologie et de Santé Publique **53**(2): 111-125.
Greenland, S. (2004). "Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies." American Journal of Epidemiology **160**(4): 301-305.
Haddak, M., C. Filou, et al. (2004). Redéfinition des victimes de la route. Premier rapport d'étape. Bron, INRETS**:** 37.
Hansen, A., C. Popkin, et al. (1991). Comparison of police and physician judgments of injury severity and inebriation in motor vehicle crash victims. 35th annual conference of the Association for the Advancement of Automative Medecine, Toronto, AAAM.
Hauer, E. and A. Hakkert (1988). "Extent and some implications of incomplete accident reporting." Transportation Research Record **1185**: 1-10.
Hay, G., N. McKeganey, et al. (1999). Methodological guidelines to estimate the prevalence of problem drug use on the local level. Lisbon, European Monitoring Centre for Drugs and Drug Addiction**:** 1-76.
Hilsenbeck, S. G., C. Kurucz, et al. (1992). "Estimation of completeness and adjustment of age-specific and age-standardized incidence rates." Biometrics **48**(4): 1249-62.
Hook, E. B. and R. R. Regal (1995). "Capture-recapture methods in epidemiology: methods and limitations." Epidemiologic Reviews **17**(2): 243-64.
Hook, E. B. and R. R. Regal (1999). "Recommendations for presentation and evaluation of capture-recapture estimates in epidemiology." Journal of Clinical Epidemiology **52**(10): 917-26; discussion 929-33.
Hook, E. B. and R. R. Regal (2000). "Accuracy of alternative approaches to capture-recapture estimates of disease frequency: internal validity analysis of data from five sources." American Journal of Epidemiology **152**(8): 771-9.
Hook, E. B. and R. R. Regal (2000). "On the need for a 16th and 17th recommendations for capture-recapture analysis." Journal of Clinical Epidemiology **53**(12): 1275-7.
Howe, G. R. (1998). "Use of computerized record linkage in cohort studies." Epidemiologic Reviews **20**(1): 112-21.

Hvoslef, H. (1994). Under-reporting of road traffic accidents recorded by the police at the international level. Norway, Public Roads Administration**:** 126.

IWGDMF (1995). "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**(10): 1047-58.

IWGDMF (1995). "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**(10): 1059-68.

James, H. F. (1991). "Under-reporting of road traffic accidents." Traffic Engineering and Control **32**: 573-583.

Jarvis, S. N., P. J. Lowe, et al. (2000). "Children are not goldfish - mark/recapture techniques and their application to injury data." Injury Prevention **6**: 46-50.

Johnson, R. L., B. A. Gabella, et al. (1997). "Evaluating sources of traumatic spinal cord injury surveillance data in Colorado." American Journal of Epidemiology **146**(3): 266-72.

Johri, M., E. Kaplan, et al. (1999). "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**(4): 136-146.

Kilgo, P., T. Osler, et al. (2003). "The Worst Injury Predicts Mortality Outcome the Best: Rethinking the Role of Multiple Injuries in Trauma Outcome Scoring." The Journal of Trauma, Injury, Infection and Critical Care **55**(4): 599-607.

Lai, P. W. (1980). "Model of injury severity allowing for different gradings of severity: Some applications using the British road accident data." Accident Analysis and Prevention **12**(3): 221-239.

Langley, J., N. Dow, et al. (2003). "Missing cyclists." Injury Prevention **9**(4): 376-379.

LaPorte, R. E., S. R. Dearwater, et al. (1995). "Efficiency and accuracy of disease monitoring systems: application of capture-recapture methods to injury monitoring." American Journal of Epidemiology **142**(10): 1069-77.

Laumon, B. (2002). (sous la direction de) Recherches coordonnées sur les traumatismes consécutifs à un accident de la circulation routière, et sur leurs causes et conséquences. Bron, INRETS**:** tome 1 : 289.

Laumon, B. and J. L. Martin (2002). "Analyse des biais dans la connaissance épidémiologique des accidents de la route en France." Revue d'Epidémiologie et de Santé Publique **50**(3): 277-285.

Laumon, B., J. L. Martin, et al. (1997). A French road accident trauma registry: first results. 41st annual conference of the Association for the Advancement of Automotive Medecine, Orlando, Florida.

Lavoie, A., L. Moore, et al. (2004). "The New Injury Severity Score: a more accurate predictor of in-hospital mortality than the Injury Severity Score." The Journal of Trauma, Injury, Infection and Critical Care **56**(6): 1312-1320.

Lenguerrand, E. (in press). L'exposition au risque routier et sa prise en compte dans les analyses épidémiologiques des accidents de la route selon la disponibilité de l'information. Th.: Epidemiologie: Lyon 1;

Lenguerrand, E., J. L. Martin, et al. (2006). "Modelling the hierarchical structure of road crash data - application to severity analysis." Accident Analysis and Prevention **38**(1): 43-53.

Lindquist, K. (2004, May 4, 2004). "How to estimate relative risk in SAS using Proc Genmod for common outcome in cohort studies." Retrieved December 9, 2004, 2004, from http://www.ats.ucla.edu/stat/sas/faq/relative_risk.htm.

Lopez, D. G., D. L. Rosman, et al. (2000). "Complementing police road-crash records with trauma registry data- an initial evaluation." Accident Analysis and Prevention **32**(6): 771-777.

Madigan, D. and J. York (1997). "Bayesian methods for estimation of the size of a closed population." Biometrika **84**(1): 19-31.

Martin, J. L., Y. Derrien, et al. (2003). Estimating relative driver fatality and injury risk according to some characteristics of cars and drivers using matched-pair multivariate analysis. 18th ESV Conference, Nagoya.

Martin, J. L., S. Lafont, et al. (2004). "Différences entre les hommes et les femmes face au risque routier." Revue d'Epidémiologie et de Santé Publique **52**(4): 357-367.

Masson, F., M. Thicoipe, et al. (1998). "Epidémiologie des traumatisés crâniens graves (Aquitaine 1996) [résumé]." Annales françaises d'anesthésie et de réanimation **17**: 921 (R218).

Masson, F., M. Thicoipe, et al. (2001). "Epidemiology of Severe Brain Injuries: A Prospective Population-Based Study." The Journal of Trauma, Injury, Infection and Critical Care **51**(3): 481-489.

Maxwell, J. C. (2000). "Methods for estimating the number of "hard-core" drug users." Substance Use and Misuse **35**(3): 399-420.

Meuleners, L. B., A. H. Lee, et al. (2006). "Estimating crashes involving heavy vehicles in Western Australia, 1999-2000: A capture-recapture method." Accident Analysis and Prevention **38**(1): 170-174.

Morris, A., M. Mackay, et al. (2003). Some injury scaling issues in UK crash research. international IRCOBI conference on the biomechanics of impact, Lisbon, IRCOBI.

Morrison, A. and D. H. Stone (2000). "Capture-recapture: a useful methodological tool for counting traffic related injuries?" Injury Prevention **6**(4): 299-304.

Neugebauer, R. and J. Wittes (1994). "Voluntary and involuntary capture-recapture samples--problems in the estimation of hidden and elusive populations." American Journal of Public Health **84**(7): 1068-9.

Nordentoft, E., C. Larsen, et al. (1989). "Færdselsuheldsstatistikkens dækningsgrad [the degree of coverage of traffic accident statistics]." Ugeskrift for Laeger **151**(43): 2808-2811.

ONISR (2005). La sécurité routière en France, bilan de l'année 2004. Paris, La documentation Française.

Orselli, J. (2003). L'analyse statistique des variations spatio-temporelles des accidents de la route. Paris, Conseil Général des Ponts et Chaussées**:** 196.

Osler, T., S. Baker, et al. (1997). "A modification of the Injury Severity Score that both improves accuracy and simplifies scoring." The Journal of Trauma: Injury, Infection, and Critical Care **43**(6): 922-926.

Ovième, F. (2003). Etude du critère "blessé grave" dans les fichiers BAAC des statistiques officielles des victimes d'accident corporel de la circulation routière dans le département du Rhône. Lyon, Université Claude Bernard**:** 70.

Quantin, C., C. Binquet, et al. (2004). "Estimation de la valeur discriminante des traits d'identification utilisés pour le rapprochement des données d'un patient." Revue d'Epidémiologie et de Santé Publique **52**(5): 431-440.

Razzak, J. A. and S. P. Luby (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. and R. Scragg (1994). "Application of capture-recapture methodology to estimate the completeness of child injury surveillance." Journal of Paediatrics and Child Health **30**(6): 513-4.

Robertson, L. S. (1992). Injury Epidemiology, Oxford University Press, USA.
Robles, S., L. Marrett, et al. (1988). "An application of capture-recapture methods to the estimation of completeness of cancer registration." Journal of Clinical Epidemiology **41**(5): 495-501.
Rosenman, K. D., A. Kalush, et al. (2006). "How much work-related injury and illness is missed by the current national surveillance system?" Journal of Occupational and Environmental Medicine **48**(4): 357-365.
Rosman, D. L. (2001). "The Western Australian road injury database (1987-1996): ten years of linked police, hospital and death records of road crashes and injuries." Accident Analysis and Prevention **33**(1): 81-88.
Rosman, D. L. and M. W. Knuiman (1994). "A comparison of hospital and police road injury data." Accident Analysis and Prevention **26**(2): 215-222.
Rosman, D. L., M. W. Knuiman, et al. (1996). "An evaluation of road crash injury severity measures." Accident Analysis and Prevention **28**(2): 163-170.
Sacco, W., E. MacKenzie, et al. (1999). "Comparison of alternative methods for assessing Injury Severity based on anatomic descriptors." The Journal of Trauma: Injury, Infection, and Critical Care **47**(3): 441-446.
Saillant, G., H. Pascal-Moussellard, et al. (2005). "Les lésions traumatiques de la moelle épinière : épidémiologie et prise en charge hospitalière." Bulletin de l'Académie Nationale de Médecine **189**(6): 1095-1106.
Samin, O. and I. Civil (1999). The New Injury Severity Score versus the Injury Severity Score in predicting patient outcome: a comparative evaluation on trauma service patients of the Auckland hospital. 43rd annual conference of the Association for the Advancement of Automative Medicine, Barcelona, Spain, AAAM.
Schouten, L. J., H. Straatman, et al. (1994). "The capture-recapture method for estimation of cancer registry completeness: a useful tool?" International Journal of Epidemiology **23**(6): 1111-6.
Sherman, G. (1981). Cancer incidence in Canada: completeness and ecological correlations. Division of Community Health: Toronto;
Shinar, D., J. R. Treat, et al. (1983). "The validity of police reported accident data." Accident Analysis and Prevention **15**(3): 175-191.
Sigal, A. (2005). Etude de l'exhaustivité du Registre de l'ARVAC à partir d'un échantillon aux Urgences/SMUR du centre hospitalier de Tarare. mémoire de DU Traumatologie. Lyon 1.
Simpson, H. (1996). Comparison of hospital data and police casualty data: a national study. UK, Crowthorne, Transport Research Laboratory**:** 37.
Spiegelman, D. and E. Hertzmark (2005). "Easy SAS Calculations for Risk or Prevalence Ratios and Differences." American Journal of Epidemiology **162**(3): 199-200.
Stevenson, M., M. Segui-Gomez, et al. (2001). "An overview of the injury severity score and the new injury severity score." Injury Prevention **7**: 10-13.
Stutts, J. and W. Hunter (1998). "Police reporting of pedestrians and bicyclists treated in hospital emergency rooms." Transportation Research Record **1635**: 88-92.
Tay, S., E. Sloan, et al. (2004). "Comparison of the New Injury Severity Score and the Injury Severity Score." The Journal of Trauma, Injury, Infection and Critical Care **56**(1): 162-164.
Tercero, F. and R. Andersson (2004). "Measuring transport injuries in a developing country: an application of the capture-recapture method." Accident Analysis and Prevention **36**(1): 13-20.
Tilling, K. and J. A. Sterne (1999). "Capture-recapture models including covariate effects." American Journal of Epidemiology **149**(4): 392-400.

Tunbridge, R. and J. Everest (1988). An assessment of the under reporting of road accident casualties in relation to injury severity. International IRCOBI conference on the biomechanics of impact, Cologne, IRCOBI.

WHO (2004). World Report on Road Traffic Injury Prevention, WHO.

Wittes, J. (1972). "On the bias and estimated variance of Chapman's two-sample capture-recapture population estimate." Biometrics **28**: 592-597.

Zou, G. (2004). "A modified poisson regression approach to prospective studies with binary data." American Journal of Epidemiology **159**(7): 702-706.

Zwane, E. N. and P. G. M. van der Heijden (2003). "Implementing the parametric bootstrap in capture-recapture models with continuous covariates." Statistics & Probability Letters **65**(2): 121-125.

ANNEXES

Table des matières

Annexes du chapitre 1
Synthèse bibliographique sur le sous-enregistrement des données des forces de l'ordre et les biais de sélection associés 3
Extrait du programme SAS : régression de Poisson modifiée, et calcul des probabilités marginales 11
Article publié dans le journal Accident Analysis and Prevention, version « pré-print » 12
Exemples de lettre et questionnaire envoyés dans le cadre de l'enquête « Registre seulement » 25
Extraits de programme SAS : analyse des réponses au questionnaire en tenant compte du plan de sondage 29

Annexes du chapitre 2
Synthèse bibliographique sur les erreurs de classement de la gravité des blessures par les forces de l'ordre 33
Extrait de programme SAS : régression de Poisson modifiée et calcul des probabilités marginales (erreurs de classement) 39
Article publié dans Journal of Trauma, version « pré-print » 40

Annexes du chapitre 3
Estimation du nombre de faux négatifs et de faux positifs, sur le résultat du chaînage des données de 2001 52
Résultats détaillés de l'analyse capture-recapture stratifiée, scénarios 1, 2 et 3 56
Article d'application de capture-recapture publié dans Accident Analysis and Prevention, version "pré-print" 58

Annexes du chapitre 4
Revue bibliographique sur la méthode de projection 69
Estimation du nombre de faux négatifs et de faux positifs sur le résultat du chaînage des données des années 1996 à 2004. 74
Résultats du modèle de capture-recapture (multinomial logit) ajusté sur 5 facteurs de biais, sur les données du Rhône 1996-2004 78
Extrait de programme SAS : modèle de capture-recapture 80
Traitement des valeurs manquantes : imputation simple 82
Adéquation des modèles de prédiction du NISS en trois catégories : NISS 1-3, NISS 4-8, NISS 9+ 83
Autres critères de santé étudiés : modèles de la procédure d'estimation 84
Détail du calcul sur la région Aquitaine de nombre de traumatisés crâniens graves d'après la littérature 97
Résultats supplémentaires du bilan France métropolitaine 98
Article soumis pour publication (estimation du bilan national) 131

Synthèse bibliographique sur le sous-enregistrement des données des forces de l'ordre et les biais de sélection associés

Il existe des articles sur le sujet depuis les années 1960. La grande majorité de ces études comparent les données des forces de l'ordre avec des données hospitalières. Certaines études, mêmes récentes, ne se basent pas sur un croisement des recensements policiers et des recensements hospitaliers, mais seulement sur une comparaison des distributions à la marge. On note aussi le peu d'analyses multivariées.

Tableau 1 : revues et méta-analyse récentes des études sur les biais de sélection des données des forces de l'ordre d'accidentologie par comparaison à des données hospitalières

Auteur, année	**pays**	**type d'analyse**	**facteurs mis en évidence**	**interactions mises en évidence**
(Elvik and Mysen 1999)	13 pays	méta-analyse	gravité, type d'usager, nombre de véhicules (1 vs 2+)	-
(Hvoslef 1994)	7 pays, 13 études	revue	gravité, type d'usager, tiers (oui/non) et antagoniste motorisé (oui/non), âge	Type d'usager × gravité, Type d'usager × Type d'antagoniste
(James 1991)	8 pays, 25 études	revue	gravité, type d'usager, tiers (oui/non) et antagoniste motorisé (oui/non), âge, hors-réseau, perception de la gravité, délai jusqu'à la consultation hospitalière	Tiers × type d'usager
(Hauer and Hakkert 1988)	8 pays, 18 études	revue	gravité, type d'usager, âge, passager/conducteur/non-occupant, nb de véhicules (1 vs 2+)	

Tableau 2 : études sur les biais de sélection des données policières d'accidentologie par comparaison à des données hospitalières ; triées sur l'année de parution (décroissante), exhaustif jusqu'en 2003 inclus

Auteur, année	**pays**	**Restriction de la population d'étude**	**effectif hôpital**	**effectif police**	**type d'analyse**	**facteurs mis en évidence**	**interactions mises en évidence**	**autres facteurs étudiés, NS**	**chaînage ?**
(Rossi, Farchi et al. 2005)	Italie		145 000	« 6 fois moins »	univarié	aucun biais étudié			non
(Sciortino, Vassar et al. 2005)	USA, San Francisco	piétons	1 300	2 000	bivarié	sexe, ethnie, moyen de transport à l'hôpital			oui
(Langley, Dow et al. 2003)	Nouvelle-Zélande	cyclistes	3 000		univarié	gravité, sexe, ethnie, durée d'hospitalisation		âge, région, jour, mois, et année accident	
Op.cit.					multivarié	gravité, âge, ethnie, durée d'hospitalisation		région, sexe ?	
(Laumon and Martin 2002)	France, Rhône		9 500	4 400	univarié	gravité, tiers (oui/non), type d'usager, âge		sexe	oui
Op.cit.					multivarié	gravité, tiers (oui/non), type d'usager	gravité × type d'usager × tiers	âge, sexe, urbain/rural	oui
(Cryer, Westrup et al. 2001)	Royaume-Uni, Sussex	blessés graves	2 700	2 900 ?	multivarié	type d'usager, conducteur/passager	type d'usager × âge	âge, sexe	oui
(Dhillon, Lightstone et al. 2001)	USA,	piétons et cyclistes vs véhicules motorisés	500	1 000	univarié	type d'usager, âge (conducteur), délit de fuite, type de collision, type de réseau, type de trafic, manœuvre, type de transport hop., blessé tête-cou		sexe, gravité, durée hosp., type d'assurance maladie, langue du foyer, région corporelle atteinte, perte de connaissance	oui

Auteur, année	pays	Restriction de la population d'étude	effectif hôpital	effectif police	type d'analyse	facteurs mis en évidence	interactions mises en évidence	autres facteurs étudiés, NS	chaînage ?
(Alsop and Langley 2001)	Nouvelle-Zélande	véhicules motorisés, victimes hospitalisées	5 000	27 200	univarié	gravité, type d'usager, nombre de véhicules (1 vs 2+) et type d'antagoniste, ethnie, âge, durée d'hospitalisation, délai d'admission, région, mois, jour d'accident		sexe, saison	oui
Op.cit.					multivarié	gravité, type d'usager, nombre de véhicules (1 vs 2+), âge, région, durée hospitalisation, délai d'admission, mois	type d'usager × tiers	sexe, ethnie, jour accident	oui
(Rosman 2001)	Australie, Ouest		46 000	142 000	multivarié	gravité, type d'usager, âge, type de collision		sexe	oui
(Lopez, Rosman et al. 2000)	Australie, Ouest	hospitalisés	500		univarié	gravité, type d'usager, nombre de véhicules (1 vs 2+), âge, sexe		agglomération /rural	oui
Op.cit.					multivarié	gravité, type d'usager, nombre de véhicules (1 vs 2+), sexe	âge × type d'usager	âge, agglomération /rural	oui
(Aptel, Salmi et al. 1999)	France, La Réunion,		1 100	?	univarié	gravité, type d'usager, urbain/rural, heure, alcoolémie, médecin d'urgence (oui/non), durée d'hospitalisation		âge, sexe, type de jour (semaine / week-end)	oui
op.cit.					multivarié	type d'usager, urbain/rural, type de jour (semaine vs week-end), heure, alcoolémie, médecin d'urgence, durée hosp. (≤6 vs >6)	urbain/rural × type de jour	âge, sexe, gravité	oui

Auteur, année	pays	Restriction de la population d'étude	effectif hôpital	effectif police	type d'analyse	facteurs mis en évidence	interactions mises en évidence	autres facteurs étudiés, NS	chaînage ?
(Stutts and Hunter 1998)	USA, 3 États	piétons vs véhicules motorisés	400		univarié	gravité, âge, alcoolémie, hors-réseau		sexe, ethnie, jour/nuit	oui
op.cit.					multivarié	gravité, âge, hors-réseau			oui
op.cit.		cyclistes vs véhicules motorisés	250		univarié	gravité	-	sexe, ethnie, âge, jour/nuit, alcoolémie, hors-réseau	oui
(Cercarelli, Rosman et al. 1996)	Australie, Ouest		2 500		univarié	gravité, âge, sexe, agglomération de Perth/rural, ethnie, statut marital	-	?	oui
(Rosman and Knuiman 1994; Rosman and Knuiman 1995)	Australie, Ouest		4 000	18 500	univarié	gravité, nb de véhicules (1 vs 2+), type d'usager, type d'hôpital, durée hosp. âge, sexe, lieu de résidence (agglomération de Perth / rural), ethnie, catégorie socio-professionnelle, statut marital,			oui
op.cit.					multivarié	gravité, type d'usager, sexe, ethnie, durée d'hospitalisation, type d'hôpital	nb de véhicules (1 vs 2+) pour les deux-roues motorisés	âge, lieu de résidence, catégorie socio-professionnelle, statut marital	oui
(Teanby 1992)	Royaume-Uni, 3 régions	piétons	100		test statistique?	âge		-	oui
(Schelp and Ekman 1990)	Suède, Skaraborg		180	40	aucun test statistique	gravité	-		non ?
(Olkkonen, Lahdenranta et al. 1990)	Finlande	cyclistes	1 895	1 235	univarié	tiers (motorisé), âge, blessure sur le tronc	-	saison, 3 autres types de blessure	non

Auteur, année	pays	Restriction de la population d'étude	effectif hôpital	effectif police	type d'analyse	facteurs mis en évidence	interactions mises en évidence	autres facteurs étudiés, NS	chaînage ?
(Agran, Castillo et al. 1990)	USA, California	enfants piétons vs véhicules motorisés	100	170	univarié	âge, hors-réseau, type d'action du véhicule, type d'action du piéton		sexe, ethnie, heure, jour, mois accident, type de véhicule motorisé, jour/nuit	oui
op.cit.		enfants cyclistes vs véhicules motorisés	100	300	univarié	hors-réseau		âge, sexe, ethnie, type d'action du véhicule, heure, jour, mois accident, type de véhicule motorisé, jour/nuit	oui
(Stutts, Williamson et al. 1988; Stutts, Williamson et al. 1990)	USA, Carolina	cyclistes	650	1300	univarié ; test statistique?	âge, sexe, gravité, tiers motorisé (oui/non), hors-réseau			oui
(Nordentoft, Larsen et al. 1989)	Danemark, Odense		3 000		univarié ; test statistique?	gravité, type d'usager, nombre de véhicules (1 vs 2+), âge			oui
(Morrison and Kjellstrom 1987)	Nouvelle-Zélande		112 500	?	univarié	type d'usager, nombre de véhicules (1 vs 2+), hors-réseau, année calendaire	âge chez les cyclistes		non
(Tunbridge and Everest 1988)	Royaume-Uni, Oxfordshire		5 600		univarié ; pas de test statistique?	gravité, type d'usager, nombre de véhicules (1 vs 2+)	cyclistes × nb de véhicules (1 vs 2+)		oui ?
(Dove, Pearson et al. 1986)	Royaume-Uni, Nottingham		700		univarié ; test statistique?	gravité, type d'usager	-	âge, sexe	oui
(Maas and Harris 1984)	Pays-Bas				pas de test statistique	type d'usager, âge, région, sexe ?	type d'usager × âge	trimestre	non

Auteur, année	pays	Restriction de la population d'étude	effectif hôpital	effectif police	type d'analyse	facteurs mis en évidence	interactions mises en évidence	autres facteurs étudiés, NS	chaînage ?
(Barancik and Fife 1985)	USA, Ohio		900		univarié	gravité, âge, type d'usager, nuit/jour, type d'assurance maladie, type de transport vers hôpital, habitant la région (oui/non)	-	sexe	oui
(Agran and Dunkle 1985)	USA, California	enfants, dans véhicule, pas d'antagoniste ou obstacle	50	20	univarié (pas de test statistique)	gravité, âge ?	-	heure, vitesse, type d'action	oui
(Sande and Thorson 1975)	Suède, région d'Uppsala		2 400		aucun test statistique	gravité, type d'usager, nombre de véhicules (1 vs 2+), âge		sexe	oui
(Bull and Roberts 1973)	Royaume-Uni, Birmingham		1 200	-	aucun test statistique	gravité, type d'usager, nombre de véhicules (1 vs 2+)			oui

Urbain/ rural : lieu de l'accident (sf autrement spécifié)

Tableau 3 : études sur les biais de sélection des données policières d'accidentologie par comparaison à des données de compagnies d'assurance ;

Etude	Pays	restriction	Effectif compagnies assurances	Effectif police	Facteurs mis en évidence	interactions mises en évidence	autres facteurs étudiés, NS	chaînage
(Nakahara and Wakai 2001)	Japon	Exclusion des accidents à un seul véhicule	1 094 611	968 565	âge, type d'usager			non

Tableau 4 : études sur les biais de sélection des données policières d'accidentologie par enquête ;
triées sur l'année de parution (décroissante), exhaustif jusqu'en 2003

Etude	pays	restriction	Effectif enquête	Type d'analyse	Facteurs mis en évidence	autres facteurs étudiés, NS
(James 1991)	Royaume-Uni	Véhicules motorisés, tous les accidents (corporels et matériels)	2 600	univarié et méthode de classification ou discrimination	gravité, type d'usager, nombre de véhicules (1 vs 2+), ampleur des dégâts matériels, gravité des autres impliqués, jour/nuit	
(Harris 1990)	Pays-Bas		67 000 mais 500 blessés dans un accident	univarié	gravité, âge, type d'usager, nombre de véhicules (1 vs 2+), antagoniste motorisé, jour/nuit	

Références :

Agran, P. F., D. N. Castillo, et al. (1990). "Limitations of data compiled from police reports on pediatric pedestrian and bicycle motor vehicle events." Accident Analysis and Prevention **22**(4): 361-70.

Agran, P. F. and D. E. Dunkle (1985). "A comparison of reported and unreported noncrash events." Accident Analysis and Prevention **17**(1): 7-13.

Alsop, J. and J. Langley (2001). "Under-reporting of motor-vehicle traffic crash victims in New-Zealand." Accident Analysis and Prevention **33**(3): 353-359.

Aptel, I., L. R. Salmi, et al. (1999). "Road accident statistics: discrepancies between police and hospital data in a French island." Accident Analysis and Prevention **31**(1-2): 101-108.

Barancik, J. and D. Fife (1985). "Discrepancies in vehicular crash injury reporting: northeastern Ohio trauma study IV." Accident Analysis and Prevention **17**(2): 147-154.

Bull, J. P. and B. J. Roberts (1973). "Road accident statistics--A comparison of police and hospital information." Accident Analysis and Prevention **5**(1): 45-53.

Cercarelli, L. R., D. L. Rosman, et al. (1996). "Comparison of accident and emergency with police road injury data." Journal of Trauma **40**(5): 805-809.

Cryer, P. C., S. Westrup, et al. (2001). "Investigation of bias after data linkage of hospital admission data to police road traffic crash reports." Injury Prevention **7**(3): 234-241.

Dhillon, P. K., A. S. Lightstone, et al. (2001). "Assessment of hospital and police ascertainment of automobile versus childhood pedestrian and bicyclist collisions." Accident Analysis and Prevention **33**(4): 529-537.

Dove, A. F., J. C. Pearson, et al. (1986). "Data collection from road traffic accidents." Archives of Emergency Medicine **3**(3): 193-198.

Elvik, R. and A. B. Mysen (1999). "Incomplete accident reporting; meta-analysis of studies made in 13 countries." Transportation Research Record **1665**: 133-140.

Harris, S. (1990). "The real number of road traffic accident casualties in the Netherlands: a year-long survey." Accident Analysis and Prevention **22**(4): 371-378.
Hauer, E. and A. Hakkert (1988). "Extent and some implications of incomplete accident reporting." Transportation Research Record **1185**: 1-10.
Hvoslef, H. (1994). Under-reporting of road traffic accidents recorded by the police at the international level. Norway, Public Roads Administration**:** 126.
James, H. F. (1991). "Under-reporting of road traffic accidents." Traffic Engineering and Control **32**: 573-583.
Langley, J., N. Dow, et al. (2003). "Missing cyclists." Injury Prevention **9**(4): 376-379.
Laumon, B. and J. L. Martin (2002). "Analyse des biais dans la connaissance épidémiologique des accidents de la route en France." Revue d'Epidémiologie et de Santé Publique **50**(3): 277-285.
Lopez, D. G., D. L. Rosman, et al. (2000). "Complementing police road-crash records with trauma registry data- an initial evaluation." Accident Analysis and Prevention **32**(6): 771-777.
Maas, M. W. and S. Harris (1984). "Police recording of road accident in-patients. Investigation into the completeness, representativity and reliability of police records of hospitalized traffic victims." Accident Analysis and Prevention **16**(3): 167-184.
Morrison, P. and T. Kjellstrom (1987). "A comparison of hospital admissions data and official government statistics of serious traffic accident injuries." New Zealand Medicine Journal **100**: 517-520.
Nakahara, S. and S. Wakai (2001). "Underreporting of traffic injuries involving children in Japan." Inj Prev **7**: 242-244.
Nordentoft, E., C. Larsen, et al. (1989). "Færdselsuheldsstatistikkens dækningsgrad [the degree of coverage of traffic accident statistics]." Ugeskrift for Laeger **151**(43): 2808-2811.
Olkkonen, S., U. Lahdenranta, et al. (1990). "Incidence and characteristics of bicycle injuries by source of information." Acta Chir Scand **156**(2): 131-6.
Rosman, D. L. (2001). "The Western Australian road injury database (1987-1996): ten years of linked police, hospital and death records of road crashes and injuries." Accident Analysis and Prevention **33**(1): 81-88.
Rosman, D. L. and M. W. Knuiman (1994). "A comparison of hospital and police road injury data." Accident Analysis and Prevention **26**(2): 215-222.
Rosman, D. L. and M. W. Knuiman (1995). "A comparison of hospital and police road injury data; erratum." Accident Analysis and Prevention **27**(1): 141.
Rossi, P. G., S. Farchi, et al. (2005). "Road Traffic Injuries in Lazio, Italy: A Descriptive Analysis From an Emergency Department-Based Surveillance System." Annals of Emergency Medicine **46**(2): 152-157.
Sande, J. and J. Thorson (1975). "An evaluation of the official Swedish statistics on seriously injured in road traffic accidents." Scand J Soc Med **3**(1): 5-11.
Schelp, L. and R. Ekman (1990). "Road traffic accidents in a Swedish community." Public Health **104**: 55-64.
Sciortino, S., M. Vassar, et al. (2005). "San Francisco pedestrian injury surveillance: Mapping, under-reporting, and injury severity in police and hospital records." Accident Analysis and Prevention **37**(6): 1102-1113.
Stutts, J. and W. Hunter (1998). "Police reporting of pedestrians and bicyclists treated in hospital emergency rooms." Transportation Research Record **1635**: 88-92.
Stutts, J. C., J. E. Williamson, et al. (1988). "Bicycle accidents: an examination of hospital emergency room reports and comparison with police accident data." Transportation Research Record **1168**: 60-71.
Stutts, J. C., J. E. Williamson, et al. (1990). "Bicycle accidents and injuries: A pilot study comparing hospital- and police-reported data." Accident Analysis and Prevention **22**(1): 67-78.
Teanby, D. (1992). "Underreporting of pedestrian road accidents." British Medical Journal **304**: 422.
Tunbridge, R. and J. Everest (1988). An assessment of the under reporting of road accident casualties in relation to injury severity. International IRCOBI conference on the biomechanics of impact, Cologne, IRCOBI.

Extrait du programme SAS : régression de Poisson modifiée, et calcul des probabilités marginales

```
title 'poisson avec estimation robuste de la variance ' ;

ods trace on;
proc genmod data=  victime9701 ;
class     id_victi MAIS NISS blesse  antagoniste agec sexe courly zonefo  reseau eclairement jour_semaine annee fo_courly fo_reseau foresGL;
format blesse $f5bless. antagoniste $tiers.  agec f5agec. niss f5niss.
reseau $f5myres. courly $f2courly.  zonefo fordre. foresGL foRGLrd.  eclairement $fjournt. jour_semaine f2typj. ;
model dansBAAC =  NISS  blesse*antagoniste  agec sexe foresGL eclairement annee  /
link=log dist=poisson ;
repeated subject = id_victi /type=unstr;
output   out      = Res_pois
        pred     = pred
        resraw   = ResRaw
        reschi   = Reschi ;
ods output GEEEmpPEst=poiss_est ;
lsmeans  NISS  blesse*antagoniste agec sexe  foresGL eclairement annee;
ods output LSMeans=LSMEANS;
run;

data poiss_est;
set poiss_est (where=(Parm NE 'Scale'));
RR=exp(estimate);
lowerCL=exp(LowerCL);
upperCL=exp(UpperCL);
format RR lowerCL upperCL 6.2 ;
run;
proc print data=poiss_est;
var parm level1  level2 RR  lowerCL upperCL; * level2 level3 ;
run;

proc contents data= LSMeans varnum;
run;
data probmeans;
set lsmeans;
prob=exp(estimate);
run;


proc print data=probmeans width=minimum;
var effect NISS agec sexe  blesse antagoniste foresGL eclairement annee prob;
run;
ods rtf close;
```

Article publié dans le journal Accident Analysis and Prevention, version « pré-print »

Amoros E, Martin JL, Laumon B, Under-reporting of road crash casualties in France. *Accident Analysis and Prevention,* 2006, 38(4), 627-635

Abstract
Introduction: Police crash data, which are the basis for safety research in most countries, are incomplete and biased. We focus here on the extent of under-reporting in France and how it is related to casualty and crash characteristics.
Methods: Police data are compared with a road trauma registry, on the Rhône county, after record-linkage. The study covers the 1997-2001 period, totalling 59714 casualties. A multivariate analysis is conducted, modelling the probability of being police-reported among casualties in the registry, as a function of crash and casualty characteristics. Results are expressed as relative risks (RR) and adjusted probabilities.
Results: Police reporting rate is 37.7%. Under-reporting varies mainly according to injury severity (RR=0.35 for slightly injured vs. severely injured), to road user type and third party involvement (yes/no): comparing casualties with no third party vs. those with one, RR=0.32 among motorcyclists whereas RR=0.78 among car occupants. Under-reporting also varies with road type, road environment (metropole/rural) and type of police force, all of which are structurally dependent.
Conclusion: Any study based on police crash data may be quite misleading. We are therefore working on obtaining unbiased estimates of road casualties figures, by extrapolating the Rhône road trauma registry to the nation-wide level.

Keywords: under-reporting, selection bias, road trauma registry, police data, multivariate analysis, relative risk

1 Introduction

In most countries, epidemiological knowledge of road crash casualties is mainly based on data collected by the police. Fatal casualties are quite clearly defined and well reported, but this is not the case for non-fatal casualties. The under-reporting of non-fatal casualties is well known (Hauer and Hakkert, 1988; James, 1991; Hvoslef, 1994; Stutts and Hunter, 1998; Aptel et al., 1999; Elvik and Mysen, 1999; Alsop and Langley, 2001; Cryer et al., 2001; Dhillon et al., 2001; Rosman, 2001; Laumon and Martin, 2002; Langley et al., 2003); the degree of under-reporting can be quite large, and furthermore it varies according to characteristics such as injury severity and road user type.

The validity of data used as a basis for epidemiological knowledge of road casualties should be assessed, and this should be done in each country since the police definition of road casualties and police reporting practice are country-specific. We analyse here the police under-reporting of non-fatal casualties in France: the degree of under-reporting and how it varies according to some crash or casualty characteristics; in other words, it consists of identifying and quantifying selection bias factors.

This can be achieved thanks to the existence of a medical road trauma registry. This registry covers all victims of road crashes in the Rhône county who seek medical attention in health facilities in the county or close surroundings. Record-linkage has been applied to the registry and police files so that casualties reported by both sources are identified. It is therefore possible to analyse the police "filter", i.e. the filter through which casualties in the registry are "selected" into being police-reported. This is done by modelling the probability of being police-reported among casualties in the registry, as a function of their crash and/or casualty characteristics. In this analysis, police under-reporting covers both under-reporting to the police and under-reporting by the police as we can not distinguish between the two. Under-reporting to the police corresponds to no-one calling the police; under-reporting by the police corresponds to the police not writing a crash report even though present at the crash scene, or omitting some of the casualties within a reported crash. The unit of the whole analysis is the casualty, not the crash.

Additionally, a survey on casualties identified in the registry but not in the police files has been set up, with the aim of having a better understanding of police under-reporting. In particular, this study aims at identifying and quantifying reasons for not calling the police.

2 Data

2.1 Police traffic crash data

The French police are required by law to write a crash report for every road crash causing at least one casualty. A road crash is officially defined as a crash involving at least one vehicle and occurring on the network open to public traffic. Skateboard or roller skate users are considered as pedestrians by the police, and, as such, are only classified as road casualties if hit by a vehicle. There is no restriction about motorised vehicles, in other words there is no exclusion criteria on bicycles. The police crash report should report all the people involved in the crash: injured (i.e. casualties) and non-injured ones. The police crash report contains detailed information on the crash, the crash environment and conditions, the vehicles involved, but it contains limited information on the people involved.
There are three main police forces in France (whose French names are 'police', 'gendarmerie' and 'Compagnie Républicaine de Sécurité') and we wish to check whether under-reporting differs between them. These 3 police forces can be named 'urban police', 'rural police' and 'urban motorway police' according to their area of control. It is however not clear-cut: the 'urban police' mostly operate in cities of a certain size; the 'rural police' mostly operate in rural areas but also in some suburbs of large cities; the urban motorway police operate on motorways in and around large cities.

2.2 The Rhône road trauma registry

A road trauma registry has been in use since 1995 in the Rhône county (Laumon et al., 1997; Charnay and Laumon, 2002), which is a large county of 1.6 million inhabitants, consisting of a large city, Lyon, its suburbs and a rural area in the northern part. The registry covers all casualties from road crashes in the Rhône county who seek medical attention in health facilities. Inclusion criteria are broader than the

police ones: off-road crashes are not excluded; roller skate, skate-board or scooter users are not considered as pedestrians but as road users using a mean of transport and are hence included, whether hit by another vehicle or not.
The registry is based on the participation of all health care facilities in the county (and its close surroundings) that may receive victims of a traffic crash: it includes some 150 health care facilities: from emergency departments, intensive care units, surgery units... to rehabilitation departments, as well as pre-hospital emergency care. The registry includes both inpatients and outpatients, i.e. all casualties, whether hospitalised or not. Information collected for each casualty consists of a few crash characteristics and of the following casualty characteristics: gender, date of birth, place of residence, hospital stay, hospital transfer if relevant, and accurate injury assessment. Indeed, for each subject, injury assessment is based on the whole set of diagnoses provided by the different health services the subject may have gone through. Plain text diagnoses are coded by the registry physicians according to the Abbreviated Injury Scale (AIS), 1990 revision. Each injury is assigned a severity score, ranging from AIS 1 (minor) to AIS 6 (beyond treatment). To measure the severity of multiple injury casualties, we use the NISS i.e. the New Injury Severity Score (Osler et al., 1997), which is defined as the sum of squares of the AIS severity scores of a patient's three most severe injuries, regardless of body region in which they occur.

2.3 Record-linkage

Police data -restricted to the Rhône county- and registry data have been linked at the casualty level using a semi-automated record-linkage procedure (Laumon and Martin, 2002; Clark, 2004). Linking variables are date and time of the crash, crash location, type of road user, date of birth (only year and month) and gender of the casualty. The procedure was not fully automated since a major linking variable (crash location) was left as unformatted free text to avoid losing any valuable information.

2.4 Additional data from a postal survey

In order to better understand police under-reporting, a postal survey was set up, on those casualties identified in the registry but not in the police files. The objectives were to estimate the frequency and the characteristics of the following subgroups: 1) the police were not alerted, 2) the police were alerted but did not intervene, 3) the police intervened but no crash report was written, 4) they intervened and a crash report was written but because of incomplete or incorrect data, the record-linkage with the registry was not successful. Situation 1 corresponds to under-reporting to the police while situations 2, 3 and 4 correspond to under-reporting by the police (sometimes named under-recording). Another objective of the survey was to estimate the frequency of the different reasons given by the casualties for not calling the police.

The study population consists of casualties injured in a road crash in 2001, identified in the registry but not in the police files, and fulfilling the police definition of a road casualty (7412 of them). A sample of 2143 casualties was drawn, using a sampling design stratified on injury severity (MAIS 1-2 vs. MAIS 3+)[1] and on road user type. A questionnaire and a cover letter were mailed to the subjects in April-may 2004; a reminder was sent one month later. We took great care in writing the questionnaire and cover letter, asking similar questions about (not) calling the police and about (not) calling the emergency care, and assuring the subjects that the data will be treated confidentially.

3 Methods

3.1 Main analysis

As previously stated, we focus on non-fatal casualties, as they suffer from a large under-reporting compared to fatal casualties, and also because the underlying reasons and characteristics of their under-reporting might be quite different. We compare the police data to the registry, which we restrict to the casualties that fulfil the police definition of road casualties: this implies that roller skate, skateboard and scooter users should be considered as pedestrians and excluded if no third party is

[1] MAIS=Maximum AIS ; at the time the study was set up, MAIS had been chosen as the overall measure of severity for casualties with multiple injuries

involved. However, we do not exclude off-road crash casualties as this off-road criteria is quite unclear and not systematically followed by the police forces.

The first part consists of measuring the degree of police under-reporting of non-fatal casualties. We compute the police reporting rate, using the definition suggested by Elvik and Mysen (as n°3) as the only theoretically correct one (Elvik and Mysen, 1999). This definition corresponds to the ratio between the number of casualties reported in the police files and the total number reported by the police and the registry, i.e. the sum over police data only, registry only, and overlap between the two. Other definitions have been used (reported as n°2 and n°1 by Elvik and Mysen) corresponding to the ratio between, respectively the number of casualties in the police files, in the police files only, and the number of casualties in hospital data.
The second part of the analysis aims at identifying and quantifying selection biases. We consider the following characteristics as potential bias factors: 1) casualty and crash characteristics: injury severity (measured by the NISS), age, gender, road user type, whether the crash involved a (human) third party: i.e. a pedestrian or someone in a vehicle, motorised or not, 2) crash environment characteristics: road type, environment: Lyon metropole vs. rural area, light (daylight vs. night), type of day (beginning of the week vs. extended week-end), 3) police force in charge of the area where the crash occurred, and finally calendar year. We compare characteristics of the casualties included in the registry to the ones that are included in both the registry and the police files, i.e. the ones that have been further selected by the "police filter". To do this, we model the probability of the casualties in the registry to be also reported in the police files, as a function of casualty, crash and police characteristics. Variables considered for inclusion in the multivariate modelling are those found in the literature or significantly associated with police under-reporting in univariate analyses. We also consider inclusion of structural interactions: the one between road user type and involvement of a third party (as pedestrians are only considered –in the police data- as road crash victims if hit by a vehicle), and the interaction between road type, metropole/rural area and police force in charge of the area where the crash occurred.

We favour relative risks (RR) to measure the association between the event of interest -being reported in the police data- and bias factors. The relative risk is defined as the ratio of probabilities of the event of interest between a subgroup of a specific characteristic and the subgroup with the reference characteristic. For instance, the pedestrians' relative risk of being police-reported is the ratio between the pedestrians' probability and the car occupants' probability (reference category) of being police-reported; a relative risk of 2 would mean that pedestrians are twice as likely as car occupants to be police-reported. It is not possible in this study to use the usual approximation of relative risk estimates by odds ratio estimates, which are easily obtained from logistic regression. Indeed, the event of interest -being police-reported- is not a rare event. In this case, different model-based methods can be used to estimate relative risks (Greenland, 2004; Spiegelman and Hertzmark, 2005), all within the Generalised Linear Models framework: log-binomial, logistic with some transformations, and a modified Poisson regression. We choose the last one: it consists of applying a Poisson model to binomial data, which would lead to an over-estimation of the RR estimate error: this is rectified by using a robust error variance procedure, the sandwich estimation (Zou, 2004). This is implemented using SAS software, GENMOD procedure (Lindquist, 2004; Spiegelman and Hertzmark, 2005). The Poisson model looks like:

$$\log(\Pr ob(\textit{being police reported})) = b_0 + b_1 X_1 + b_2 X_2 + ... + b_j X_j$$

with X_1, X_2, X_j being some crash and/or casualty characteristics.
We also provide results in terms of adjusted (standardised) probabilities from the multivariate model; these correspond to marginal probabilities for a given characteristic, if the data were balanced. These probabilities are computed by standardising the data over the other bias factors (more precisely by averaging over levels of the other categorical variables, setting equal weights). This is achieved using LSMEANS option in GENMOD procedure, SAS software.

3.2 Survey analysis

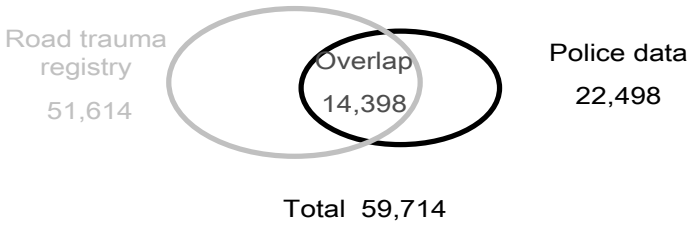
We check for discrepancies between respondents and non–respondents (in the broad sense: those refusing to respond and those whose address is missing or invalid). For this purpose we conduct a multivariate logistic regression of the probability of response (yes vs. no or missing/invalid address), that takes into account the stratified sampling design. This is implemented with SURVEYLOGISTIC procedure, SAS software. We find that respondents and non-respondents (in the broad sense) differ on a number of characteristics (see results section) indicating that the respondents are not representative of the study population. Hence, we can not estimate the size of the different subgroups of under-reporting, and we restrict our analyses to a few broad frequency results among the respondents without extrapolating to the whole study population. We estimate the proportion of respondents who report having called the police, and among the others, the relative frequencies of the reasons given for not calling the police. Proportions and their confidence intervals are estimated by taking account of the stratified sampling design, using SURVEYMEANS procedure, SAS software. We did not make any further sampling weight adjustment to compensate for the non-responses, as we can not assume that the non-respondents are "missing at random".

4 Results

4.1 Under-reporting and selection bias

The registry reports 53 988 non-fatal casualties from road crashes in the Rhône county during the 1997-2001 period. Of these we exclude 2374 roller skate, skateboard or scooter users not hit by a vehicle, leading to 51 614 casualties in the registry. The police reports 22498 non-fatal casualties. Of these, 14 398 are identified in both sources, leading to a total of 59 714 non-fatal casualties (figure 1). The police reporting rate is 37.7%

Figure 1: distribution of non-fatal casualties according to the road trauma registry and police files, in the Rhône county, 1997-2001



For further analysis, we have to exclude 678 casualties because of missing NISS, so that the analysis is based on 50 936 subjects. In the univariate modelling, all the examined factors are significantly associated with police under-reporting. In the multivariate modelling, one variable -type of day- is no longer associated with police under-reporting. The multivariate model includes injury severity, age, gender, road user type, third party involvement, light conditions, road type, crash environment, police force area and calendar year, as selection bias factors, each one being adjusted for all the others:

- Under-reporting is inversely and strongly associated with injury severity: there is a clear gradient of decreasing probability of being police-reported with decreasing NISS categories, with relative risk as low as 0.38 for NISS 1-3 (minor) casualties.
- Under-reporting is inversely and slightly associated with age: the younger the casualties, the lower the probability of being reported to/by the police.

- Under-reporting is slightly associated with gender: female casualties are 0.95 times less likely to be reported than males.
- Under-reporting is strongly related to both road user type and involvement of a third party. Within crashes involving a third party, occupants of vans, trucks, buses or trams are 1.17 times more likely to be reported compared to car occupants, and cyclists are 0.75 times less likely to be reported than car occupants ; pedestrians, motorcyclists and car occupants (in crashes involving a third party) do not display different reporting probabilities. Casualties in crashes without a third party involved are less likely to be police-reported compared to casualties with a third party involved, and even less so if they are motorcyclists (RR=0.32) and much less so if they are cyclists (RR=0.06). This pattern is shown in Figure 2**.**
- Casualties in crashes occurring in the daylight are slightly less likely (RR=0.92) to be police-reported than casualties involved in night-time crashes.
- Under-reporting is strongly associated with the combination of road type, crash environment (metropole vs. rural area) and police force area (figure 3): police-reporting decreases with decreasing road level i.e. from motorways, to state and county roads, local roads, and down to off-road. Crashes in areas controlled by the 'rural police' are less reported than crashes in areas controlled by the 'urban police', within a given road type. The comparison of reporting rates between the metropole and the rural area is not straightforward: it depends on the police force control area.
- Under-reporting rate varies slightly with calendar year.
- Finally, whatever the characteristic considered, the "unknown" category displays a rather low relative risk.

Figure 2: adjusted (standardised) probabilities of being police-reported, among registry casualties, according to road user type and involvement of a third party (with/without)

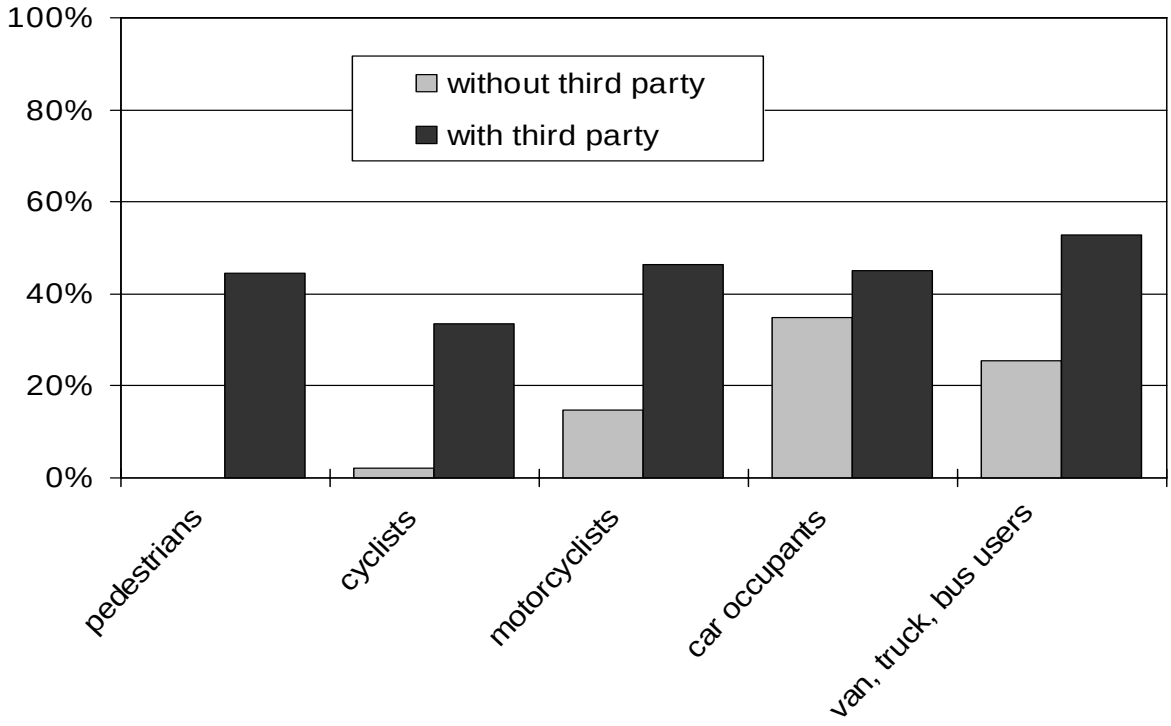
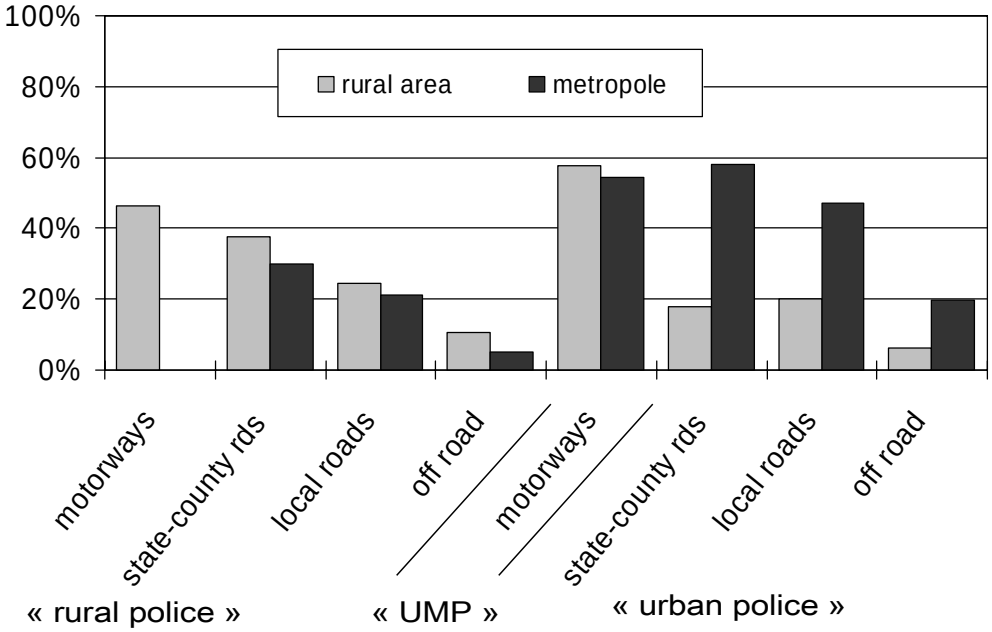


Figure 3: adjusted (standardised) probabilities of being police-reported, among registry casualties, according to police force control area, road type and road environment (rural/metropole)



Detailed results of the multivariate model are displayed in table 1: variables significantly associated with police under-reporting (which are all included in the model), adjusted (standardised) probabilities, and adjusted relative risks; variables are displayed in the following order: casualty, crash, environment and general characteristics.

Table 1: adjusted relative risks and adjusted (standardised) probabilities of being police-reported; multivariate analysis (50936 non-fatal casualties, Rhône county, 1997-2001)

		n	adj. prob.	RR	95% C.I.	
injury severity:						
NISS 1-3		37 873	15.1 %	0.38	0.36	0.41
NISS 4-8		8 632	24.3 %	0.62	0.58	0.66
NISS 9-15		2 822	33.1 %	0.84	0.79	0.90
NISS 16-24		997	37.6 %	0.96	0.89	1.03
NISS 25-75		612	39.3 %	1.00	-	-
casualty age:						
0-13 years old		5712	27.1 %	0.95	0.90	1.00
14-17		5116	26.8 %	0.93	0.89	0.98
18-24		12564	27.8 %	0.97	0.94	1.00
25-64		25111	28.6 %	1.00	-	-
65 and above		2433	30.9 %	1.08	1.03	1.13
casualty gender						
female		19441	27.5 %	0.95	0.92	0.97
male		31495	29.0 %	1.00	-	-
road user type	third party					
pedestrians	with	4742	44.4 %	0.99	0.95	1.03
cyclists	with	1651	33.6 %	0.75	0.69	0.81
motorcyclists	with	4924	46.5 %	1.03	0.99	1.07
car occupants	with	20744	45.0 %	1.00	-	-
van, truck, bus, tram users	with	476	52.7 %	1.17	1.07	1.29
unknown (& other)	with	44	55.0 %	1.22	0.90	1.67
cyclists	without	5284	2.1 %	0.06	0.05	0.08
cyclists	with	-	-	1.00	-	-
motorcyclists	without	5127	14.7 %	0.32	0.29	0.35
motorcyclists	with	-	-	1.00	-	-
car occupants	without	6870	34.9 %	0.78	0.75	0.81
car occupants	with	-	-	1.00	-	-
van, truck, bus, tram users	without	675	25.5 %	0.48	0.41	0.57
van, truck, bus, tram users	with	-	-	1.00	-	-
unknown (& other)	without	399	36.3 %	0.66	0.46	0.96
unknown (& other)	with	-	-	1.00	-	-
police force area x road type x environment:						
‘urban motorways police’, motorways		4508	55.1 %	1.16	1.12	1.21
‘rural police’, motorways, rural area		117	46.3 %	0.98	0.81	1.18
‘rural police’, state & county roads, rural area		3027	37.7 %	0.80	0.76	0.83
‘rural police’, state & county roads, metropole		474	29.9 %	0.63	0.56	0.71
‘rural police’, local roads		7221	22.5 %	0.48	0.45	0.50
‘rural police’, off road		810	7.6 %	0.16	0.11	0.24
‘urban police’, state & county roads, metropole		266	58.1 %	1.23	1.10	1.37
‘urban police’, road except motorways, rural area		1523	19.6 %	0.41	0.37	0.47
‘urban police’, local roads, metropole		18764	47.3 %	1.00	-	-
‘urban police’, off road		837	18.4 %	0.39	0.32	0.48
unknown		13749	18.3 %	0.39	0.37	0.40
light						
day light		27059	28.4 %	0.92	0.89	0.95
night		10107	30.8 %	1.00	-	-
unknown		13770	25.7 %	0.83	0.80	0.86
calendar year:						
1997		9921	26.9 %	0.98	0.94	1.02
1998		10379	28.8 %	1.05	1.01	1.09
1999		10704	29.1 %	1.06	1.02	1.10
2000		9595	28.9 %	1.05	1.01	1.10
2001		10337	27.4 %	1.00	-	-

from a Poisson regression with robust error estimation; RR=relative risk; C.I. = Confidence Intervals

4.2 Survey results

Out of the 2143 sampled subjects, 756 subjects had no address or one that was invalid (either incomplete or incorrect, or no longer valid because the subject has moved), 840 subjects did not reply (or refused to: 20 of them), and 547 subjects sent back the completed questionnaire, i.e. 39.4% of those who were reached by postal mail.

The multivariate logistic regression comparing the non-respondents (in the broad sense) to the respondents includes age, gender, injury severity, safety device, road type, living environment as significantly associated with response to the questionnaire (yes vs. no in the broad sense). Compared to respondents, non-respondents are on average rather young (18-24 years old), male, slightly or moderately injured (NISS<15), not wearing the compulsory seatbelt or helmet, involved in a crash occurring in streets or local roads, and living in a urban environment.

Among the 547 responses, it turns out that 49 (8.9%) do not fulfil the study criteria (bicycle or scooter accident at home/school, road crash outside the Rhône county, or sport or occupational accident exclusively) and for another 3 responses, there has been a confusion with another road crash. This leaves 495 exploitable questionnaires. Among these respondents, 47.3% (C.I. 43.5%-51.1%) claim to have called the police, 39.4% (C.I. 35.7%-43.1%) report to have not called the police and 13.3% (C.I. 10.6%-16.0%) do not remember or do not answer. Table 2 displays the reasons given by the respondents for not calling the police.

Table 2: reasons given for NOT calling the police (among 230 respondents from the postal survey who did not alert the police)

	n	Prop. (%)	95% C.I. (%)	
1) it was not judged to be useful, because:	167	74.7	69.7	79.7
1a) there was no third party involved	*79*	*32.5*	*28.1*	*36.9*
1b) a third party was involved but an amicable agreement was written for insurance purposes	*46*	*23.4*	*19.4*	*27.4*
1c) nobody seemed to be seriously injured	*66*	*30.8*	*25.2*	*36.4*
2) nobody thought of calling them	22	8.5	5.3	11.6
3) reluctance to call them	8	3.5	1.3	5.7
4) other reason	9	4.4	1.8	6.9
5) unknown	19	7.1	4.3	9.9

The sum is over 100% as respondents could give several answers

5 Discussion

The strength of this study lies mostly in the Rhône road trauma registry. This registry has been acknowledged by the French National Registry Committee. The study is based on a five-year period, on a county with a large population, and hence on a large number of subjects.
The underlying assumption of the study is that each police force having a centralised national structure (not regional) and receiving centralised instructions, each one has the same way of reporting crashes and casualties all over France. Under this assumption, the study results on selection biases can be extrapolated to the target population of non-fatal casualties in France.

The first study result is that police reporting is quite low. It is however consistent with the literature (Hauer and Hakkert, 1988; James, 1991; Hvoslef, 1994; Elvik and Mysen, 1999) especially since the Rhône road trauma registry has a broad coverage with both inpatients and outpatients.

The second main result concerns the selection biases, i.e. the characteristics that are associated with the police-reporting probability, and the extent of these biases as quantified in the multivariate analysis, each bias being adjusted for the others. The major selection bias factors are injury severity, road user type combined with third party involvement (yes/no), and the combination of road type, metropole vs. rural area and police force control area. Police-reporting probability increases with increasing injury severity: it is somehow trivial that the higher the severity, the more likely the people involved -as well as the police- will feel that the casualty is a serious one and should be reported. This

bias factor has been widely reported (Hauer and Hakkert, 1988; James, 1991; Hvoslef, 1994; Elvik and Mysen, 1999).
Police-reporting probability varies with road user type, adjusted for the other factors; this is also a common bias factor (Hauer and Hakkert, 1988; James, 1991; Hvoslef, 1994; Elvik and Mysen, 1999). In particular, cyclists suffer from a low reporting probability: extremely low when no third party is involved, but still lower than other road users when a third party is involved: bicycle crashes are not always considered as road traffic crashes by the people involved and/or by the police even when a third party is involved. This is also the case in other countries (Hvoslef, 1994; Elvik and Mysen, 1999), even in countries of Northern Europe where riding a bike is widespread and really considered as a mean of transport.
Police-reporting probability is higher for casualties in crashes involving a third party compared to crashes without (adjusted for other factors) and this can be explained by the issue of responsibility and legal procedure. Among such crashes, the reporting probability is even higher for people injured in vans, trucks, buses… where there could be the further issue of occupational accident, and hence even stronger pressure to comply with legal procedures. The bias factor on the involvement of a third party is similar to what is found in other studies (Hauer and Hakkert, 1988; James, 1991; Hvoslef, 1994; Elvik and Mysen, 1999) that show that multi-vehicle crashes are better reported than single-vehicle crashes.

The interaction of road type, environment (metropole vs. rural area) and police force control area with police under-reporting (adjusted for the other factors) seems quite complex, but it is largely explained by their structural interdependency (and we will not go into further detail here as this is specific to France). The association between higher reporting probability with higher road level is probably due to some expected higher severity on higher level roads (because of higher speed). In the under-reporting reviews, road type has not been considered as such, only off-road has, and it is associated with lower reporting rates (James, 1991). Association of police-reporting rate with urban/rural areas has been explored in a few studies, but no clear picture has come out: no significant difference was found in the reporting rates between metropole and rural areas in Australia (Lopez et al., 2000); a lower reporting rate has been indicated (Hvoslef, 1994) in rural areas compared to urban areas, explained by the longer distance to emergency units. Variation of police-reporting probability according to the type of police force, if relevant, has apparently not been explored.

There are a number of minor selection biases i.e. characteristics (age, gender, light, calendar year) that are weakly but significantly associated with the police-reporting probability in the multivariate model; this is favoured by the large number of observations the study is based on. One of these minor biases is the increased police-reporting probability with increased age, and this is coherent with the reviews (Hauer and Hakkert, 1988; James, 1991), the first one indicating a gradient with age, and the other indicating that children are less reported but probably so because of their road user type (mainly bicycle). This is confirmed in our study as the extent of this bias factor is quite small in the multivariate analysis (in which the road user type is accounted for) compared to the univariate analysis (not presented). The remaining effect may be to do with an expected higher fragility at older age.
The slightly lower police-reporting probability for casualties in daylight crashes might also be due to a higher expected severity for night-time crashes compared to daytime crashes.
The association between police-reporting probability and gender is not easy to explain, given that other factors such as injury severity and road user type are adjusted for. Still, male and female crash risks display quite a number of differences that can not be entirely explained by different road user types or other characteristics (Martin et al., 2004). Most studies have not found this selection bias. It is however very small.

There is possibly yet another selection bias factor: even though all the people involved in the crash should be reported, it is very likely that injured drivers are better police-reported than injured passengers, since the crash report should identify a driver for each vehicle involved.

“Unknown” categories display rather low reporting probabilities compared to others; this is partly due to lower record-linkage probabilities: casualties with little information are more difficult to identify

and hence to link between the registry and the police data. Typically, unknown road type, or unknown police force control area correspond to missing or incomplete crash location, while crash location is a major piece of information in the record-linkage process.

Results of the postal mail survey on casualties identified in the registry but not in the police files are consistent with the only similar survey we found (Simpson, 1996), in which the main reasons given by the casualties for not reporting to the police were that the crash was too trivial, or that no third party was involved.
One should however be very cautious in the interpretation of the survey results. Indeed the respondents differ from the non-respondents and those with invalid address on a number of characteristics, as displayed in the results section. Furthermore, we can assume that among non-respondents (in the broad sense), a larger proportion (than among respondents) have not called the police, and that the reasons for not having done so differ, with presumably a larger proportion of subjects "reluctant to" and among these, presumably a proportion having committed a driving offence. The reasons are that, firstly, the comparison of both non-respondents (in the broad sense) to the respondents shows that non-respondents are more likely to have not worn the compulsory seatbelt or helmet at the time of the crash (information from the registry) indicating that they are more likely not to abide by the law. Secondly, if they were reluctant to call the police at the time of the crash, it is rather logical that they were also reluctant to give their address at the hospital (information source of the registry), or later on to answer the questionnaire, irrespective of its wording. This implies that, in the study population i.e. casualties identified in the registry but not in the police files, the proportion that has not called the police must be higher than a half.
The survey also indicates that a small but significant proportion of the study population do not fulfil the registry or the police criteria and this could not have been assessed previously because of incomplete or incorrect data in the registry.

6 Conclusion

The large extent of police under-reporting of non-fatal casualties and the associated selection biases means that police data give the wrong picture. Slight casualties are under-represented; their consequences in terms of health care, disabilities and associated costs are not as dramatic as serious casualties, but since there are many more of them, they do represent a public health issue. Casualties involved in crashes without a third party are under-represented, and even more so for some road-user types (motorcyclists and cyclists) leading to distorted comparisons between road user types. Traffic safety comparisons across geographical areas are also distorted because of the selection bias on road type, environment (urban/rural) and police force. Any result of descriptive or etiologic study based on police crash data may be quite misleading. Furthermore, these biases may lead to setting the wrong targets in prevention campaigns. For instance it has been shown in the US (Abdel-Aty et al., 2005) that analyses of road design effect on crash types provide different results whether they are based on long-form police reports (crashes involving injury or felony) only, or on both long-form and short-form reports (mostly property-damage only crashes).

These words of caution raise the issue of obtaining unbiased estimates of the number of casualties, according to their injury severity, and according to a number of characteristics. This is part of our objectives for the near future. As there is no national registry of road trauma casualties in France, we are working on extrapolating the Rhône registry. We aim at providing nation-wide estimates of the number of NISS slight and serious casualties as a whole and according to characteristics such as age, gender, road user type, as well as to provide a nation-wide estimate of the distribution of injury types. The first step is to model the relationship between the Rhône road trauma registry and the Rhône police data and to estimate both police severity classification biases (to be presented) and police selection biases that account for under-reporting (presented here). From the estimation of these biases, we will construct correction factors to be applied to the nation-wide police data; more precisely these correction factors will give different weights to police data depending on their characteristics: severity, road user type, third party involvement, reporting police force... i.e. characteristics found to be

selection and/or classification bias factors. We will hence obtain a nation-wide extrapolation of the Rhône road trauma registry.

Acknowledgements
We wish to thank the following people for having participated in the data collection and data entry, in the framework of the Association for the Registry of road traffic casualties in the Rhône (ARVAC, president Banssillon V, director Ndiaye A, scientific consultant for the registry, Laumon B) and in the framework of INRETS-UMRESTTE: Ait Idir T, Ait Si Selmi T, Alloatti D, Andrillat M, Artru F, Asencio Y, Assossou I, Auzaneau F, Bagès-Limoges F, Bagou G, Balogh C, Banssillon G, Barnier N, Barth X, Basset M, Bec JF, Bejui J, Bel JC, Bérard E, Bérard J, Bernard JC, Berthet N, Bertrand JC, Besson L, Biot B, Biot V, Blanc C, Blanchard J, Bœuf C, Boisson D, Bonjean M, Bost J, Bouchedor C, Bouletreau P, Boyer V, Breda Y, Bret P, Brilland R, Bussery S, Cabet N, Caillot JL, Cannamela A, Caregnato B, Carre M, Catala Y, Chagnon PY, Chantran C, Chardon P, Charnay P, Chatelain P, Chattard S, Chavane H, Chazot G, Chettouane I, Chevreton N, Chevrillon E, Chevrillon S, Chiron M, Chotel P, Cochard P, Combe C, Contamin B, Coppard E, Cot T, Crettenet Z, Cristini A, Dal Gobbo B, De Angelis MP, Decourt L, Delfosse A, Demazière J, Deruty R, Desjardins G, Dohin B, Emonet A, Escarment J, Eyssette M, Fallavier L, Fanton L, Felten D, Feuglet P, Fifis N, Figura J, Fisher G, Fischer LP, Floccard B, Floret D, Fournier G, Fredenucci JF, Freidel M, Gadegbeku B, Galin L, Gaillard P, Gallon M, Garnier N, Garzanti A, Gaussorgues P, Gautheron V, Genevrier M, Gibaud F, Gillet Y, Goubsky A, Granger M, Grattard P, Gueniaud PY, Guenot C, Guignand M, Haddak M, Hamel D, Heckel T, Jacquemard C, Joffre T, Kohler R, Lablanche C, Lafont S, Lagier C, Lapierre B, Laplace MC, La Rosa C, Laurent R, Lebel M, Leblay G, Le-Xuan I, Lieutaud T, Lille R, Linné M, Lucas R, Machin B, Maello E, Malicier D, Mangola B, Marduel YN, Marie-Catherine M, Martin JL, Martinand G, Marty F, Messikh C, Meyer F, Meyrand S, Molard S, Morel-Chevillet E, Mioulet E, Minjaud F, Mollet C, Monnet J, Moreno S, Moyen B, Ndiaye A, Neidhart JP, Ngandu E, Ny S, Ould T, Paget P, Paillot JC, Paris D, Patay B, Pauget P, Peillon D, Perrin-Blondeau D, Petit P, Piton JL, Plantier M, Pornon P, Pramayon C, Quelard B, Rezig M, Rigal F, Robert D, Rode G, Romanet JP, Rongieras F, Roset C, Rousson A, Roussouli P, Roux H, Ruhl C, Salamand J, Sametzky P, Sayegh K, Sbraire N, Scappaticci N, Schiele P, Schneider M, Simonet C, Sindou M, Soldner R, Soudain M, Stagnara J, Stamm D, Suc B, Taesch MC, Tasseau F, Tell L, Thomas M, Tilhet-Coartet S, Toukou JC, Trifot M, Vallet G, Vancuyck A, Vergnes I, Verney MP, Voiglio EJ, Vourey G, Vuillard J, Westphal M, Willemen L.

References

Abdel-Aty, M., Keller, J., Brady, P., 2005. Analysis of the types of crashes at signalized intersections using complete crash data and tree-based regression. In: proceedings of the 84th annual meeting of the Transport Research Board, Washington DC.

Alsop, J., Langley, J., 2001. Under-reporting of motor-vehicle traffic crash victims in New-Zealand. Accident Analysis and Prevention 33(3), 353-359.

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(1-2), 101-108.

Charnay, P., Laumon, B., 2002. Le registre des victimes d'accidents de la circulation routière du Rhône [The trauma registry of road casualties in the Rhône county], INRETS, Bron, 0204, p.119.

Clark, D. E., 2004. Practical introduction to record linkage for injury research. Injury Prevention 10(3), 186-191.

Cryer, P. C., Westrup, S., Cook, A. C., Ashwell, V., Bridger, P., Clarke, C., 2001. Investigation of bias after data linkage of hospital admission data to police road traffic crash reports. Injury Prevention 7(3), 234-241.

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(4), 529-537.

Elvik, R., Mysen, A. B., 1999. Incomplete accident reporting; meta-analysis of studies made in 13 countries. Transportation Research Record 1665, 133-140.

Greenland, S., 2004. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. American Journal of Epidemiology 160(4), 301-305.

Hauer, E., Hakkert, A., 1988. Extent and some implications of incomplete accident reporting. Transportation Research Record 1185, 1-10.

Hvoslef, H., 1994. Under-reporting of road traffic accidents recorded by the police at the international level, Public Roads Administration, Norway, p.126.

James, H. F., 1991. Under-reporting of road traffic accidents. Traffic Engineering and Control 32, 573-583.

Langley, J., Dow, N., Stephenson, S., Kypri, K., 2003. Missing cyclists. Injury Prevention 9(4), 376-379.

Laumon, B., Martin, J. L., 2002. Analyse des biais dans la connaissance épidémiologique des accidents de la route en France [Bias in the epidemiological knowledge of road crashes in France]. Revue d'Epidémiologie et de Santé Publique 50(3), 277-285.

Laumon, B., Martin, J. L., Collet, P., Chiron, M., Verney, M. P., Ndiaye, A., Vergnes, I., 1997. A French road accident trauma registry: first results. In: 41st annual proceedings of the Association for the Advancement of Automotive Medecine, Orlando, Florida, pp. 127-137.

Lindquist, K., 2004. How to estimate relative risk in SAS using Proc Genmod for common outcome in cohort studies. Accessed December 9, 2004, http://www.ats.ucla.edu/stat/sas/faq/relative_risk.htm.

Lopez, D. G., Rosman, D. L., Jelinek, G. A., Wilkes, G. J., Sprivulis, P. C., 2000. Complementing police road-crash records with trauma registry data- an initial evaluation. Accident Analysis and Prevention 32(6), 771-777.

Martin, J. L., Lafont, S., Chiron, M., Gadegbeku, B., Laumon, B., 2004. Differences between males and females in traffic accident risk in France. Revue d'Epidémiologie et de Santé Publique 52(4), www.e2med.com/resp.

Osler, T., Baker, S., Long, W., 1997. A modification of the Injury Severity Score that both improves accuracy and simplifies scoring. The Journal of Trauma: Injury, Infection, and Critical Care 43(6), 922-926.

Rosman, D. L., 2001. The Western Australian road injury database (1987-1996): ten years of linked police, hospital and death records of road crashes and injuries. Accident Analysis and Prevention 33(1), 81-88.

Simpson, H., 1996. Comparison of hospital data and police casualty data: a national study, Transport Research Laboratory, UK, Crowthorne, TRL report 173, p.37.

Spiegelman, D., Hertzmark, E., 2005. Easy SAS Calculations for Risk or Prevalence Ratios and Differences. American Journal of Epidemiology 162(3), 199-200.

Stutts, J., Hunter, W., 1998. Police reporting of pedestrians and bicyclists treated in hospital emergency rooms. Transportation Research Record 1635, 88-92.

Zou, G., 2004. A modified poisson regression approach to prospective studies with binary data. American Journal of Epidemiology 159(7), 702-706.

Exemples de lettre et questionnaire envoyés dans le cadre de l'enquête « Registre seulement »

Papier à entête de IUMRESTTE

Paul DUPONT

42 RUE D'INKERMANN
69006 LYON

Nos réf. 200108020001/1 RU-GE2

Bron, le 24 Mai 2004

Mr Paul DUPONT,

Notre laboratoire de recherche travaille en sécurité routière, afin de mieux comprendre comment les accidents surviennent, quels types de blessures les personnes subissent, et comment se passe la prise en charge des blessés. Le but à terme est d'améliorer la prévention et la prise en charge des victimes d'accidents de la route.

Notre laboratoire travaille en collaboration avec l'Association pour le Registre des Victimes d'Accidents de la Circulation routière du Rhône (ARVAC). Ce registre recense les victimes de la route, et leurs blessures, des plus graves aux plus légères. Il fonctionne grâce à la participation des services médicaux hospitaliers du département.

C'est dans ce cadre que nous avons eu connaissance que vous avez été blessé lors d'un accident en février 2001. Veuillez nous excuser de raviver des épisodes douloureux, mais votre participation est primordiale pour faire avancer les connaissances en sécurité routière.

Notre but est de mieux connaître les interventions des pompiers, SAMU, et forces de l'ordre lors d'un accident de la route, et, lorsqu'ils n'interviennent pas, de mieux comprendre pourquoi.

Nous vous prions donc de bien vouloir répondre au questionnaire ci-joint. Une fois le questionnaire complété, merci de l'envoyer grâce à l'enveloppe 'T' ci-jointe (sans mettre de timbre).

Les renseignements que vous fournirez seront traités de façon confidentielle ; ils ne seront communiqués à aucune structure juridique, policière, d'assurance, ou autre. Lors de l'analyse, les informations seront traitées de façon anonyme, et en aucun cas de façon individuelle. Le Registre a reçu l'autorisation de la CNIL (Commission Nationale de l'Informatique et des Libertés).

Par avance, nous vous remercions de l'attention que vous porterez à ce courrier, et nous vous prions d'accepter nos salutations les meilleures. Si vous souhaitez des informations complémentaires, vous pouvez nous joindre au 04 72 14 25 33.

Emmanuelle Amoros
Chercheur, Responsable de l'étude

Bernard Laumon
Médecin, Responsable du Registre

Pour information : journées « portes ouvertes » de l'INRETS les 16 et 17 Octobre 2004, sur le site de Bron

Votre identité :

Paul DUPONT sexe : masculin

Indiquez votre date de naissance : Jour | Mois | Année

Accident dans lequel vous avez été blessé(e) en <u>février 2001</u> :

Merci de complèter ou de confirmer les informations ci-dessous

Indiquez la date de l'accident : Jour | Mois | Année

Indiquez l'heure : …..….. (même approximative)

Votre accident s'est produit sur la commune de VILLEFRANCHE-SUR-SAONE
Est-ce exact ? ☐ oui ☐ non ; si non, corrigez : ……

Précisez le lieu de l'accident (arrondissement, quartier, croisement entre rue …. et rue …., rue …, juste avant…, départementale N°… , lieu-dit …., périphérique… , autoroute… vers sortie ….) :

……

……

Vous étiez en voiture
Est-ce exact ? ☐ oui ☐ non ; si non, corrigez : …….……

Vous êtes entré(e) en collision avec : ……..……

Intervenants sur les lieux de l'accident :

<u>Les services suivants ont-ils été **appelés** ?</u>

pompiers	☐ oui	☐ non	☐ vous ne savez pas
SAMU (ou SMUR)	☐ oui	☐ non	☐ vous ne savez pas
police (ou gendarmerie)	☐ oui	☐ non	☐ vous ne savez pas

<u>Ces services sont-ils venus sur place ?</u>

pompiers	☐ oui	☐ non	☐ vous ne savez pas
SAMU (ou SMUR)	☐ oui	☐ non	☐ vous ne savez pas
police (ou gendarmerie)	☐ oui	☐ non	☐ vous ne savez pas

suite au dos

Si les secours médicaux (pompiers, SAMU) n'ont pas été appelés, indiquez pourquoi, selon vous :
(vous pouvez cocher plusieurs cases)

☐ cela n'a pas été jugé utile, car :
- ☐ personne ne paraissait avoir de blessure grave
- ☐ autre raison ; précisez : …….…….

☐ personne n'a pensé à le faire
☐ réticence à les appeler
☐ autre raison ; précisez, si vous le souhaitez : ……..…….
☐ vous ne savez pas

Si la police (ou gendarmerie) n'a pas été appelée, indiquez pourquoi, selon vous :
(vous pouvez cocher plusieurs cases)

☐ cela n'a pas été jugé utile car :
- ☐ il n'y avait pas d'autre personne impliquée dans l'accident (pas d'antagoniste)
- ☐ un constat amiable d'accident avait été rempli
- ☐ personne ne paraissait avoir de blessure grave
- ☐ autre raison ; précisez : ……..

☐ personne n'a pensé à le faire
☐ réticence à les appeler
☐ autre raison ; précisez, si vous le souhaitez :……..……
☐ vous ne savez pas

Si la police (ou gendarmerie) est venue, a-t-elle établi un procès-verbal ? (c'est-à-dire un constat officiel, autre que le constat amiable, et qui doit en principe être rédigé lorsqu'il y a au moins un blessé)

☐ oui ☐ non ☐ vous ne savez pas

Si la police (ou gendarmerie) est venue, avez-vous été satisfait(e) de son intervention ?

rapidité :	☐ satisfaisant	☐ non satisfaisant
prise en charge :	☐ satisfaisant	☐ non satisfaisant

Si les pompiers sont venus, avez-vous été satisfait(e) de leur intervention ?

rapidité :	☐ satisfaisant	☐ non satisfaisant
prise en charge :	☐ satisfaisant	☐ non satisfaisant

Si le SAMU (ou SMUR) est venu, avez-vous été satisfait(e) de son intervention ?

rapidité :	☐ satisfaisant	☐ non satisfaisant
prise en charge :	☐ satisfaisant	☐ non satisfaisant

Précisez, si vous le souhaitez, en quoi vous avez été satisfait(e) ou mécontent(e) de ces interventions :

……

……

……

Le questionnaire est terminé. Nous vous remercions vivement de votre participation.

Extraits de programme SAS : analyse des réponses au questionnaire en tenant compte du plan de sondage

```
*manipulation pr créer variable strate, pour le calcul des poids;
data enquete ;
set enquete;
if put(antagoniste, $ftiers.)= 'avec tiers' then tiers=1;
   else if put(antagoniste, $ftiers.)= 'sans tiers' then tiers=2;
if put(MAIS, f2smais.)='MAIS 1' then MAIS2cat=1;
   else if put(MAIS, f2smais.)='MAIS 2+' then MAIS2cat=2;
if put(blesse, $f5typeu.)='piéton ou roller' then uc=1;
   else if put(blesse, $f5typeu.)='vélo' then uc=7;
   else if put(blesse, $f5typeu.)='deux roues motorisée' then uc=6;
   else if put(blesse, $f5typeu.)='voiture' then uc=2;
   else if put(blesse, $f5typeu.)='autre/mq' then uc=9;
*création de la variable strate;
strate= put(MAIS2cat, 1.)||put(uc, 1.)||put(tiers,1.);
run;

proc freq data=enquete ;
tables strate ;
run;


*calcul des poids correspondant au plan de sondage stratifié;
data RegSeulTotals;
input  MAIS2cat uc tiers _total_ sample;
datalines;
1 1 1 391 89
1 7 1 177 41
1 7 2 605 138
1 6 1 450 103
1 6 2 672 153
1 2 1 2420 550
1 2 2 844 192
1 9 1 57 13
1 9 2 256 59
2 1 1 124 65
2 7 1 49  26
2 7 2 311 162
2 6 1 156 82
2 6 2 328 171
2 2 1 254 132
2 2 2 220 115
2 9 1 11 6
2 9 2 87 46
;
*1 1 2 0 0
2 1 2 0 0  ;

run;

data regseultotals;
set regseultotals;
strate= put(MAIS2cat, 1.)||put(uc, 1.)||put(tiers,1.);
tot=_total_;
f=sample/tot;
w=1/f;
if strate in ('112','212') then delete;
run;
```

```
proc format;
value fanta
1='avec tiers'
2='sans tiers';

proc print data=regseultotals;
format uc f5uc. tiers  fanta. ;
run;


*merge avec enquete pour fournir les poids;
proc sql;
create table enquete as
select r.*, t.f, t.w
from enquete r
 join regseultotals t on (r.strate=t.strate);

proc freq data=enquete;
table uc*tiers;
format uc f5uc. tiers  fanta. ;
run;


*****************************************************************************;
* definition des enquêtés inclus ds analyse;

data enquete;
set enquete;
if rep=1 and special='' then inclus=1 ;
run;

*****************************************************************************;


proc format;
value $fyndet
'1'='oui'
'2'=' non'
'8'=' sans objet'
'9'=' NSP'
'99'=' NR';
value $fyn
'1'='oui'
'2'=' non'
'8','9','99'=' NSP, NR...';
;

* analyse police_appel ;
proc surveymeans data=enquete total=regseultotals ;
stratum uc tiers mAIS2cat;
class police_appel  uc tiers  mAIS2cat;
format police_appel $fyn.;
var police_appel;
weight w;
domain inclus;
run;

* analyse police_appel, selon variables de stratification ;
proc surveymeans data=enquete total=regseultotals;
stratum uc tiers mAIS2cat;
```

```
class police_appel uc tiers mAIS2cat;
format police_appel $fyn. uc f5uc. ;
var police_appel;
weight w;
domain inclus*uc;
run;

proc surveymeans data=enquete total=regseultotals;
stratum uc tiers mAIS2cat;
class police_appel uc tiers mAIS2cat;
format police_appel $fyn. tiers fanta. ;
var police_appel;
weight w;
domain inclus*tiers;
run;

proc surveymeans data=enquete total=regseultotals;
stratum uc tiers mAIS2cat;
class police_appel uc tiers mAIS2cat;
format police_appel $fyn. mAIS2cat f2smais.;
var police_appel;
weight w;
domain inclus*mais2cat;
run;
*analyse de police-PV;
proc surveymeans data=enquete total=regseultotals;
stratum uc tiers mAIS2cat;
class pol_PV uc tiers mAIS2cat;
format pol_PV $fyn. ;
var pol_PV;
weight w;
domain inclus;
run;

[…]


*******************************************************************************;
* pour estimer  les proportions de répondants hors critères BAAC;

* definition des enquêtés inclus ds analyse;

data enquete;
set enquete;
if rep=1  then inclus=1 ; *mis à jour;
run;

proc freq data=enquete;
tables inclus/missing;
run;


proc univariate data=enquete (where=(inclus=1)) ;
var w;
run;


* analyse proprotion de spécial ie hors critères;
title 'proportions de special ';
proc surveymeans data=enquete total=regseultotals ;
```

```
stratum uc tiers mAIS2cat;
class special  uc tiers  mAIS2cat;
format special $special. ;
var special;
weight w;
domain inclus;
run;

* analyse proprotion de spécial, selon variables de stratification, ici type usager ;
proc surveymeans data=enquete total=regseultotals;
stratum uc tiers mAIS2cat;
class special uc tiers mAIS2cat;
format special $special.  uc f5uc. ;
var special;
weight w;
domain inclus*uc;
run;

**********************************************************************************************;

* analyse multivariée pour comparer repondants a tous les autres;

*permettra de donner les caracteristiques des autres que repondants;

proc format ;
value  rep_autres
1,7='répondants'
2,3,4,5,6,8,9 ='non rép et adr mq/NPAI'
;

proc surveylogistic data=retour  total=regseultotals  ;
stratum uc tiers mAIS2cat;
class retour_code uc tiers(desc) mAIS2cat mais agec sexe reseau acc_urb eclairement zonefo zus
adr_zone adr_courly/ param=glm;
format retour_code rep_autres. mais f3mais. reseau $f6myres. tiers fanta. agec  f5agec. uc f5uc.
adr_zone zone.  ;
* ref= ;
model retour_code= tiers  ;
weight w;
ods output ParameterEstimates=est;

run;

data est;
set est;
OR=exp(estimate);
low=exp(estimate-1.96*StdErr);
up=exp(estimate+1.96*StdErr);
run;
proc print data=est;
var variable classval0  OR low up;
run;
```

Synthèse bibliographique sur les erreurs de classement de la gravité des blessures par les forces de l'ordre

Cette bibliographie contient deux tableaux, l'un portant sur les erreurs de classement proprement dites, l'autre, sur les caractéristiques associées à ces erreurs

Tableau 5 : études sur la concordance du classement de la gravité des victimes de la circulation routière par les forces de l'ordre avec un classement médical, présentées par année décroissante

Auteur, année	pays	restriction	Effectif	méthode	Résultats chiffrés	conclusion	Chaînage ?
(Morris, Mackay et al. 2003)	Royaume-Uni, sondage	occupants de voiture	13 600	classement policier de gravité (indemne, blessé léger, grave, tué) vs MAIS et ISS	34% de ceux classés « blessé grave » par la police sont en fait MAIS 1	« faible concordance ; les catégories de gravité de la police ne correspondent pas assez aux niveaux d'AIS »	Échantillon de données policières, avec codage ultérieur selon l'AIS
(Farmer 2003)	USA, échantillon NASS[2]	Conducteurs de véhicules particuliers motorisés ?	9 900	Classement MAIS vs classement policier (KABCO)	49% des conducteurs classés « avec blessures handicapantes » par la police étaient MAIS 0 ou 1	« le classement de la gravité est trop inexact pour la recherche appliquée »	Oui (rapports de police avec NASS)
(Oviève 2003)	France, Rhône (1998-2000)		7 900	Classement policier vs durée d'hospitalisation (0-6 jours vs 6+)	Se=0,68, VPP=0,64, Sp=0,95, VPP=0,96	« le critère de gravité des BAAC ne semble donc pas très satisfaisant »	oui
(Laumon and Martin 2002)	France, Rhône (1996)		3 000	Classement policier (graves vs légers) vs MAIS (évalué sur les données du Registre)	3,4% des blessés légers (critère policier) sont en fait MAIS 3, et 54,3% des blessés graves (critère policier) sont MAIS 1 ou 2. En considérant MAIS 3+ comme seuil de référence, le classement policier est évalué : Se=85,6%, Sp=80,3%, VPP=45,7%	« la notion de blessé grave habituellement retenue [dans les données policières] s'avère biaiser, sinon la réalité accidentologique, du moins la perception que l'on peut en avoir »	oui

[2] NASS= National Automotive Sampling System

Auteur, année	pays	restriction	Effectif	méthode	Résultats chiffrés	conclusion	Chaînage ?
(Lopez, Rosman et al. 2000)	Australie, ouest	Exclus : piétons et cyclistes se blessant seuls	400	Comparaison du classement policier et de l'hospitalisation (oui/non)	Le classement policier est correct pour 78% des blessés		Oui (police avec registre hospitalier)
(Aptel, Salmi et al. 1999)	France, la Réunion		1 100	Classement policier (graves vs légers) vs données hospitalières : durée d'hospitalisation (0-6 jours vs 6+)	Sensibilité=77%, spécificité=61%	« les niveaux de gravité policières ne sont pas en concordance avec les durées effectives d'hospitalisation ; sur-estimation de la gravité »	oui
(Austin 1995)	UK, Humberside		1 000	classement policier (catégories de blessés légers et graves) vs données hospitalières : hospitalisation (oui/non) et description des blessures	9,3% des blessés classés « graves » par les forces de l'ordre devraient être classés comme légers ; 13,2% de ceux classés « légers » devraient classés comme graves	« le nombre de blessés graves devrait être augmenté de 34,7% ; la gravité lésionnelle devrait être codée d'après les données hospitalières	oui
(Hansen, Popkin et al. 1991)	USA, North Carolina		1 300	Comparaison du classement KABCO effectué par la police et par des médecins séparément	Concordance chez 49,5% des victimes ; coef. Spearman=0.57 ; Se=74%, Sp= 78%, VPP=57%, VPN=88%	« les résultats sont en faveur de l'utilisation des données policières de gravité »	Oui (police et « trauma center »)
(Agran, Castillo et al. 1990)	USA, California	Enfants piétons ou cyclistes heurtés par un véhicule motorisé	140	Comparaison entre classement policier et ISS (évalué à l'hôpital)	Coef. corrélation de Spearman=0,50. 50% des blessés graves (au sens ISS) sont mal classés par police	Faible corrélation entre les deux classements ; les blessés graves au sens ISS sont souvent mal classés	oui

Auteur, année	pays	restriction	Effectif	méthode	Résultats chiffrés	conclusion	Chaînage ?
(Tunbridge and Everest 1988)	Royaume-Uni, Oxfordshire		5 600	classement policier (léger et grave) vs MAIS (0-1 vs 2+)	aucun	"la concordance est bonne ; les regroupements de MAIS donne une bonne approximation des niveaux de gravité utilisées par le ministère de transports »	Il semblerait (rien d'explicite)
(Saunders and Wheeler 1987)	Royaume-Uni, Dorset (généralistes)		21	classement policier (en légers et graves) vs description des blessures par des généralistes	2 sur 21 sont incorrectement classés	-	oui
(Barancik and Fife 1985)	USA, Ohio		500	Proportion de ceux classés indemnes par la police, en fonction du type de soins (hospitalisation (oui/non))	Le sous-classement en « indemne » concerne 5% des hospitalisés et 17% de ceux traités aux urgences (et non-hospitalisés)	« les limites des données accidentologiques policières doivent être considérées à chaque utilisation »	oui
(Shinar, Treat et al. 1983)	USA, Indiana		200	Comparaison des données policières et étude détaille d'accidents ;	Coef. incertitude=0,25 (loin de 1) ; gravité sous-estimée dans 30% des cas (classés à tort « matériels » au lieu de « corporel »)	« les données policières de gravité sont inadéquates ; la gravité était sous-estimée dans plus de 30% des cas »	oui
(Sande and Thorson 1975)	Suède, région d'Uppsala		2 400	classement policier vs hospitalisation (oui/non) et durée d'hospitalisation		Concordance faible	
(Bull and Roberts 1973)	Royaume-Uni, Birmingham		1 200	Classement policier vs classement par l'hôpital en fonction de l'échelle policière (légers et graves)		« pour un petit nombre de cas, le classement est erroné »	oui

Tableau 6 : études sur les caractéristiques liées aux concordances ou discordances entre classements policier et médical sur la gravité des blessés

Auteur, année	pays	restriction	Effectif	méthode	Facteurs associés	Autres facteurs étudiés, NS	Résultats chiffrés	conclusion	Chaînage ?	notes
(Farmer 2003)	USA, échantillon (NASS)	Conducteurs de véhicules particuliers motorisés ?	9 900	Proportion de MAIS 0-1 en fonction de la catégorie KABCO (classement policier)	Gravité (MAIS), région, jour/nuit, type de collision, sexe et âge du conducteur	Type de jour (semaine/week-end), année de mise en circulation du véhicule (?)	% de MAIS 0-1 parmi ceux avec « blessures immobilisantes » : 53% chez les femmes vs 44% chez les hommes ; 37% chez les 65 ans et + mais 50% chez les 25-44 ans	« le classement de la gravité est trop inexact pour la recherche appliquée »	oui (rapports de police avec NASS)	Univarié semble-t-il
(Ovième 2003)	France, Rhône (1996)		7 900	Concordance classement policier et durée d' hospitalisation	âge, sexe, type d'usager, type de réseau, jour/nuit	tiers	OR femmes =1,43 vs hommes, OR piétons=0,50, 2-roues motorisé : 0,49, vs automobilistes		oui	

Auteur, année	pays	restriction	Effectif	méthode	Facteurs associés	Autres facteurs étudiés, NS	Résultats chiffrés	conclusion	Chaînage ?	notes
(Laumon and Martin 2002)	France, Rhône (1998-2000)		3 000	Probabilité d'être classé « blessé grave » par la police, en fonction du MAIS ?	MAIS × type d'usager		Parmi les victimes MAIS 1, les piétons ont 80% de chances de plus qu'un automobiliste d'être classé « blessé grave »	« Le fait que l'intensité de ce biais de classement varie selon le type d'usager encourage plus encore à une révision «de la notion de blessé grave telle qu'elle est aujourd'hui définie en France »		
(Lopez, Rosman et al. 2000)	Australie, ouest	Exclus : piétons et cyclistes se blessant seuls	400	Concordance classement police et hospitalisation	Gravité (ISS), sexe	Type usager, nb de véhicules impliqués, agglomération / rural	les hommes ont plus de chances d'être correctement classés que les femmes : 87% vs 76% (OR=2,05) ; les blessés ISS>9 ont plus de chances d'être correctement classés que les blessés ISS 1-4 (OR entre 5,17 et 11,29)	« L'amélioration des rapports d'accidents avec l'information détaillée sur les lésions et sur les soins provenant du registre de traumatologie fournirait des mesures plus fiables »	oui	Univarié, et multivarié (logistique)
(Hansen, Popkin et al. 1991)	USA, North Carolina	?	1 300	Concordance entre classement des forces de l'ordre et classement des médecins (sur même échelle KABCO)	Gravité (KABCO), alcoolémie (sobre/alcoolisé)	-	Concordance des deux classement chez 48% des alcoolisés mais chez 57% des sobres	Concordance meilleure chez les blessés les plus graves, et les plus légers (sur l'échelle à 5 niveaux) ; pas de surévaluation systématique de la gravité des impliqués alcoolisés. « Les résultats sont en faveur de l'utilisation des données policières de gravité et de alcoolémie »	oui	

Auteur, année	pays	restriction	Effectif	méthode	Facteurs associés	Autres facteurs étudiés, NS	Résultats chiffrés	conclusion	Chaînage ?	notes
(Tunbridge and Everest 1988)	Royaume-Uni, Oxfordshire		5 600	Comparaison des distributions marginales des classement police (légers et graves) et MAIS (0-1 et 2+)	-	type d'usager			Oui ? pas explicité	
(Barancik and Fife 1985)	USA, Ohio		500	Sous-classement : « indemne » mais traité à l'hôpital		âge			oui	univarié

Références

Agran, P. F., D. N. Castillo, et al. (1990). "Limitations of data compiled from police reports on pediatric pedestrian and bicycle motor vehicle events." Accident Analysis and Prevention **22**(4): 361-70.

Aptel, I., L. R. Salmi, et al. (1999). "Road accident statistics: discrepancies between police and hospital data in a French island." Accident Analysis and Prevention **31**(1-2): 101-108.

Austin, K. (1995). "The identification of mistakes in road accident records: part 2, casualty variables." Accident Analysis and Prevention **27**(2): 277-282.

Barancik, J. and D. Fife (1985). "Discrepancies in vehicular crash injury reporting: northeastern Ohio trauma study IV." Accident Analysis and Prevention **17**(2): 147-154.

Bull, J. P. and B. J. Roberts (1973). "Road accident statistics--A comparison of police and hospital information." Accident Analysis and Prevention **5**(1): 45-53.

Farmer, C. M. (2003). "Reliability of police-reported information for determining crash and injury severity." Traffic Injury Prevention **4**(1): 38-44.

Hansen, A., C. Popkin, et al. (1991). Comparison of police and physician judgments of injury severity and inebriation in motor vehicle crash victims. 35th annual conference of the Association for the Advancement of Automative Medecine, Toronto, AAAM.

Laumon, B. and J. L. Martin (2002). "Analyse des biais dans la connaissance épidémiologique des accidents de la route en France." Revue d'Epidémiologie et de Santé Publique **50**(3): 277-285.

Lopez, D. G., D. L. Rosman, et al. (2000). "Complementing police road-crash records with trauma registry data- an initial evaluation." Accident Analysis and Prevention **32**(6): 771-777.

Morris, A., M. Mackay, et al. (2003). Some injury scaling issues in UK crash research. international IRCOBI conference on the biomechanics of impact, Lisbon, IRCOBI.

Oviève, F. (2003). Etude du critère "blessé grave" dans les fichiers BAAC des statistiques officielles des victimes d'accident corporel de la circulation routière dans le département du Rhône. Lyon, Université Claude Bernard**:** 70.

Sande, J. and J. Thorson (1975). "An evaluation of the official Swedish statistics on seriously injured in road traffic accidents." Scand J Soc Med **3**(1): 5-11.

Saunders, R. and T. Wheeler (1987). "Reported accidents and the GP." Highways and transportation **34**(12): 20-22.

Shinar, D., J. R. Treat, et al. (1983). "The validity of police reported accident data." Accident Analysis and Prevention **15**(3): 175-191.

Tunbridge, R. and J. Everest (1988). An assessment of the under reporting of road accident casualties in relation to injury severity. International IRCOBI conference on the biomechanics of impact, Cologne, IRCOBI.

Extrait de programme SAS : régression de Poisson modifiée et calcul des probabilités marginales (erreurs de classement)

```
title 'poisson avec estimation robuste de la variance ' ;
proc genmod data=victcm9701 ;
class   id_victi MAIS NISS typeUsag tiers AgeA sexe corps annee courly;
format grav01 GRAV01f.  MAIS f5mais. iss f7iss. NISS f5iss. typeUsag f5typeu. tiers Btiers. courly $f2gl. AgeA f5age. corps f3ordre.  local f5agglo. cadm f5reseau. int f2int. lum flum. typj f2typj. atm f2atm. ;
model grav01 =  NISS  typeusag   corps  annee     / link=log dist=poisson  ;
repeated subject = id_victi /type=unstr;
output   out       = Res_pois
        pred      = pred
        resraw    = ResRaw
        reschi    = Reschi ;
ods output GEEEmpPEst=poiss_est ;
lsmeans NISS typeusag  corps  annee / CL;
ods output LSMEans=lsmeans;
run;

ods trace off;

data poiss_est;
set poiss_est (where=(Parm NE 'Scale'));
RR=exp(estimate);
lowerCL=exp(LowerCL);
upperCL=exp(UpperCL);
format RR lowerCL upperCL 6.2 ;
run;

*ods rtf body='E:\these\biais\multivarié.rtf' style=minimal;
proc print data=poiss_est;
var parm level1    RR  lowerCL upperCL; * level2 level3 ;
run;
ods rtf close;
proc contents data= LSMeans varnum;
run;
data probmeans;
set lsmeans;
prob=exp(estimate);
lowerprob=exp(lowerCL);
upperprob=exp(upperCL);
*lowerCL=exp(estimate-1.96* StdErr);
*upperCL=exp(estimate+1.96* StdErr);
run;

ods rtf body='E:\these\biais\prob ajustées.rtf' style=minimal;
proc print data=probmeans width=minimum;
var effect NISS  typeusag  corps   prob ; *lowerprob upperprob; *que valent ces IC???.;
run;
ods rtf close;
```

Article publié dans Journal of Trauma, version « pré-print »

Amoros E, Martin JL, Chiron M, Laumon B, Road crash casualties: characteristics of police injury severity mis-classification. *The Journal of Trauma: Injury, Infection, and Critical Care,* 2007, 62(2), 482-490

Abstract

Background: In most countries, epidemiological knowledge of road crash injury is mainly based on police data, as they very often are the only available data at the nation-wide level. However their validity is of some concern. We focus here on the police severity classification of 'serious' and 'slight' casualties in France. We want to know how the police classification compares with a trauma severity scale, so that we could correctly interpret police based studies.
Method: The study is based on the Rhône county (population 1.6 M) over the 1997-2001 period. Police data have been linked with a road trauma registry, so that both police and NISS classifications are available on 14342 casualties. The police classification of 'slight' and 'serious' casualties is compared to the registry classification grouped into NISS 1-15 and NISS 16-75 categories. We conduct multivariate analyses of the probability of police severity mis-classification, over and under-classification, as a function of crash and casualty characteristics.
Results: Kappa is estimated at 0.41; the sensitivity of the police classification is 72% and the positive predictive value is 35%. Pedestrians and motorcyclists casualties are the most likely to be over-classified (RR=1.4 and RR=1.2 respectively compared to car occupants). The 'rural police' are more likely to over-classify than the other police forces (RR=3.1). Over-classification decreases over the 1997-2001 period while under-classification increases.
Conclusion: these mis-classification characteristics must be kept in mind when interpreting severity results based on police data. We are working on obtaining unbiased nation-wide estimates of severity figures.

Abbreviations used:
AIS = Abbreviated Injury Scale; NISS= New Injury Severity Score; ISS= Injury Severity Score; MAIS = Maximum AIS severity code; RR = Relative Risk; CI = Confidence Interval;

INTRODUCTION

In most countries, epidemiological knowledge of road crash casualties is mainly based on data collected by the police. Police definition of injury severity is often based on 3 categories: fatal casualties, non-fatal serious and slight casualties. The police definition of slight and serious casualties is based either on hospital admission, hospital stay (length of), type of injury, medical treatment required (length of), disability to work (length of), or a combination of these[1]. Since the police data often are the only road casualties data available at the nation-wide level, and hence the basis for most epidemiological knowledge, their quality must be assessed. Some studies[2-8] have analysed the validity of the police severity classification, i.e. the police classification has been compared with a physician's implementation of the same definition or a very similar one. Other studies[9-14] have analysed the discordance between the police severity classification and a recognised and internationally used trauma classification, based on injury assessment, namely the Abbreviated Injury Scale (AIS). The overall severity of casualties was measured either by the MAIS (Maximum AIS severity code) or by the ISS (Injury Severity Score). A few studies[5, 13, 15] have further explored whether the discordance between police and trauma-based classification was related to some casualty or crash characteristics, such as road user type, gender, injury severity itself.

Police severity definition and police organisation being country-specific, it is necessary that the quality of police severity classification be assessed in every country. In France, the police definition of serious and slight casualties is based on the length of hospital stay with a cut-off point at 6 days. The validity of the French police classification has already been studied : it has been compared to the casualties' actual length of hospital stay in two studies: in the first one[8], conducted in an overseas territory, La Reunion, sensitivity was estimated at 77% and specificity at 61%. In mainland France, in the Rhône county[16], over the 1998-2000 period, sensitivity was estimated at 68%, positive predictive value at 64%, specificity at 95%, and negative predictive value at 96%. Some of these values are quite far from 1; this confirms what we have been told, namely that the police do not always inquire about the length of hospital stays, at it can be quite laborious. Instead they quite often evaluate severity directly on the crash scene or categorise it using the number of days of ITT (temporary total incapacity; this is evaluated by physicians for legal purposes and it is recorded in police reports). Besides, even when length of hospital stay is used, it is not a very good measure of injury severity since it is related to characteristics other than severity itself (individual ones such as age, comorbidity[17]… and broader ones such as hospital policy). We will therefore not study any further the relationship between police classification and length of hospital stay. We focus directly on the relationship between the police severity classification and injury severity itself, as police classification is what is available at the nation-wide level, and injury severity is what we are interested in. We need to know how they relate to each other so that we can correctly interpret, in terms of injury severity, studies based on police data.
Injury severity is measured here by the New Injury Severity Score (NISS), based on the widely used and recognised trauma scale, the Abbreviated Injury Scale (AIS). We further group NISS scores into 2 categories, as close as possible in definition to the police categories of slight and seriously injured, so that we can study discordance between the two classifications. The first part of the paper evaluates the extent of discordance, or in other words police mis-classification. The second part identifies and quantifies how much police mis-classification is related to some crash or casualty characteritics. More specifically it identifies and quantifies how much some categories of casualties and/or crashes are more likely to have their severity over or under-classified by the police.

MATERIALS AND METHODS

This study can be achieved thanks to the existence of a road trauma registry, to which the police data are compared to. The registry provides a complete injury assessment coded with the Abbreviated Injury Scale. The study is conducted on casualties identified as common to the police file and the registry, for whom both classifications are hence available.
The target population consists of non-fatal casualties from road crashes occurring in France. We focus on non-fatal casualties and leave out fatal casualties as fatalities benefit from rather reliable figures and rather clear definition**.** The source population consists of non-fatal casualties from road crashes that occurred in the Rhône county, and that have been reported by both the police and the Rhône road

trauma registry, during the 1997-2001 period. The Rhône county is a large county of 1.6 million inhabitants; it consists of a large city, Lyon, its suburbs and a rural area in the north part. The study population includes 14342 subjects.

Police traffic crash data

The French police are required by law to write a crash report for every road crash causing at least one casualty. A road crash is defined as a crash occurring on the network open to public traffic and involving at least one vehicle. Skateboard or roller skate users are considered as pedestrians by the police, and are only considered as road casualties if hit by a vehicle.

The crash report contains detailed information on the crash: crash environment, conditions, vehicles involved, but information on the people involved is limited. The definition of injury severity of non-fatal casualties is based on the length of hospital stay: casualties requiring a hospital stay of 6 days or more are categorised as 'seriously injured', whereas casualties requiring less than 6 days of hospital stay are categorised as 'slightly injured'.

There are three main police forces in France (whose French names are 'police', 'gendarmerie' and 'Compagnie Républicaine de Sécurité') and we wish to check whether severity classification differs between those. These 3 police forces can be named 'urban police', 'rural police' and 'urban motorways police' according to their area of control. It is however not clear-cut: the 'urban police' mostly operate in cities of a certain size; the 'rural police' mostly operate in rural areas but also in some part of large city suburbs; the 'urban motorways police' operate on motorways in and around large cities.

The Rhône road trauma registry

A road trauma registry[18, 19] has been in use since 1995 in the Rhône county. The registry is based on the participation of all public and private health care facilities of the county and its close surroundings that may receive victims of a traffic crash: it includes some 150 health care facilities, from emergency departments, intensive care units, surgery... to rehabilitation departments. It covers all casualties from road crashes occurring in the Rhône county that seek care in health facilities. Both inpatients and outpatients are registered, i.e. all casualties are registered whether they are hospitalised or not. This registry has been approved by the French National Registry Committee.

Information collected for every casualty consists of a few crash characteristics and of the following casualty characteristics: gender, date of birth, place of residence, hospital stay, hospital transfer if relevant, and injury assessment. For every subject, injury assessment is based on the whole set of diagnoses provided by the different health services the subject has gone through. Plain text diagnoses are coded by the registry physician according to the Abbreviated Injury Scale (AIS), 1990 revision. Moreover, each injury is assigned a severity code, ranging from AIS 1 (minor) to AIS 6 (currently beyond treatment). The severity of a casualty with multiple injuries can be measured with the MAIS, the ISS[20] or with the NISS[21], and there has been quite some debate about the choice of severity measure[22-29]. The ISS and the NISS take better account of multiple injuries than the MAIS, and the NISS has no constraint about accounting only for the worst injuries from different body regions. The NISS seems to correlate better with the probability of survival[21-24]. We hence choose to use the NISS as the reference measure to compare the police classification to. Nonetheless, we also repeat the analyses with the ISS and with the MAIS to check whether the results are sensitive to the chosen AIS-based severity scale.

Data linkage

Comparison of police severity classification with the NISS severity classification is conducted on casualties on whom both classifications are available, that is to say on casualties identified as common to both sources. Police and registry casualty data have been linked using a semi-automated record-linkage procedure[30, 31]. Linking variables are date and time of crash, crash location, type of road user, date of birth (only year and month) and gender. The procedure was not fully automated since a major linking variable (crash location) was left as unformatted free text to avoid losing any valuable information.

Analysis

The first part of the analysis consists of studying the extent and direction of the discordance between the police severity classification and the NISS classification. Firstly, we compute in each NISS category, the proportion of casualties that have been classified as 'seriously injured' by the police. We use 5 NISS categories, with the choice of cut-off points corresponding to the jumps in AIS for a single injury: NISS 1-3, 4-8, 9-15, 16-24, and 25-75. These are the same as Rosman[32] used to group ISS. Secondly, by grouping the NISS scores into two, we estimate the sensitivity, specificity, positive and negative predictive values of the police severity classification. The cut-off point of a NISS of 16 was chosen, corresponding to a single AIS 4 injury, as the closest equivalent, by construction, to the police cut-off point defined at 6 days of hospital stay. This choice is based on the descriptive analysis[33] of injuries recorded in the Rhône road trauma registry, which indicates that AIS 3 injuries are most often displaced fractures, and that AIS 4 injuries are most often head injuries and thorax injuries (bilateral lung contusion). The former do not usually lead to 6 days or more of hospital stay, whereas the latter do. This is supported by another study[34] based on the Rhône trauma registry, which shows that the length of hospital stay median is 4 days among ISS 9-15 casualties and 13 days among ISS 16-24 casualties. Hence the choice of cut-off point at NISS 16, so that the NISS 1-15 and NISS 16-75 categories are the closest equivalent, in terms of definition, to the police 'slight' and 'serious' severity categories. This leads us to switch terminology from 'discordance' to 'mis-classification' and further to 'over-classification' and 'under-classification'. Over-classification is used for NISS 1-15 casualties classified as 'seriously injured' by the police, and under-classification for NISS 16-75 casualties classified as 'slightly injured' by the police.

The second part of the analysis consists of identifying crash and casualty characteristics associated with police severity mis-, over- and under-classification, and quantifying the extent of these associations. This is done respectively by modelling 1) the probability of discordance between the two classifications 2) the probability among NISS 16-75 casualties of being classified as 'slightly injured' by the police, and 3) the probability among NISS 1-15 casualties of being classified as 'seriously injured' by the police. We model each of these probabilities as a function of casualty, crash, and police characteristics. Potentially related characteristics that we explore are 1) casualty characteristics: injury severity itself (5 NISS categories), casualty age (5 categories), gender, type of road user, 2) crash characteristics: whether it involved a (human) third party or not, road type, environment (in-town vs. out-of-town, or town size), type of day (beginning of the week vs. end of the week), light conditions (daylight vs. darkness), weather conditions (good vs. bad), junction (yes vs. no), and lastly 3) external characteristics: calendar year and reporting police force. We first conduct univariate analyses: analyses of the association of each characteristic with mis/under/over-classification. In the construction of the multivariate model, we consider for inclusion those characteristics that are either found relevant in the literature or significant (at the 5% error level) in the univariate analyses. We finally keep those that are significantly associated with the outcome, adjusted for the other factors.

In order to measure the extent of the association of a characteristic with the outcome, we favour relative risks (RR) to odds ratios as they are much easier to interpret. One can not use here the usual approximation of RR estimates by odds ratio estimates (easily obtained from logistic regression). The required condition is not satisfied: the outcome - mis, under or over-classification - is not a rare event. In this case, different model-based methods can be used to estimate relative risks [35, 36]: log-binomial, logistic with some transformations, and a modified Poisson regression. We choose the latter i.e. Poisson regression with robust error variance estimation[37] and we conduct it with GENMOD procedure (generalized linear models), SAS software (SAS Institute Inc., Cary, NC)[35, 38].

We also provide results in terms of adjusted (standardised) probabilities from the final multivariate model. These correspond to marginal probabilities for a given characteristic, if the data were balanced. These probabilities are computed by standardising the data over the other bias factors (more precisely by averaging over levels of the other categorical variables, setting equal weights). This is achieved using least-squares means (LSMEANS) option in GENMOD procedure.

RESULTS

In the Rhône county, on the 1997-2001 period, the police files include 23080 casualties, the registry 54661. Record-linkage has lead to 14974 casualties identified as common to both sources. Of these linked casualties, 14446 are categorised as non-fatal casualties (according to the police classification). Some 104 (0.7%) casualties are further excluded because of missing NISS (recorded in the registry but with no or incomplete injury assessment). The study is hence based on 14342 subjects.

Discordance between police severity and NISS classifications

Casualties' probability of being classified as 'seriously injured' by the police in each NISS category is displayed in figure 1. This probability increases with increasing NISS. Similar distributions are obtained when displaying the percentages according to the ISS or according to the MAIS categories (data not shown).

Comparing the police classification of 'slight' and 'serious' to the 2 groups of NISS 1-15 and NISS 16-75 casualties (table 1), the Kappa measure of concordance is estimated at 0.41. Sensitivity is estimated at 71.7% (C.I. 69.1% - 74.4%); in other words, slightly more than two thirds of NISS 16-75 casualties are classified as 'seriously injured' by the police. Conversely, positive predictive value is estimated at 35.0% (C.I. 33.0% - 37.0%); i.e. among those classified as 'seriously injured' by the police, one third are NISS 16-75 casualties. Specificity is estimated at 89.1% (C.I. 88.6% - 89.6%), that is to say, most NISS 1-15 casualties are classified as 'slightly injured' by the police. Conversely, negative predictive value is estimated at 97.5% (C.I. 97.2% – 97.8%); i.e. among those classified as 'slightly injured' by the police, almost all of them are NISS 1-15 casualties.

Figure 1. Proportion of non-fatal casualties classified as 'seriously injured' vs. 'slightly injured' by the police, according to NISS category (14342 casualties common to police and registry data, Rhône county, 1997-2001)

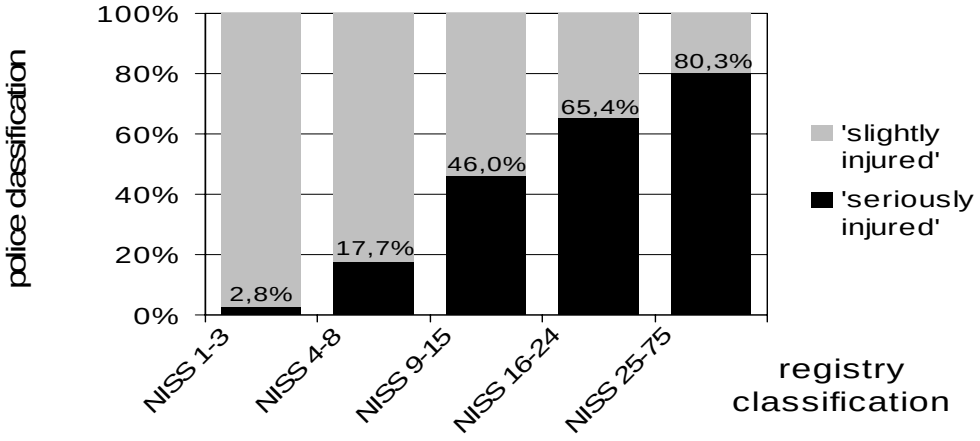


Table 1: distribution of non-fatal casualties according to the police and to the registry severity classifications, Rhône county, 1997-2001

	Registry classification		
Police classification	NISS 1-15	NISS 16-75	total
'slightly injured'	11815	306	12121
'seriously injured'	1444	777	2221
total	13259	1083	14342

Mis-classification characteristics

In the univariate analyses, almost all variables explored are significantly associated with police severity mis-classification. Exceptions are: third party and type of day. In the multivariate modelling, the following characteristics are no longer significantly associated with police mis-classification: gender, road type, crash environment (out-of-town vs. in-town, or town size), light conditions, weather conditions, and junction.

Characteristics significantly associated with police severity mis-classification in the multivariate model are: injury severity itself (measured by NISS in 5 categories), casualty's age, road user type, police force reporting the crash and calendar year. Adjusted probabilities and adjusted relative risks are given in table 2. Probability of police severity mis-classification shows an inverse-U shape association with NISS, with a relative risk as high as 12.2 for NISS 9-15 casualties compared to NISS 1-3 casualties. Probability of police severity mis-classification displays a slight increase with casualty's age, with a relative risk of 1.3 for casualties aged 65 years and over compared to 0-13 year old casualties. Pedestrian and motorcyclist casualties have a slightly higher probability of being mis-classified, with a relative risk of 1.4 and 1.3 respectively, compared to injured car occupants. The 'urban police' are slightly more likely (RR=1.2) to mis-classify casualties' severity than the 'urban motorways police', whereas the 'rural police' are twice as likely to do so. Lastly, the probability of police severity mis-classification was slightly higher in 1997 (RR=1.3), compared to 2001.

Table 2: characteristics associated with police severity mis-classification, from a multivariate analysis among 14342 non-fatal casualties (Rhône county, 1997-2001)

	n	adjusted probability	adjusted RR	95% C.I.	
injury severity:					
NISS 1-3	8895	0.03	1.00		
NISS 4-8	2883	0.16	5.28	4.55	6.12
NISS 9-15	1481	0.36	12.23	10.57	14.15
NISS 16-24	621	0.25	8.64	7.19	10.39
NISS 25-75	462	0.15	4.97	3.90	6.34
age:					
0-13 years old	991	0.13	1.00		
14-17 years old	1038	0.12	0.94	0.75	1.18
18-24 years old	3493	0.14	1.11	0.91	1.35
25-64 years old	7809	0.16	1.21	1.01	1.45
65 years old and over	1011	0.17	1.33	1.09	1.63
road user type:					
pedestrians	2038	0.18	1.36	1.20	1.53
pedal cyclists	485	0.13	1.02	0.82	1.25
motorcyclists	2561	0.17	1.26	1.14	1.39
car occupants	8701	0.13	1.00		
other road users	557	0.11	0.85	0.64	1.13
reporting police force:					
'urban motorways police'	2479	0.10	1.00		
'rural police'	3022	0.24	2.34	1.98	2.76
'urban police'	8841	0.13	1.25	1.06	1.48
calendar year:					
1997	2612	0.17	1.28	1.12	1.45
1998	3152	0.14	1.09	0.97	1.24
1999	3016	0.14	1.05	0.92	1.20
2000	2763	0.14	1.06	0.93	1.21
2001	2799	0.13	1.00		

from a Poisson regression with robust error estimation

Over-classification characteristics

In the univariate analyses, all explored variables are significantly associated with the probability of police severity over-classification. In the multivariate modelling, the probability of police over-classification is significantly related to injury severity (as measured by NISS), casualty's age, road user type, reporting police force and calendar year. Adjusted over-classification probabilities and adjusted relative risks are displayed in table 3. The probability among NISS 1-15 casualties of being over-classified as 'seriously injured' by the police is related to injury severity itself: NISS 4-8 casualties are 5 times more likely to be over-classified than NISS 1-4 casualties, and NISS 9-15 casualties are 11 times more likely to be so.

The probability of police over-classification of casualties is related to their road user type: injured motorcyclists and injured pedestrians are respectively 1.2 and 1.4 times more likely to be over-classified as "seriously injured" than injured car occupants.

Probability of over-classification is related to the reporting police force: the 'rural police' are three times more likely to over-classify casualties than the 'urban motorways police'.

The probability of police over-classification decreases over the years 1997-2001. If we assume a linear decrease over the 1997-2001 period, we obtain a relative risk of 0.91 (C.I. 0.89-0.94) between any two consecutive years.

Table 3: characteristics associated with police severity over-classification, from a multivariate analysis among13259 non-fatal NISS 1-15 casualties (Rhône county, 1997-2001)

	n	adjusted probability	adjusted RR	95% C.I.	
injury severity:					
NISS 1-3	8895	0.03	1.00		
NISS 4-8	2883	0.15	4.89	4.22	5.68
NISS 9-15	1481	0.33	10.94	9.43	12.68
age:					
0-13	926	0.09	1.00		
14-17	948	0.10	1.06	0.82	1.35
18-24	3254	0.12	1.25	1.00	1.55
25-64	7226	0.12	1.33	1.08	1.63
65 years old and over	905	0.14	1.47	1.17	1.84
road user type:					
pedestrians	1816	0.15	1.42	1.25	1.63
pedal cyclists	436	0.11	1.01	0.81	1.26
motorcyclists	2219	0.13	1.24	1.11	1.38
car occupants	8260	0.11	1.00		
other road users	528	0.08	0.78	0.57	1.07
reporting police force:					
'urban motorways police'	2376	0.07	1.00		
'rural police'	2506	0.23	3.10	2.58	3.73
'urban police'	8377	0.09	1.16	0.96	1.40
calendar year:					
1997	2409	0.14	1.49	1.30	1.71
1998	2941	0.12	1.22	1.07	1.40
1999	2782	0.11	1.16	1.00	1.34
2000	2532	0.11	1.11	0.96	1.29
2001	2595	0.10	1.00		

from a Poisson regression with robust error estimation

Under-classification characteristics

In the univariate analyses, all explored variables except age, sex, and weather conditions are significantly associated with the probability of police severity under-classification. In the multivariate analysis, probability of police under-classification of severity is significantly associated with injury severity itself (as measured by NISS), road type, reporting police force and calendar year. Adjusted under-classification probabilities and adjusted relative risks are given in table 4.

The probability of police under-classification among NISS 16-75 casualties is related to their injury severity: NISS 16-24 casualties are 1.7 times more likely to be under-classified than 25-75 NISS casualties. The probability of police under-classification is related to the road type where the crash occurred: casualties on county roads, local roads, and to a lesser extent on state roads have a higher probability of being under-classified than casualties on motorways.
The probability of under-classification is related to the reporting police force: the 'urban motorways police' and the 'urban police' are three times as likely to under-classify casualties as the 'rural police'.
The probability of police under-classification increases over the years 1997-2001. If we assume a linear increase over the 1997-2001 period, we obtain a relative risk of 1.15 (C.I. 1.08-1.23) between any two consecutive years. These results are adjusted for the other factors, in particular for the NISS. The 25-75 NISS category being quite broad, we estimate the median NISS in each subgroup of police force, road type and year: it is very stable across subgroups, from 33.5 to 35.

Table 4: characteristics associated with police severity under-classification, from a multivariate analysis among 1083 non-fatal NISS 15-75 casualties (Rhône county, 1997-2001)

	n	adjusted probability	adjusted RR	95% C.I.	
injury severity:					
NISS 16-24	621	0.29	1.69	1.39	2.07
NISS 25-75	462	0.17	1.00		
road type:					
motorways	78	0.13	1.00		
state roads	135	0.24	1.84	1.02	3.32
county roads	422	0.28	2.16	1.12	4.14
local roads	406	0.28	2.18	1.11	4.27
other	42	0.22	1.69	0.72	3.93
reporting police force:					
'urban motorways police'	102	0.32	3.06	1.90	4.93
'rural police'	516	0.10	1.00		
'urban police'	465	0.32	3.11	2.31	4.18
calendar year:					
1997	203	0.18	1.00		
1998	211	0.18	1.03	0.72	1.46
1999	234	0.21	1.16	0.84	1.61
2000	231	0.26	1.46	1.06	1.99
2001	204	0.30	1.65	1.21	2.26

from a Poisson regression with robust error estimation

DISCUSSION

The first result of the study is the extent of the police severity mis-classification compared to the NISS. It is consistent with other countries' comparisons of police severity classification with severity measures based on the AIS, using either the ISS or the MAIS. A UK study[9] found that within the police category of 'serious' casualties, about one third are in fact ISS 1-4 and another third are ISS 4-9. A more recent UK[14] study also finds that a third of those classified as 'serious' by the police are in fact ISS 1-4. A small US study[12] finds that among 52 casualties with ISS $\geq$ 9, half of them were not classified as 'severe' by the police.
Even though there is some discordance between the police and the NISS classifications, there is still a strong association between the two, as shown in figure 1 and in the multivariate models. Casualties at the two ends of the NISS scale i.e. the most slightly and the most severely injured casualties are less prone to mis-classification. This is trivial and expected. The discordance and yet strong association between the police severity classification and the NISS are not surprising, given the criteria used by the police. The length of hospital stay or the number of days of ITT do not measure injury severity itself, but they are related to it and hence to the NISS, there is some concordance, though far from reaching 100%. When the police evaluate severity on the crash scene, it is likely that their classification is influenced by some broad injury characteristics such as the number of injuries, the region(s) injured and maybe the type(s) of injury so that their classification is related to some extent to injury severity and hence to the NISS.

Concerning the overall direction of the mis-classification, it is rather tempting to compare the numbers of over and under-classified casualties (respectively 1444 and 306 in table 1) and summarize the issue (and minimize it) by saying that globally the police tend to over-estimate severity. However, this is erroneous. One should take into account the police under-reporting of casualties. It is inversely related to injury severity, in France[30, 39] and in other countries[40], i.e. the lower the severity the lower the probability of being police-reported. This under-reporting pattern means that the number of NISS 1-15 casualties should be multiplied by a much larger correcting factor than the number of NISS 16-75 casualties, so that it is not possible to make an overall assessment of the police severity mis-classification.

The second and most important results of the present study are the general crash and casualty characteristics that the police severity under- and over-classifications are associated with. Pedestrians and motorcyclists have a slightly higher probability of being over-classified.: they are part of so-called vulnerable road users, that is to say those not protected by a vehicle body, and who, at equal impact force, are expected to be more seriously injured than 4-wheel vehicle occupants. It is hence not surprising that this expected higher severity plays a role in the police classification of severity. In an Australian study[15], road user type was not significantly associated with the agreement between police classification and hospital admission (yes/no) but the study was based on only 406 subjects. Road user type was not explored in the other two similar studies. A French study also based on the Rhône county but restricted to 1996[30] data already found that pedestrians had a higher probability of being classified as 'seriously injured' by the police (adjusted for MAIS).
Casualty's age is associated with a slight increase in over-classification probability. This is not surprising since age is a risk factor for both severity and longer hospital stay[17]. In the Australian study[15], the age of motor vehicle passengers is inversely associated with the probability of a correct police classification. Our result is consistent with this. On the contrary, in the US study[13], age above 65 years old shows a negative association with over-statement of injury severity.
There are large discrepancies in severity classification between the three police forces. A US study[3] analysing the validity of police crash data found some differences between the State, County and Municipal police agencies. They attributed part of these to their different areas of control. This has been explored in the present study. In the modelling of mis- and over-classification, the variables road type and crash environment (in town vs. out of town) are no longer associated with police severity classification as soon as the reporting police force is also included in the model. In the multivariate modelling of under-classification, the association with road type remains significant. Keeping this association in the model, the police forces adjusted relative risks are still large. This means that distortions in the classification associated with different areas of control (different road types or urban/rural distribution) explain at most some part but not all of the discrepancies in the police forces classification practices. Also, there is no remaining confounding effect of injury severity that could explain these discrepancies: we have checked that the median NISS of the wide NISS 25-75 subgroup is indeed very stable across police forces. In the UK, differences between police forces in severity classification have also been identified ; a study[41] has shown that the extent of variation among police force areas is higher than the extent of severity variation among different crash characteristics (urban/rural combined with day/night).
Over the time period considered 1997-2001, over-classification decreases but at the same time under-classification increases. This can be read as one general trend: casualties are more often classified by the police as 'slightly injured' (in a given NISS category). This might partly be due to the overall tendency of reduced length of hospital stays because of reduced public hospital fundings (and improved medical care and technology). This can only account for the (probably small) proportion of casualties whose severity is classified according to this criteria.

The study has its strengths and limitations. The NISS is not the gold standard that would perfectly measure the true injury severity. The NISS has been chosen as one severity scale based on a widely used and recognised injury assessment scale (the AIS) so that we could compare the low-quality police severity classification to it. Analyses have been replicated using the ISS or the MAIS so that we have checked that the results found were not very sensitive to the severity measure (among those based on

AIS). The accuracy of the AIS coding of injuries in the registry should be quite good: the registry's physician has of course been trained in AIS coding; she is also a member of the European Interest Group on Injury Scaling. NISS scores have to be grouped into 2 categories so that discordance between the police and the NISS classifications could be assessed, and it has been further labelled mis-classification and split into under and over-classification. These are simplifications but they are needed and yet informative.
The strength of this study lies mainly in the Rhône road trauma registry. This registry is not restricted to a few hospitals but it covers all the hospitals of a large county and its close surroundings. It includes both inpatients and outpatients. Furthermore the study is based on a five-year period, and hence on a large number of subjects.
The underlying assumption of the study is that each police force, having a centralised national structure (not regional) and receiving centralised instructions, have the same way of classifying crashes and casualties all over France. Under this assumption, the present study results can be extrapolated to the target population of casualties from road crashes occurring all over France.

The results found lead to some recommendations in the interpretation of studies and in the setting up of intervention campaigns. Severity classification being associated with some casualty and crash characteristics, the description of injury severity of different subgroups (road user types for instance) can be misleading. This also applies to any correlated characteristic such as gender -correlated with road user type-, or urban/rural area -correlated with police force-. More generally, the interpretation of any study of severity must be very cautious. In interpreting French police-based data, one should keep in mind that the severity of the following subgroups is over-estimated: pedestrians and motorcyclists, older casualties, people injured in areas controlled by the 'rural police'. Conversely, the severity of the following casualties is under-estimated: people injured in areas controlled by the 'urban police' and 'urban motorways police'. For instance, it is common knowledge that rural road crashes are more severe than urban ones. In France this knowledge is based on police data, so that the relative risk between the two types of crashes should be corrected with the new piece of information that the 'rural police' are 3 times more likely to over-classify casualties' severity than the 'urban police'. This should be further corrected with under-reporting coefficients as police under-reporting of casualties is related to police force[39]. Also, as there is a trend in the police probability of classifying more often casualties (in a given NISS category) as slightly injured, we must be very cautious if a decreased severity in police data is observed. Also, we must take these police severity over and under-classification into account when setting up intervention campaigns, in order to choose the appropriate target groups and conduct a valid evaluation of these campaigns.

All these words of caution about the interpretation of studies based on police data raise the issue of obtaining unbiased estimates of severity, on the traffic crash casualty population as a whole, and on different casualty subgroups. This is our next objective. We are working on obtaining nation-wide estimates of the number of slight and serious casualties (as measured by the NISS) as a whole, and according to age, gender, road user type, as well as to estimate the distribution of injury types. The first step consists of modelling the relationship between the registry and the police data at the Rhône level, identifying and quantifying selection biases (to be presented[39]) and classification biases (presented here). Based on these, the next step consists of constructing correction factors to be applied to the nation-wide police data to obtain a nation-wide picture of the road trauma registry.

Acknowledgements
We wish to thank the following people for having participated in the data collection and data entry, as a member of the association for the registry of road traffic casualties in the Rhône (ARVAC, president Banssillon V, director Ndiaye A) or as a member of INRETS-UMRESTTE (Laumon B, scientific consultant for the registry) : Ait Idir T, Ait Si Selmi T, Alloatti D, Andrillat M, Artru F, Asencio Y, Assossou I, Auzaneau F, Bagès-Limoges F, Bagou G, Balogh C, Banssillon G, Barnier N, Barth X, Basset M, Bec JF, Bejui J, Bel JC, Bérard E, Bérard J, Bernard JC, Berthet N, Bertrand JC, Besson L, Biot B, Biot V, Blanc C, Blanchard J, Bœuf C, Boisson D, Bonjean M, Bost J, Bouchedor C, Bouletreau P, Boyer V, Breda Y, Bret P, Brilland R, Bussery S, Cabet N, Caillot JL, Cannamela A, Caregnato B, Carre M, Catala Y, Chagnon PY, Chantran C, Chardon P, Charnay P, Chatelain P, Chattard S, Chavane H, Chazot G, Chettouane I, Chevreton N, Chevrillon E, Chevrillon S, Chiron M, Chotel P, Cochard P, Combe C, Contamin B, Coppard E, Cot T, Crettenet Z, Cristini A, Dal Gobbo B, De Angelis MP, Decourt L, Delfosse A, Demazière J, Deruty R, Desjardins G, Dohin B, Emonet A, Escarment J, Eyssette M,

Fallavier L, Fanton L, Felten D, Feuglet P, Fifis N, Figura J, Fisher G, Fischer LP, Floccard B, Floret D, Fournier G, Fredenucci JF, Freidel M, Gadegbeku B, Galin L, Gaillard P, Gallon M, Garnier N, Garzanti A, Gaussorgues P, Gautheron V, Genevrier M, Gibaud F, Gillet Y, Goubsky A, Granger M, Grattard P, Gueniaud PY, Guenot C, Guignand M, Haddak M, Hamel D, Heckel T, Jacquemard C, Joffre T, Kohler R, Lablanche C, Lafont S, Lagier C, Lapierre B, Laplace MC, La Rosa C, Laurent R, Lebel M, Leblay G, Le-Xuan I, Lieutaud T, Lille R, Linné M, Lucas R, Machin B, Maello E, Malicier D, Mangola B, Marduel YN, Marie-Catherine M, Martin JL, Martinand G, Marty F, Messikh C, Meyer F, Meyrand S, Molard S, Morel-Chevillet E, Mioulet E, Minjaud F, Mollet C, Monnet J, Moreno S, Moyen B, Ndiaye A, Neidhart JP, Ngandu E, Ny S, Ould T, Paget P, Paillot JC, Paris D, Patay B, Pauget P, Peillon D, Perrin-Blondeau D, Petit P, Piton JL, Plantier M, Pornon P, Pramayon C, Quelard B, Rezig M, Rigal F, Robert D, Rode G, Romanet JP, Rongieras F, Roset C, Rousson A, Roussouli P, Roux H, Ruhl C, Salamand J, Sametzky P, Sayegh K, Sbraire N, Scappaticci N, Schiele P, Schneider M, Simonet C, Sindou M, Soldner R, Soudain M, Stagnara J, Stamm D, Suc B, Taesch MC, Tasseau F, Tell L, Thomas M, Tilhet-Coartet S, Toukou JC, Trifot M, Vallet G, Vancuyck A, Vergnes I, Verney MP, Voiglio EJ, Vourey G, Vuillard J, Westphal M, Willemen L.

References

1. OCDE-IRTAD. Definitions and data availability: compilation and evaluation of A-level Roads and hospitalised victims in OECD countries- accident and injury definitions. Germany: BASt; 1998. IRTAD special report.
2. Bull JP, Roberts BJ. Road accident statistics--A comparison of police and hospital information. *Accid Anal Prev.* 1973;5:45-53.
3. Shinar D, Treat JR, McDonald ST. The validity of police reported accident data. *Accid Anal Prev.* 1983;15:175-191.
4. Barancik J, Fife D. Discrepancies in vehicular crash injury reporting: northeastern Ohio trauma study IV. *Accid Anal Prev.* 1985;17:147-154.
5. Hansen A, Popkin C, Campbell B, et al. Comparison of police and physician judgments of injury severity and inebriation in motor vehicle crash victims. In: Proceedings of the 35th annual conference of the Association for the Advancement of Automative Medecine, 1991:249-260.
6. Rosman DL, Knuiman MW. A comparison of hospital and police road injury data. *Accid Anal Prev.* 1994;26:215-222.
7. Austin K. The identification of mistakes in road accident records: part 2, casualty variables. *Accid Anal Prev.* 1995;27:277-282.
8. Aptel I, Salmi LR, Masson F, et al. Road accident statistics: discrepancies between police and hospital data in a French island. *Accid Anal Prev.* 1999;31:101-108.
9. Dove AF, Pearson JC, Weston PA. Data collection from road traffic accidents. *Arch Emerg Med.* 1986;3:193-198.
10. Tunbridge R, Everest J. An assessment of the under reporting of road accident casualties in relation to injury severity. In: Proceedings of the International IRCOBI conference on the biomechanics of impact, 1988:101-109.
11. Nordentoft E, Larsen C, Jørgensen H. Færdselsuheldsstatistikkens dækningsgrad [the degree of coverage of traffic accident statistics]. *Ugeskr Laeger.* 1989;151:2808-2811.
12. Agran PF, Castillo DN, Winn DG. Limitations of data compiled from police reports on pediatric pedestrian and bicycle motor vehicle events. *Accid Anal Prev.* 1990;22:361-70.
13. Farmer CM. Reliability of police-reported information for determining crash and injury severity. *Traffic Inj Prev.* 2003;4:38-44.
14. Morris A, Mackay M, Wodzin E, et al. Some injury scaling issues in UK crash research. In: Proceedings of the international IRCOBI conference on the biomechanics of impact, 2003:283-291.
15. Lopez DG, Rosman DL, Jelinek GA, et al. Complementing police road-crash records with trauma registry data- an initial evaluation. *Accid Anal Prev.* 2000;32:771-777.
16. Oviève F. Etude du critère "blessé grave" dans les fichiers BAAC des statistiques officielles des victimes d'accident corporel de la circulation routière dans le département du Rhône ['seriously injured' criteria in the police data of road casualties in the Rhône county]. Lyon: Université Claude Bernard; 2003. Rapport UMRETTE No. 0304.
17. Bergeron E, Lavoie A, Moore L, et al. Comorbidity and age are both independent predictors of length of hospitalization in trauma patients. *Can J Surg.* 2005;48:361-366.
18. Charnay P, Laumon B. Le registre des victimes d'accidents de la circulation routière du Rhône [The trauma registry of road casualties in the Rhône county]. Bron: INRETS; 2002. Rapport UMRETTE No. 0204.
19. Laumon B, Martin JL, Collet P, et al. A French road accident trauma registry: first results. In: Proceedings of the 41st annual conference of the Association for the Advancement of Automotive Medecine, 1997:127-137.
20. Baker S, O'Neill B, Haddon W, et al. The Injury Severity Score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma.* 1974;14:187-196.

21. Osler T, Baker S, Long W. A modification of the Injury Severity Score that both improves accuracy and simplifies scoring. *J Trauma.* 1997;43:922-926.
22. Brenneman F, Boulanger B, McLellan B, et al. Measuring Injury Severity: Time for a Change? *J Trauma.* 1998;44:580-582.
23. Frankema S, Steyerberg E, Edwards M, et al. Comparison of Current Injury Scales for Survival Chance Estimation: An Evaluation Comparing the Predictive Performance of the ISS, NISS, and AP Scores in a Dutch Local Trauma Registration. *J Trauma.* 2005;58:596-604.
24. Lavoie A, Moore L, LeSage N, et al. The New Injury Severity Score: a more accurate predictor of in-hospital mortality than the Injury Severity Score. *J Trauma.* 2004;56:1312-1320.
25. Kilgo P, Osler T, Meredith W. The Worst Injury Predicts Mortality Outcome the Best: Rethinking the Role of Multiple Injuries in Trauma Outcome Scoring. *J Trauma.* 2003;55:599-607.
26. Tay S, Sloan E, Zun L, et al. Comparison of the New Injury Severity Score and the Injury Severity Score. *J Trauma.* 2004;56:162-164.
27. Sacco W, MacKenzie E, Champion H, et al. Comparison of alternative methods for assessing Injury Severity based on anatomic descriptors. *J Trauma.* 1999;47:441-446.
28. Samin O, Civil I. The New Injury Severity Score versus the Injury Severity Score in predicting patient outcome: a comparative evaluation on trauma service patients of the Auckland hospital. In: Proceedings of the 43rd annual conference of the Association for the Advancement of Automative Medicine, 1999:1-15.
29. Stevenson M, Segui-Gomez M, Lescohier I, et al. An overview of the injury severity score and the new injury severity score. *Inj Prev.* 2001;7:10-13.
30. Laumon B, Martin JL. Analyse des biais dans la connaissance épidémiologique des accidents de la route en France [Bias in the epidemiological knowledge of road crashes in France]. *Rev Epidemiol Sante Publique.* 2002;50:277-285.
31. Clark DE. Practical introduction to record linkage for injury research. *Inj Prev.* 2004;10:186-191.
32. Rosman DL, Knuiman MW, Ryan GA. An evaluation of road crash injury severity measures. *Accid Anal Prev.* 1996;28:163-170.
33. Chiron M, Gadegbeku B, Ndiaye A. Les conséquences sanitaires des accidents de la route [Health outcomes of road crashes]. In: ONISR, ed. *La sécurité routière en France, bilan de l'année 2003*. Paris: La documentation Française; 2004:195-200.
34. Haddak M, Filou C, Ndiaye A, et al. Redéfinition des victimes de la route [Towards a new definition of road casualties]. Premier rapport d'étape. Bron: INRETS; 2004. Rapport UMRETTE No. 0405.
35. Spiegelman D, Hertzmark E. Easy SAS Calculations for Risk or Prevalence Ratios and Differences. *Am J Epidemiol.* 2005;162:199-200.
36. Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am J Epidemiol.* 2004;160:301-305.
37. Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol.* 2004;159:702-706.
38. Lindquist K. How to estimate relative risk in SAS using Proc Genmod for common outcome in cohort studies. May 4, 2004. Available at: http://www.ats.ucla.edu/stat/sas/faq/relative_risk.htm. Accessed December 9, 2004
39. Amoros E, Martin JL, Laumon B. Under-reporting of road crash casualties in France. *Accid Anal Prev.* 2006; 38:627-635
40. Hvoslef H. Under-reporting of road traffic accidents recorded by the police at the international level. Norway: Public Roads Administration; 1994. IRTAD special report.
41. Lai PW. Model of injury severity allowing for different gradings of severity: Some applications using the British road accident data. *Accid Anal Prev.* 1980;12:221-239.

Estimation du nombre de faux négatifs et de faux positifs, sur le résultat du chaînage des données de 2001

Afin d'améliorer le résultat du chaînage entre le registre et les BAAC, nous estimons le pourcentage et le nombre de faux négatifs et faux positifs à l'aide de calculs de probabilité.

Faux négatifs
Un faux négatif est une paire d'enregistrements non reliés, à tort, autrement dit une paire d'enregistrements du même individu que l'on n'a pas rapprochés.

P(faux négatif)
=P(ne pas rapprocher alors qu'il s'agit d'enregistrements d'une même victime)
=P(ne pas rapprocher / même victime)
=1-P(rapprocher/même victime)
=1-P(concordance sur toutes ou certaines variables / même victime)

Concordance « sur toutes ou certaines variables de chaînage » est précisément défini. Les variables de chaînage sont :

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

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, définis par les années. Cela suppose l'absence d'erreur sur cette variable. Pour les accidents ayant lieu en fin d'année (d'après les BAAC), les enregistrements du registre des jours 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 décidons de relier la paire si :
- A) les sept variables de chaînage concordent, ou
- B) seulement six variables concordent, ou
- C) 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.

Ces trois situations (A, B, C) étant disjointes, la probabilité de l'ensemble est la somme des trois probabilités et chacune, sous l'hypothèse d'indépendance entre les variables de chaînage, est le produit des probabilités de concordance pour les variables de chaînage prises individuellement.

En langage mathématique :
P (relier la paire/même victime)
=P (7 variables concordent / même victime) + P (exactement 6 variables concordent / même victime) + P (exactement 5 variables concordent / même victime)

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

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

$$\left[\sum_{(j,k)=(4,6),(4,7),(6,7)} \left(\prod_{i=1, i\neq j, i\neq k}^{i=7} P(\text{var}_i \ concorde / même \ victime) \times P(\text{var}_j \ discorde / même \ victime) \times P(\text{var}_k \ dicorde / même \ victime \right) \right]$$

D'après Brenner et Schmidtmann, si on dispose du « gold standard », i.e. si on connait le sous-ensemble M des paires correspondant à une même victime, et le sous-ensemble U des paires correspondant à deux victimes distinctes, 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 argument, qui conduit aussi à modérer ces proportions : chacune de ces probabilités est égale à 1-taux d'erreur (Howe 1998), et cela ne parait pas raisonnable 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 7 : 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 2ème prénom	88,8%
jour de naissance	85,1%
mois de naissance	93,3%
année de naissance	77,3%
région de naissance	98,1%

Nous choisissons donc de diminuer les proportions observées de 5% (valeur arbitraire). Les proportions de concordance sont données par le tableau ci-dessous :

Tableau 8 : proportions de concordance sur les variables de chaînage, observées sur paires reliées, Registre et BAAC, Rhône, 2001, et quantification des faux négatifs

variables de chaînage	concordance parmi les paires reliées (n=2813)	minoration de 5%
jour accident	97,3%	92%
mois accident	99,9%	95%
année accident	*100%*	*100%*
lieu de l'accident (commune)	82,9%	78%
mois de naissance	95,8%	91%
année de naissance	96,5%	92%
sexe	97,1%	92%
type d'usager	96,1%	91%
probabilité de concordance / ensemble M	96,4%	87,0%
probabilité de faux négatifs	3,6%	13,0%
effectif de faux négatifs (% appliqué à l'ensemble des paires reliées)	109	396
effectif de faux négatifs (% appliqué à l'ensemble des paires possibles)	148	537

Pour le lieu d'accident, nous estimons les proportions de concordance sur la commune uniquement (i.e. sans inclure les informations sur la route ou la rue… trop complexes à prendre en compte). Le faible taux de concordance observé sur la variable commune (par rapport aux autres) s'explique par le fait qu'il est obtenu par une seule comparaison (automatisée) des codes INSEE des communes, alors que dans la pratique, deux communes limitrophes sont plutôt considérées comme concordantes si les éventuels noms de rue indiqués ne l'infirment pas.

La probabilité de faux négatifs est intéressante en soi, mais nous souhaitons aussi obtenir l'effectif des faux négatifs, La probabilité de faux négatifs est définie sur l'ensemble M des paires qui correspondent au même individu. Cet ensemble n'étant pas connu, nous pouvons de nouveau l'approximer par l'ensemble de paires reliées, ou bien, si l'on veut un majorant, par le plus grand ensemble possible de liens, qui correspond à la

situation où toutes les victimes du plus petit ensemble (les BAAC ici) sont reliées à une victime de l'autre ensemble (le Registre).
Nous avons choisi de l'approximer par l'ensemble des paires reliées (qui nous parait le plus proche de l'ensemble M). Nous retenons ainsi l'estimation de 396 faux négatifs.

Faux positifs :
Un faux positif est une paire d'enregistrements reliés, à tort, autrement dit une paire d'enregistrements de deux victimes distinctes que l'on a rapprochés

P (faux positifs)
=P (rapprocher alors qu'il s'agit d'enregistrements de 2 victimes distinctes)
=P (rapprocher / victimes distinctes)
=P (concordance sur toutes ou certaines variables / victimes distinctes)

La concordance « sur toutes ou certaines variables » est celle définie précédemment, de façon précise en fonction des sept variables de chaînage.

La concordance sur une variable alors qu'il s'agit de victimes distinctes correspond à une concordance par hasard. 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 pour le mois et environ 1/30 pour le jour (Brenner and Schmidtmann 1996).

Pour les autres variables, les proportions de concordance sont estimées par les proportions observées sur l'ensemble (U) des paires d'enregistrements de victimes distinctes si l'on dispose du gold standard, A défaut, il est approximé par l'ensemble des paires non-reliées, voire par l'ensemble de toutes les paires possibles, car la différence entre ces deux sous-ensembles est négligeable : leurs tailles respectives sont $N_A*N_B - N_A$ et N_A*N_B, avec N_A négligeable devant N_A*N_B. Nous choisissons d'approximer par l'ensemble des paires non-reliées, soit 47 220 438.

Concernant le lieu de l'accident, nous minorons la proportion obtenue 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 à 1% est arbitrairement choisi.

Tableau 9 : proportions de concordance sur les variables de chaînage, observées sur l'ensemble des 47 220 438 paires non-reliées, registre et BAAC, Rhône, 2001, et quantification des faux positifs

Variables de chaînage	Concordance parmi les paires non-reliées	Après minoration (lieu d'accident)
Jour accident	3,2%	3,2%
Mois accident	8,4%	8,4%
Année accident	100,0%	100,0%
Lieu de l'accident (commune)	8,0%	1,0%
Mois de naissance	8,3%	8,3%
Année de naissance	2,0%	2,0%
sexe	52,6%	52,6%
type d'usager	38,2%	38,2%
Probabilité de concordance / ensemble U	1,1E-5*	2,1E-6*
Probabilité de faux positifs	1,1E-5*	2,1E-6*
Effectif de faux positifs	486	97

* : la probabilité de faux positifs est très faible ; elle est définie sur l'ensemble des paires d'enregistrements d'individus distincts, dont le nombre s'élève à 47 million.

Nous vérifions que la proportion observée de concordance sur le jour de l'accident (3,2%) est proche de 1/30 (soit 3,33%), celles observées sur le mois d'accident (8,4%), le mois de naissance (8,3%) proches de 1/12 soit 8,33%.

Il est donc estimé un effectif de 97 faux positifs.

Remarque :
Les critères de rapprochement sont une écriture un peu simplifiée de la procédure de rapprochement telle qu'elle est pratiquée.
Cela ne prend pas en compte le fait qu'on rapproche aussi au niveau de l'accident, et que, lorsqu'il y a plusieurs victimes dans un accident, cela peut aider au rapprochement d'une des victimes : si trois victimes sur quatre dans un accident sont aisément reliables (toutes les variables concordent), on rapprochera plus souvent la 4ème victime (même si discordance sur jour de l'accident et mois de naissance par exemple) que si c'était une unique victime dans un accident.
Autre situation non prise en compte : si six variables concordent mais que la discordance porte sur le lieu d'accident, on ne rapproche pas systématiquement ; on rapproche si les deux communes sont limitrophes (et que le nom de rue ou de route n'infirme pas), ou si la commune est non renseignée (côté registre), mais en pratique, il est apparu que la décision prise pour cette dernière situation dépend de la personne effectuant le rapprochement.

Références :
Brenner, H. and I. Schmidtmann (1996). "Determinants of homonym and synonym rates of record linkage in disease registration." Methods of Information in Medicine **35**(1): 19-24.
Howe, G. R. (1998). "Use of computerized record linkage in cohort studies." Epidemiologic Reviews **20**(1): 112-21.
Newcombe, H. B. (1988). Handbook of record linkage: methods for health and statistical studies, administration, and business. Oxford, Oxford university press.

Résultats détaillés de l'analyse capture-recapture stratifiée, scénarios 1, 2 et 3

Tableau 10 : résultats détaillés de capture-recapture, scénario 1, Rhône 2001

Type d'usager et tiers	Gravité lésionnelle	Effectif BAAC	Effectif Registre	Effectif intersection	Effectif union des deux	Effectif capture-recapture	Taux de couverture BAAC	Taux de couverture du registre	Taux de couverture du regroupement (union des 2)
piétons	NISS 1-3	286	589	187	687	897	31,9%	65,7%	76,6%
	NISS 4-8	138	192	101	230	264	52,3%	72,7%	87,2%
	NISS 9+	125	138	99	164	174	71,8%	79,3%	94,2%
cyclistes	NISS 1-3	62	836	47	850	1090	5,7%	76,7%	78,0%
	NISS 4-8	21	322	15	328	448	4,7%	71,9%	73,2%
	NISS 9+	18	75	14	79	96	18,8%	78,1%	82,4%
Usagers de 2RM	NISS 1-3	299	682	225	756	905	33,0%	75,4%	83,5%
Avec tiers	NISS 4-8	191	226	117	300	368	51,9%	61,4%	81,5%
	NISS 9+	149	177	125	201	211	70,6%	83,9%	95,3%
Sans tiers	NISS 1-3	27	699	14	712	1310	2,1%	53,4%	54,4%
	NISS 4-8	34	266	16	285	555	6,1%	47,9%	51,4%
	NISS 9+	25	110	22	113	123	20,3%	89,4%	91,8%
automobilistes	NISS 1-3	1576	3644	1019	4201	5636	28,0%	64,7%	74,5%
Avec tiers	NISS 4-8	325	440	228	538	629	51,7%	70,0%	85,6%
	NISS 9+	158	179	116	222	245	64,5%	73,1%	90,6%
Sans tiers	NISS 1-3	322	1108	211	1219	1691	19,0%	65,5%	72,1%
	NISS 4-8	130	244	89	285	357	36,4%	68,3%	80,0%
	NISS 9+	82	149	71	161	174	47,1%	85,6%	92,6%
autres	NISS 1-3	134	414	71	477	776	17,3%	53,4%	61,5%
	NISS 4-8	21	98	15	105	137	15,3%	71,5%	76,2%
	NISS 9+	13	44	13	44	44	29,5%	100,0%	100,0%

Tableau 11 : résultats détaillés de capture-recapture, scénario 2, Rhône 2001

Type d'usager et tiers	Gravité lésionnelle	Effectif BAAC	Effectif Registre	Effectif intersection	Effectif union des deux	Effectif capture-recapture	Taux de couverture BAAC	Taux de couverture du registre	Taux de couverture du regroupement (union des 2)
piétons	NISS 1-3	285	591	197	679	854	33,4%	69,2%	79,5%
	NISS 4-8	139	196	106	229	257	54,2%	76,1%	89,1%
	NISS 9+	127	142	104	165	174	73,0%	81,9%	95,1%
cyclistes	NISS 1-3	62	824	49	837	1037	6,0%	79,5%	80,7%
	NISS 4-8	21	317	16	323	423	5,0%	75,0%	76,3%
	NISS 9+	18	75	15	79	93	19,9%	80,7%	84,7%
usagers de 2RM	NISS 1-3	302	685	236	751	875	34,5%	78,3%	85,8%
avec tiers	NISS 4-8	189	230	123	296	353	53,4%	65,3%	83,9%
	NISS 9+	153	183	131	204	212	71,9%	86,1%	96,1%
sans tiers	NISS 1-3	26	688	15	699	1199	2,2%	57,3%	58,3%
	NISS 4-8	33	263	17	279	505	6,6%	52,1%	55,3%
	NISS 9+	25	110	23	112	120	21,1%	91,1%	93,0%
automobilistes	NISS 1-3	1567	3648	1070	4145	5341	29,3%	68,3%	77,6%
avec tiers	NISS 4-8	326	448	239	535	611	53,4%	73,4%	87,6%
	NISS 9+	160	184	122	222	241	66,2%	76,3%	92,0%
sans tiers	NISS 1-3	321	1102	221	1202	1595	20,1%	69,1%	75,3%
	NISS 4-8	130	246	93	283	342	37,9%	72,0%	82,6%
	NISS 9+	85	152	74	162	173	48,9%	87,7%	93,8%
autres	NISS 1-3	131	411	75	467	716	18,2%	57,4%	65,2%
	NISS 4-8	21	98	16	103	131	16,4%	74,8%	79,1%
	NISS 9+	14	44	14	44	44	31,3%	100,0%	100,0%

tableau 12 : résultats détaillés de capture-recapture, scénario 3, Rhône, 2001

Type d'usager et tiers	Gravité lésionnelle	Effectif BAAC	Effectif Registre	Effectif intersection	Effectif union des deux	Effectif capture-recapture	Taux de couverture BAAC	Taux de couverture du registre	Taux de couverture de l' union des 2
piétons	NISS 1-3	284	593	207	670	811	35.0%	73.2%	82.6%
	NISS 4-8	141	200	112	229	251	55.9%	79.4%	91.0%
	NISS 9+	129	147	109	167	174	74.4%	84.6%	96.1%
cyclistes	NISS 1-3	63	811	52	822	983	6.4%	82.4%	83.6%
	NISS 4-8	21	312	17	317	398	5.4%	78.4%	79.6%
	NISS 9+	19	74	15	78	89	21.0%	83.5%	87.1%
usagers de 2RM avec tiers	NISS 1-3	306	689	249	745	846	36.2%	81.4%	88.2%
	NISS 4-8	186	235	129	292	338	55.2%	69.5%	86.4%
	NISS 9+	157	189	138	207	214	73.4%	88.3%	96.9%
sans tiers	NISS 1-3	26	675	15	685	1090	2.4%	61.9%	62.9%
	NISS 4-8	32	259	18	273	455	7.0%	56.8%	59.9%
	NISS 9+	26	109	24	111	118	22.4%	92.5%	94.3%
automobilistes avec tiers	NISS 1-3	1558	3652	1127	4084	5049	30.9%	72.3%	80.9%
	NISS 4-8	327	457	252	532	593	55.2%	77.0%	89.7%
	NISS 9+	161	189	128	222	238	67.8%	79.6%	93.5%
sans tiers	NISS 1-3	319	1096	233	1182	1500	21.3%	73.1%	78.8%
	NISS 4-8	130	248	98	280	328	39.6%	75.7%	85.4%
	NISS 9+	87	154	78	163	172	50.7%	89.7%	94.9%
autres	NISS 1-3	127	408	79	457	658	19.3%	62.0%	69.4%
	NISS 4-8	21	97	17	102	124	17.3%	78.2%	82.1%
	NISS 9+	14	44	14	44	44	32.9%	100.0%	100.0%

Article d'application de capture-recapture publié dans Accident Analysis and Prevention, version "pré-print"

Amoros E, Martin JL, Laumon B, Estimating non-fatal road casualties in a large French county, using the capture-recapture method. *Accident Analysis and Prevention,* 2007 May;39(3):483-490

ABSTRACT

Introduction: non-fatal road casualties are under-reported, with official figures coming from the police. In the French Rhône county, a road trauma registry constitutes a second source of reporting but its completeness needs to be assessed. We also wish to estimate an incidence rate of non-fatal road casualties that is corrected for under-count.

Methods: having two sources of reporting available, we can apply the capture-recapture method under certain conditions. To take into account different reporting probabilities among casualties, the capture-recapture analysis is stratified according to injury severity (New Injury Severity Score=NISS), road user type and human third party. To evaluate the sensitivity of the capture-recapture estimate on the number of matched casualties between the police file and the registry, three scenarios of record-linkage are considered.

Results: For serious (NISS 9+) non-fatal road casualties, the police ascertainment rate is at most 57%, the registry ascertainment rate is at most 87%, and the aggregate ascertainment rate is at most 95%. The ascertainment-corrected incidence rate for serious (NISS 9+) non-fatal road casualties is at least 65 per 100 000.

Conclusion: the degree of completeness of the registry for serious casualties is rather high, though not satisfactory.

Keywords: capture-recapture, ascertainment rate, record-linkage, road trauma registry, injury severity, under-reporting

1. INTRODUCTION

Non-fatal road casualties are known to be under-reported, with official figures usually coming from the police forces. Hospital discharges are sometimes used as a second source of reporting. When two or more sources exist, it is possible, under certain conditions, to use the capture-recapture method to estimate the number of cases missed by all sources and hence obtain an incidence rate that is corrected for under-count. The capture-recapture method has been used in epidemiology (IWGDMF 1995a;1995b; Gallay et al. 2002; Hook and Regal 1995) including road injury (Aptel et al. 1999; Dhillon et al. 2001; Meuleners et al. 2006; Razzak and Luby 1998; Roberts and Scragg 1994; Tercero and Andersson 2004) with two papers showing more sceptical views (Jarvis et al. 2000; Morrison and Stone 2000).

In France, in the Rhône county, which is a large area of 1.6 million inhabitants, a road trauma registry has been in operation since 1996, providing an additional source of reporting to the police's. We apply the capture-recapture method to evaluate both the completeness of the registry and to obtain an estimate of the incidence rate of non-fatal road casualties correcting as much as possible for under-count. A capture-recapture analysis stratified on injury severity measured by the New Injury Severity Score, road user type and third party involvement (yes/no) is conducted. One of the conditions of the capture-recapture method is the perfect identification of subjects common to both sources. This is always difficult to reach and to assess. In this study, common casualties are identified by a semi-automated probabilistic record-linkage. Its validity is assessed by two methods: an additional record-linkage using surnames and first names, and a method inspired by Brenner (1994) and Brenner and Schmidtmann (1996) based on the probability of agreement of the linking variables. Based on this, we elaborate three possible scenarios depending on the number of linked casualties, and we apply the capture-recapture method on each of these.

2. MATERIAL AND METHODS

The study is based on the 2001 annual data of non-fatal road casualties, injured in crashes that occurred in the Rhône county (irrespective of the place of residence of the casualties), from the road trauma registry and from the police file. The 2001 annual data are studied as 2001 is the last year before the registry recording procedure has been modified. The Rhône county ("département") is an area of 1.6 million inhabitants, consisting of a large city (Lyon), its suburbs and a rural area in the northern part.

2.1 Police records

The French police are required by law to write a crash report for every road crash causing at least one casualty. A road crash is officially defined as a crash involving at least one vehicle and occurring on the network open to public traffic. Skateboard, roller skate or scooter users are considered as pedestrians by the police, and, as such, are only classified as road casualties if hit by a vehicle. There is no restriction about motorised vehicles, in other words there is no exclusion criteria for bicycles. The police crash report should include everyone involved in the crash, and classify them as non-injured, slightly injured, severely injured (depending on whether the length of hospital stay is less or more than 6 days) or deceased. Fatalities and involved non-injured are excluded from this study.

Police crash reports are extensive paper reports ("procès-verbaux"); these are later recorded as computer files in a shorter version; in particular, they are made anonymous. It is these computer files ("BAAC") that are commonly available and referred to as "police file".

2.2 The Rhône road trauma registry

A road trauma registry has been in operation since 1995 (Laumon et al. 1997) in the Rhône county. The registry covers all casualties from road crashes in the Rhône county who seek medical attention in health facilities. Inclusion criteria are broader than those of the police in terms of crash characteristics: off-road crashes are not excluded; roller skate, skate-board or scooter users are not considered as pedestrians but as road users and are hence included, whether hit by another vehicle or not. In this study, common criteria on road user type are applied so that they are excluded if injured on their own.

The registry is based on the participation of all health care facilities in the county (and its close surroundings) that could receive victims of a traffic crash: it includes some 270 health care facilities: from pre-hospital emergency care, emergency departments, intensive care units, surgery units... to

rehabilitation departments. The registry includes both inpatients and outpatients, i.e. all casualties, whether hospitalised or not. Information collected for each casualty consists of a few crash characteristics and of the following casualty characteristics: gender, date of birth, place of residence, hospital stay, hospital transfer (if relevant), and accurate injury assessment. Indeed, for each casualty, injury assessment is based on the whole set of diagnoses provided by the different health services the subject may have visited. Plain text diagnoses are coded by the registry physicians according to the Abbreviated Injury Scale (AIS), 1990 revision. Each injury is assigned a severity score, ranging from AIS 1 (minor) to AIS 6 (beyond treatment). To measure the severity of multiple injury casualties, we use the NISS i.e. the New Injury Severity Score (Osler et al., 1997), which is defined as the sum of squares of the AIS severity scores of a patient's three most severe injuries, regardless of body region in which they occur.

2.3 Record-linkage and additional surveys

A semi-automated probabilistic record-linkage (Clark 2004) procedure has been applied on the police and registry data. Linking variables are date and time of the crash, crash location, type of road user, date of birth (year and month only) and gender of the casualty. The procedure was not fully automated since a major linking variable (crash location) was left as unformatted free text to avoid losing any valuable information; crash location typically consists of town/village (or nearest one) and street(s)/road(s) name. The decision to link a pair is taken when at least five major linking variables agree (the minor ones being month of birth, road user type and gender). This record-linkage procedure has been a standard procedure conducted on the police and registry data between 1996 and 2001.

Following this "standard" record linkage on the 2001 data, two additional surveys have been conducted on casualties identified in one source only, i.e. one survey on the casualties identified in the registry but not identified in the police file ("registry only") and the other survey on the casualties identified in the police file but not identified in the registry ("police only"). One of the purposes was to check whether this was really the case or a failure of the record-linkage. Some results of the survey on casualties in the "registry only" were previously reported (Amoros et al. 2006), indicating that half of them alerted the police, but also that there was a response bias towards more law abiding citizens. 25% of the respondents believed the police established a crash report, which, if this were the case, when applied to the whole set of casualties in the "registry only" would outnumber the casualties in the "police only"; this means that it is impossible to know how many of these "registry only" casualties really had a police crash report established.

The survey on casualties in the "police only" consisted in going back to the paper crash reports in order to collect names and medical data if available. It appeared that medical certificates were often included, allowing the coding of injury severity into the New Injury Severity Score (by the registry physicians). The names allowed us to conduct an additional record-linkage.

The additional record-linkage made use of first name and surname (married name, if applicable, for women) as these were available from both the police paper crash reports and from the registry (from 2000 onwards). Names were compared using the Soundex code to allow for typing, spelling, or transliteration mistakes. This additional record-linkage was conducted on records that remain un-linked from the previous "standard" record-linkage (i.e. from "police only" and from "registry only"). Pairs were pre-selected on matching first name and surname (Soundex coded) and then linked if also matching on year and month of birth, date of crash and place of crash.

Finally, inspired by two papers from by Brenner (1994) and Brenner and Schmidtmann (1996), we have assessed the quality of our record-linkage procedure by estimating the number of false positives and false negatives i.e. the number of pairs linked whereas corresponding to two distinct casualties and the number of non-linked pairs whereas corresponding to the same casualty (same crash and same individual). These estimations are based on probability computations and approximations of these. These are probability of disagreement of the linking variables among pairs of records from the same casualties, and probability of agreement (by chance) among pairs of records from distinct casualties. The details are given in the appendix.

2.4 Capture-recapture

As we have only two sources, we apply the simple 2-list capture-recapture method. As we conduct a stratified analysis and some of the strata are quite sparse, we use the nearly unbiased estimator (Wittes

1972), $\hat{n} = \frac{(n_A + 1) \times (n_B + 1)}{(n_{AB} + 1)} - 1$, with n_A the number of non-fatal casualties in the police computer file, n_B the number in the registry, and n_{AB} the number of matched casualties.

The capture-recapture method is based on four assumptions (Gallay et al. 2002; Hook and Regal 1995). They will be all developed in the discussion. One of the assumptions is named "homogeneity of capture" and means a homogeneous probability among all casualties of being recorded, by a given source. Different probabilities among casualties, typically probabilities depending on the severity of injury, can be taken into account by stratification. A previous study of police under-reporting indicates a number of crash and casualty characteristics associated with different probabilities of reporting (Amoros et al. 2006). The major ones are: injury severity, road user type, whether a human third party was involved in the crash (yes/no), and the combination between road type, urban / rural area and type of police force. Minor police under-reporting factors are: age and gender, light conditions (day vs. night). Registry under-reporting factors are assumed to be injury severity and road user type, with cyclists being less viewed as road users. We choose to conduct a capture-recapture analysis stratified on injury severity, road user type, and human third party (yes/no). We are limited by the number of variables we can simultaneously stratify on (three variables is usually a maximum) and by the data available in both sources. Injury severity is measured by the NISS. Road user type is defined by 5 categories: pedestrians, cyclists, motorcyclists, car occupants, others (van, truck, bus…). Human third party (yes/no) strata are only constructed for car occupants and motorcyclists, as there are too few observations in the cyclists and "others" strata in the police computer file to split them further. Injury severity as measured by the NISS is available among the registry casualties, and among "the police only" casualties thanks to the study going back to the police paper crash reports: it is only available for those with a medical certificate, which are however most of them and we assume we can apply the distribution of NISS evaluated on those casualties to those without a medical certificate (and also to those without an available paper crash report).

The fourth assumption of the capture-recapture method is the perfect identification of the common casualties between the two lists, in other words a perfect matching. As already mentioned, this is always difficult to achieve and to assess. We conduct a sensitivity analysis: we apply the capture-recapture method on 3 scenarios. The first scenario corresponds to our "standard" record-linkage procedure. The second scenario corresponds to the standard situation updated with the additional record-linkage using first name and surname. This second scenario is an estimation: only a subset of "police only" records had the original paper crash report available so that the additional linked records are an extrapolation from the proportion linked in this subset. The third scenario corresponds to the standard situation updated with an estimation of the false negatives and the false positives.

In order to apply the capture-recapture method with stratification, it is necessary to have (for each scenario) the distribution of casualties' characteristics -NISS, road user type, third party- in each subset ("police only", "registry only", overlap between the registry and the police). Since scenarios two and three are estimated scenarios, we have to assume some distributions to be applied in each subset. We assumed the same distributions as in scenario 1: indeed, differences between the scenarios are only differences in efficiency of the record-linkage procedure; in other words, the additional matched casualties found with the names have no reason to be different in terms of NISS, road user type or third party to the casualties previously found with the standard record-linkage procedure. The distribution of these NISS, road user type and third party characteristics for each subset are provided in the results section.

Ascertainment rates of the registry, police file and of the aggregate are estimated; they are defined as the ratio between the number of cases reported by the registry, by the police, by the aggregate respectively, and the number of cases estimated by capture-recapture. Ascertainment-corrected incidence rates are estimated; the numerator is the number of cases estimated by capture-recapture and the denominator is the population of the Rhône county in 2001 provided by INSEE, the French National Institute of Statistics and Economics Studies.

3. RESULTS

In 2001 in the Rhône county, the police record 4135 non-fatal casualties and the registry records 10636 non-fatal casualties that fulfil the police criteria of a road crash casualty. The first scenario,

corresponding to the "standard" record-linkage based on date of crash, detailed place of crash, year and month of birth, gender and road user type of the casualty, identifies 2813 casualties common to both sources.
The additional survey on 1322 "police only" records is based on only 868 available paper crash reports (because of technical problems). We have checked that they are however representative of the whole subset as regards to police severity variable (slight vs. serious). Out of these 868 paper crash reports, all have surname and first name available, and 528 have a medical certificate available, on which NISS is evaluated.
The additional record-linkage on first name and surname identifies 94 additional matches out of these 868 police records with available paper crash reports. This corresponds to 10.8% of them, so that, out of the whole set of 1322 "police only" records, we estimate a total of 148 matches on names, additional to the "standard" record-linkage: scenario 2 has hence a total of 2956 linked casualties between the registry and the police file.
The third scenario is the "standard" scenario updated with the estimation of 97 false positives and 396 false negatives based on the probabilities computations and approximations given in the appendix. This yields a total of 3112 matches between the registry and police file.
Table 1 provides the distribution of NISS, road user type, and human third party in each subset: "police only", "registry only", overlap between the police and the registry, based on scenario 1 and which is used in scenario 2 and 3.
Results of the stratified capture-recapture analyses displayed by NISS categories but aggregated over road user type and third party strata are given in table 2, for each scenario. Detailed results (for each strata) for scenario 3 are given in table 3. Detailed results for scenario 1 and 2 are available upon request.

Table 1: distribution of NISS, road user type, (human) third party, of non-fatal casualties, all scenarios

	Police only n=1 322	Registry only n=7 823	Police and registry (overlap) n=2 813
Injury severity			
NISS 1-3	69.6%	79.2%	63.0%
NISS 4-8	21.7%	15.5%	20.6%
NISS 9+	8.7%	5.3%	16.4%
	100.0%	100.0%	100.0%
Road user type			
pedestrians	12.3%	6.8%	13.8%
cyclists	1.9%	14.8%	2.7%
motorcyclists	15.5%	21.0%	18.4%
car occupants	65.1%	51.6%	61.6%
others	5.2%	5.8%	3.5%
	100.0%	100.0%	100.0%
Third party			
with	83.7%	55.8%	83.8%
without	16.3%	44.2%	16.2%
	100.0%	100.0%	100.0%

Table 2: results of stratified capture-recapture, aggregated on road user type and third party, by scenario

	Police computer file	Road trauma registry	Overlap	Aggregate	Capture-recapture	Police ascertain. rate	Registry ascertain. rate	Aggregate ascertain. rate	Ascertain.-corrected incidence rate (per 100 000)
Scenario 1									
NISS 1-3	2 705	7 972	1 774	8 903	12 305	22.0%	64.8%	72.4%	768
NISS 4-8	861	1 791	580	2 071	2 757	31.2%	64.9%	75.1%	172
NISS 9+	570	873	459	984	1 067	53.4%	81.8%	92.2%	67
Total	4 135	10 636	2 813	11 858	16 129	25.6%	65.9%	74.1%	1 007
Scenario 2									
NISS 1-3	2 694	7 949	1 864	8 780	11 616	23.2%	68.4%	75.6%	725
NISS 4-8	860	1 798	610	2 048	2 621	32.8%	68.6%	78.1%	164
NISS 9+	581	889	483	987	1 057	55.0%	84.1%	93.4%	66
Total	4 135	10 636	2 956	11 815	15 294	27.0%	69.5%	77.3%	954
Scenario 3									
NISS 1-3	2 683	7 924	1 962	8 645	10 937	24.5%	72.5%	79.0%	683
NISS 4-8	859	1 806	642	2 023	2 487	34.5%	72.6%	81.4%	155
NISS 9+	594	906	508	992	1 048	56.7%	86.5%	94.7%	65
Total	4 135	10 636	3 112	11 659	14 471	28.6%	73.5%	80.6%	903

Table 3: scenario 3, results of stratified capture-recapture

road user type and third party	Injury severity	Police computer file	Registry	Overlap	Aggregate	Capture-recapture	Police ascertain. rate	Registry ascertain. rate	Aggregate ascertain. rate
pedestrians									
	NISS 1-3	284	593	207	670	811	35.0%	73.2%	82.6%
	NISS 4-8	141	200	112	229	251	55.9%	79.4%	91.0%
	NISS 9+	129	147	109	167	174	74.4%	84.6%	96.1%
cyclists									
	NISS 1-3	63	811	52	822	983	6.4%	82.4%	83.6%
	NISS 4-8	21	312	17	317	398	5.4%	78.4%	79.6%
	NISS 9+	19	74	15	78	89	21.0%	83.5%	87.1%
motorcyclists									
with third party	NISS 1-3	306	689	249	745	846	36.2%	81.4%	88.2%
	NISS 4-8	186	235	129	292	338	55.2%	69.5%	86.4%
	NISS 9+	157	189	138	207	214	73.4%	88.3%	96.9%
no third party	NISS 1-3	26	675	15	685	1090	2.4%	61.9%	62.9%
	NISS 4-8	32	259	18	273	455	7.0%	56.8%	59.9%
	NISS 9+	26	109	24	111	118	22.4%	92.5%	94.3%
car occupants									
with third party	NISS 1-3	1558	3652	1127	4084	5049	30.9%	72.3%	80.9%
	NISS 4-8	327	457	252	532	593	55.2%	77.0%	89.7%
	NISS 9+	161	189	128	222	238	67.8%	79.6%	93.5%
no third party	NISS 1-3	319	1096	233	1182	1500	21.3%	73.1%	78.8%
	NISS 4-8	130	248	98	280	328	39.6%	75.7%	85.4%
	NISS 9+	87	154	78	163	172	50.7%	89.7%	94.9%
others									
	NISS 1-3	127	408	79	457	658	19.3%	62.0%	69.4%
	NISS 4-8	21	97	17	102	124	17.3%	78.2%	82.1%
	NISS 9+	14	44	14	44	44	32.9%	100.0%	100.0%

In all three scenarios, the police ascertainment rate shows an increase with NISSS categories, from around 23% in the NISS 1-3 category to 55% in the NISS 9+ category. In all 3 scenarios, the registry ascertainment rate is the same in the NISS 1-3 and NISS 4-8 categories, at 65% in scenario 1 and at 73% in scenario 3. The registry ascertainment rate in the NISS 9+ category is above 80% in all scenarios. The aggregate ascertainment rate is above 90% in all three scenarios. The ascertainment-corrected incidence rate of NISS 1-3 casualties is at least 683 per 100 000, the ascertainment-corrected incidence rate of NISS 4-8 casualties is at least 155 per 100 000 and the ascertainment-corrected incidence rate of NISS 9+ casualties is at least 65 per 100 000.

4. DISCUSSION

The capture-recapture method is based on four assumptions. The first assumption is that the population is closed i.e. that there are no entries or losses between the two recordings. This is largely met since there is very little delay between the police attending the crash scene and the casualty's visit to the hospital: at most a few days for the most slightly injured casualties.

The second assumption is that there is independence between the two sources. We do believe that there is some positive dependence between the police file and the registry among the serious casualties. Indeed if the police are first at the scene of a serious crash, they will call the emergency care. The reverse, i.e. the emergency care alerting the police may be slightly less true. This is a dependency "at the scene" meaning that the dependency in terms of recorded lists would be less strong. When there is a dependency and it is known to be positive, as it is here, it has been established (IWGDMF 1995a) that the capture-recapture estimator is biased downwards. Therefore the obtained estimates of the total number of non-fatal casualties must be interpreted as lower bounds. The estimates of ascertainment rates must be interpreted as upper bounds.
The third assumption about homogeneous reporting probability is not met, but stratification is a way to deal with varying probabilities. The most important factors associated with varying reporting probabilities have been taken into account as stratifying variables, the major one being injury severity. Moreover, injury severity has been measured by a trauma severity scale, namely the New Injury Severity Score, based on detailed injury description provided and coded by physicians. Some other important factors associated in France with varying police reporting probabilities (Amoros et al. 2006) are the combination of road type, urban/rural area, type of police force; they play a role in the police reporting probability but one can doubt that they play a role in the registry reporting probability once injury severity is taken into account.
The fourth assumption of perfect identification of common cases is difficult to evaluate.
The additional record-linkage with names did find some more matches but not many more. These could not have been found with the "standard" record-linkage as they appeared to contain 2 or 3 errors in the major linking variables or a missing value for the place of crash (most often missing in the registry). We are confident with the efficiency of this additional record-linkage since it is based on many informative variables: month and year of birth, first and surname to identify the casualty and date and location to identify the crash, with location being very detailed: town and street/road name. It is however still probable that we fail to identify a number of pairs from the same casualties as the Soundex code is not optimal for French, South-European and North-African names which constitute the French population. Scenario 3 is therefore the most realistic scenario even though it is based on some approximations. Moreover, the three scenarios yield estimates that are within a reasonnable range.
By definition the police aim to report all casualties from road traffic crashes, whereas the registry aims to report all casualties that seek care in health care facilities, in other words, at hospitals. Many of the slight casualties will not go to the hospital but instead will go and see their general practitioner or take care of their injuries themselves. Therefore we can not interpret the registry ascertainment rate as a measure of completeness of the registry in the NISS 1-3 and NISS 4-8 categories. It is only in the serious casualties' category that the registry ascertainment rate has its full meaning. The same is hence true for the aggregate ascertainment rate.
The police report at most 25% of the slight road traffic casualties and at most 57% of the serious ones. The registry reports at most 87% of the serious casualties, and altogether, the registry and the police report at most 95% of the serious road traffic casualties. The ascertainment-corrected incidence rate of NISS 9+ road non-fatal casualties is at least 65 per 100 000, in the Rhône county, France.
The capture-recapture method has allowed us to quantify the degree of completeness of the registry for serious non-fatal casualties. It is rather high, though not satisfactory; the aggregate ascertainment rate is satisfactory. The lack of completeness of the registry must be due, in the health care units, to the non-identification of the circumstances of injuries as road crashes, or merely to the oversight of filling registry notification forms.

Acknowledgments:

We wish to thank the following people for having participated in the data collection and data entry in the framework of the Association for the Registry of road traffic casualties in the Rhône (ARVAC : president YN Martin) or in the framework of INRETS-UMRESTTE (B. Laumon, scientific consultant for the registry, and A. Ndiaye, physician, director of the registry) : 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, Bec JF, Bejui J, Bel JC, Bérard E, Bérard J, Bernard JC, Berthet N, Bertrand JC,

Besson L, Biot B, Biot V, Blanc C, Blanchard J, Bœuf C, Boisson D, Bonjean M, Bost J, Bouchedor C, Bouletreau P, Boyer V, Breda Y, Brilland R, Bussery S, Cabet N, Caillot JL, Cannamela A, Caregnato B, Carre M, Catala Y, Chagnon PY, Chambost M, Chantran C, Chardon P, Charnay P, Chatelain P, Chattard S, Chavane H, Chazot G, Chettouane I, Chevreton N, Chevrillon E, Chevrillon S, Chiron M, Chotel P, Cochard P, Combe C, Contamin B, Coppard E, Cot T, Crettenet Z, Cristini A, Dal Gobbo B, De Angelis MP, Decourt L, Delfosse A, Demazière J, Deruty R, Desjardins G, Devaux J, Dohin B, Emonet A, Escarment J, Eyssette M, Fallavier L, Fanton L, Felten D, 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, Goubsky A, Granger M, Grattard P, Gueniaud PY, Guenot C, Guérin AC, Guignand M, Haddak M, Hamel D, Heckel T, Herzberg G, Jacquemard C, Javouhey E, Joffre T, Kohler R, Lablanche C, Lafont S, Lagier C, Lapierre B, Laplace MC, La Rosa C, Laurent R, Lebel M, Leblay G, Le-Xuan I, Lieutaud T, Lille R, Linné M, Lucas R, Machin B, Maello E, Malicier D, Mangola B, Marduel YN, Marie-Catherine M, Martin JL, Martinand G, Marty F, Mazouzi S, Messikh C, Meyer F, Meyrand S, Molard S, Morel-Chevillet E, Mioulet E, Minjaud F, Mollet C, Monnet J, Moyen B, Neidhart JP, Ngandu E, Ny S, Ould T, Paget P, Paillot JC, Paris D, Patay B, Pauget P, Peillon D, Perrin-Blondeau D, Petit P, Piton JL, Plantier M, Pornon P, Pramayon C, Quelard B, Raquin L, Rezig M, 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, Sbraire N, Scappaticci N, Schiele P, Schneider M, Simonet C, Sindou M, Soldner R, Soudain M, Stagnara J, Stamm D, Suc B, Taesch MC, Tasseau F, Tell L, Thomas M, Tilhet-Coartet S, Tissot E, Toukou JC, Trifot M, Vallee B, Vallet G, Vancuyck A, Vergnes I, Verney MP, Voiglio EJ, Vourey G, Vuillard J, Westphal M, Willemen L.

References

Amoros, E., Martin, J. L. and Laumon, B., 2006. Under-reporting of road crash casualties in France. Accident Analysis and Prevention 38 (4), 627-635.

Aptel, I., Salmi, L. R., Masson, F., Bourdé, A., Henrion, G. and Erny, P., 1999. Road accident statistics: discrepancies between police and hospital data in a French island. Accident Analysis and Prevention 31 (1-2), 101-108.

Brenner, H., 1994. Application of capture-recapture methods for disease monitoring: potential effects of imperfect record linkage. Methods of Information in Medicine 33 (5), 502-506.

Brenner, H. and Schmidtmann, I., 1996. Determinants of homonym and synonym rates of record linkage in disease registration. Methods of Information in Medicine 35 (1), 19-24.

Clark, D. E., 2004. Practical introduction to record linkage for injury research. Injury Prevention 10 (3), 186-191.

Dhillon, P. K., Lightstone, A. S., Peek-Asa, C. and Kraus, J. F., 2001. Assessment of hospital and police ascertainment of automobile versus childhood pedestrian and bicyclist collisions. Accident Analysis and Prevention 33 (4), 529-537.

International Working Group for Disease Monitoring and Forecasting (IWGDMF), 1995a. Capture-recapture and multiple-record systems estimation I: History and theoretical development. American Journal of Epidemiology 142 (10)

International Working Group for Disease Monitoring and Forecasting (IWGDMF), 1995b. Capture-recapture and multiple-record systems estimation II: Applications in human diseases. American Journal of Epidemiology 142 (10), 1059-1068.

Gallay, A., Nardone, A., Vaillant, V. and Desenclos, J. C., 2002. [The capture-recapture applied to epidemiology: principles, limits and application]. Revue d'Epidémiologie et de Santé Publique 50 (2), 219-232.

Hook, E. B. and Regal, R. R., 1995. Capture-recapture methods in epidemiology: methods and limitations. Epidemiologic Reviews 17 (2), 243-264.

Howe, G. R., 1998. Use of computerized record linkage in cohort studies. Epidemiologic Reviews 20 (1), 112-121.

Jarvis, S. N., Lowe, P. J., Avery, A., Levene, S. and Cormack, R. M., 2000. Children are not goldfish - mark/recapture techniques and their application to injury data. Injury Prevention 6, 46-50.

Laumon, B., Martin, J. L., Collet, P., Chiron, M., Verney, M. P., Ndiaye, A. and Vergnes, I., 1997. A French road accident trauma registry: first results. In: 41st annual conference of the Association for the Advancement of Automotive Medecine, Orlando, Florida, 127-137

Meuleners, L. B., Lee, A. H., Cercarelli, L. R. and Legge, M., 2006. Estimating crashes involving heavy vehicles in Western Australia, 1999-2000: A capture-recapture method. Accident Analysis and Prevention 38 (1), 170-174.

Morrison, A. and Stone, D. H., 2000. Capture-recapture: a useful methodological tool for counting traffic related injuries? Injury Prevention 6 (4), 299-304.
Newcombe, H. B., 1988. Handbook of record linkage: methods for health and statistical studies, administration, and business. Oxford university press, Oxford.
Razzak, J. A. and 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. and Scragg, R., 1994. Application of capture-recapture methodology to estimate the completeness of child injury surveillance. Journal of Paediatrics and Child Health 30 (6), 513-514.
Tercero, F. and Andersson, R., 2004. Measuring transport injuries in a developing country: an application of the capture-recapture method. Accident Analysis and Prevention 36 (1), 13-20.
Wittes, J., 1972. On the bias and estimated variance of Chapman's two-sample capture-recapture population estimate. Biometrics 28, 592-597.

Appendix: estimating the frequency of false positives and false negatives

The reasoning in this appendix is largely inspired by the papers of Brenner (1994) and Brenner and Schmidtmann (1996);

1. False negatives

P (false negative) =
P (deciding a non-link / records from the same casualties)
=P (disagreement on all or some of the linking variables / same casualties)
=1-P (agreement on all or some of the linking variables / same casualties)

"on all or some of the linking variables " is defined as follows: the linking variables are:
1) day of crash, 2) month of crash, 3) place of crash, 4) month of birth of the casualty, 5) year of birth, 6) gender, 7) road user type
Comment 1: Year of crash is not a linking variable since it is used as a blocking variable, i.e. the record-linkage procedure is conducted by blocks of annual data. This assumes that there is no error on this variable.
Comment 2: if at least one of the variables of the comparison has a missing value, we consider the result of the comparison as a disagreement.
We decide to link if:
- all the 7 variables agree
- only 6 variables agree
- only 5 variables agree, provided that disagreement is on 2 variables among road user type, gender and month of birth of the casualty.

These 3 situations being separate, the probability of the whole is the sum of the probabilities, and each one, assuming independence between the linking variables, is the product over the linking variables, of the probability of agreement. Each probability is estimated by the observed proportion.
This is a simplification of the linking decision since 1) in the linking process we could use information from linking the crash first, 2) in the computation of the observed proportion of agreement, place of location is simplified to the postcode of town or village; in the actual record-linkage process it is the full detail (town/village and street/road name) that is used.

As we do not have a gold standard, we can use the subset of the linked pairs as a rough approximate of the subset of pairs from the same casualties. Because of this and because we have been rather cautious in deciding a link, the observed proportions are over-estimations and should be scaled down. This is confirmed with the figures found in the literature (Howe 1998; Newcombe 1988). An arbitrary 5% reduction is applied. The results obtained are provided in the table below:

Linking variables	Agreement among the linked pairs (n=2 813)	-5% reduction
Day of crash	97.3%	92%
Month of crash	99.9%	95%
Year of crash (blocking variable)	*100%*	*100%*
Place of crash (town/village postcode)	82.9%	78%
Month of birth	95.8%	91%
Year of birth	96.5%	92%
Gender	97.1%	92%
Road user type	96.1%	91%
Probability of agreement / same casualties	96.4%	87%
Probability of false negatives	3.6%	13%
Frequency of false negatives	109	396

This yields an estimate of 396 false negatives.

2. False positives

P (false positive) =
=P (deciding a link/records from distinct casualties)
=P (agreement on all or some of the linking variables/ distinct casualties)
"on all or some of the linking variables" is defined as above.

These probabilities correspond to the probability of agreement by chance. For linking variables such as month or day of birth, or of crash, whose modalities are uniformly distributed (providing large enough files), these probabilities of agreement by chance would be 1/12 for month and about 1/30 for day. For other linking variables, these can be approximated by the observed proportions of agreement of the linking variables, with the subset of pairs from distinct casualties being approximated by the subset of pairs that we have not linked.

As already mentioned for place of crash, only the proportion of agreement on town/village can be computed (using postcodes). However, in the record-linkage process, it is agreement on town/village and street(s)/road(s) name that matters so that agreement by chance must be very low. Therefore, the probability must be reduced; an arbitrary 1% probability is chosen. Results are provided in the table below.

Linking variables	Agreement among the non-linked pairs (n=47 220 438)	reduction for place of crash
Day of crash	3.2%	3.2%
Month of crash	8.4%	8.4%
Year of crash (blocking variable)	*100.0%*	*100.0%*
Place of crash	8.0%	1.0%
Month of birth	8.3%	8.3%
Year of birth	2.0%	2.0%
Gender	52.6%	52.6%
Road user type	38.2%	38.2%
Probability of agreement / distinct casualties	1.1E-5*	2.1E-6*
Probability of false positives	1.1E-5*	2.1E-6*
Frequency of false positives	486	97

***** the probability of false positives is defined on the subset of pairs of records of distinct casualties whose frequency is 47 million

It is estimated that 97 false positives are expected with this record-linkage procedure.

Revue bibliographique sur la méthode de projection

Cadre fourni par la théorie des sondages

Nous avons abordé ce sujet par l'angle de la théorie des sondages, et dans ce cadre, par la notion de redressement (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.

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

avec h=identifiant des post-strates

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.

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 Ui résidus de somme nulle

et donc l'estimateur redressé s'écrit :

$\hat{\bar{Y}} = \bar{y} \times \frac{\bar{X}}{\bar{x}}$, avec $R \approx \frac{\bar{y}}{\bar{x}}$, 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 Ui résidus de somme nulle

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

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.

Revue bibliographique des études épidémiologiques : projection du régional au national

L'extrapolation qui nous intéresse, c'est l'extrapolation d'indicateurs de santé (incidence en particulier, prévalence) 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.

Nous avons trouvé des publications en épidémiologie des addictions (drogues), du cancer, des traumatismes… et nous avons classé les méthodes rencontrées.

1) 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).

2) Extrapolation en ajustant sur facteurs de confusion

Extrapolation en ajustant sur catégories d'âge et de sexe

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, Cannan 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, et 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 sur âge et sexe ; il s'agit aussi de standardisation indirecte, selon la terminologie en épidémiologie, où le nombre attendus 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).

Extrapolation en ajustant sur des facteurs de confusion quantitatifs
Principe :
Une relation entre l'incidence d'une maladie et des caractéristiques régionales est observée et modélisée, et on applique ensuite cette relation à la population générale.

Exemples :
- estimation de l'incidence de cancer sur tous les Etats-Unis à partir de la relation entre l'incidence de cancer et des variables socio-économiques et ethniques ; relation établie pour certains Etats, puis appliquée aux autres Etats (Frey, Feuer 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, Schau et al. 2005).

Détail de la méthode (avec variables agrégées):
Etape 1 : obtention, sur plusieurs régions, de l'incidence observée et de variables corrélées agrégées et observées ;
Etape 2 : modélisation de l'incidence (observée) en fonction des variables régionales observées ; la modélisation est de type modèle linéaire généralisé. ;
Etape 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 (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 de référence et dans la région d'application (ou dans le territoire tout entier).

3) Extrapolation à partir d'une morbidité corrélée ou de la mortalité associée

a) Extrapolation en fonction d'un seul indicateur similaire
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 au Royaume-Uni (Gossop, Strang et al. 1994)

- estimation de l'incidence des cancers (par site) 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, Esteve 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'Etat de Caroline du Nord (au nombre de 10 et 15 hôpitaux) et appliquée à tout l'Etat (Stutts, Williamson et al. 1990).
- 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, Uhry et al. 2004).

Un exemple un peu différent : l'indicateur similaire est exactement le même mais évalué par une autre source : estimation du nombre de victimes en fonction du nombre de victimes recensées par la police, par application de taux de correction moyens (et semble-t-il avec post-stratification sur le type d'usager, et sur la gravité) : relation observée sur 6 études (sans doute 6 régions ou 6 hôpitaux), et appliquée à la Grande-Bretagne (James 1991).

Détail de la méthode utilisant l'incidence d'une morbidité associée :
Etape 1 : obtention, sur une (ou des) région(s), de l'incidence d'intérêt et de l'incidence d'une morbidité similaire et corrélée ;
Etape 2 : modélisation de l'incidence d'intérêt (observée) en fonction de l'incidence (observée) de cette morbidité similaire ; la modélisation est de type proportionnelle ;
Etape 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.

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.

<u>b) Extrapolation en fonction d'un indicateur similaire et de caractéristiques macroscopiques</u>

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) 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 et appliquée à toute la France (Menegoz, Black 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, Hickman et al. 2001).

Cela correspond au redressement par régression (et 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.

Intervalles de confiance

Quelques articles fournissent des résultats avec des intervalles de confiance, mais aucun parmi ceux qui font de la projection par régression.

Articles qui présentent des intervalles de confiance. :
(Jones, Cannan et al. 2000) concernant les fractures : taux d'incidence standardisé estimé avec son intervalle de confiance
(Gossop, Strang et al. 1994) : le ratio (entre effectifs d'héroinomanes et de cocainomanes) est estimé avec sa variance et son intervalle de confiance ; le nombre de cas prévalents est aussi accompagné d'un intervalle de confiance
(Frischer, Hickman et al. 2001) à partir d'enquêtes par sondage : effectifs prévalents avec intervalle de confiance ; résultats avec autres méthodes : pas d'intervalle de confiance (mais des « fenêtres » d'après les résultats de différentes études)
(Nachbaur, Uhry et al. 2004) : taux d'incidence (d'accidents de la vie courante) et leurs intervalles de confiance.

Articles sans présentation d'intervalles de confiance :
(Jensen, Esteve et al. 1990) : taux annuels d'incidences (de cancer), sans intervalle de confiance
(Frey, Feuer et al. 1994) : taux et effectifs incidents (de cancer), sans intervalle de confiance
(Takala 1999) : effectifs incidents (d'accident professionnels mortels), sans intervalle de confiance
(Menegoz, Black et al. 1997) : effectifs incidents (de cancer), sans intervalle de confiance
(Foster and Johnson 2001) : effectifs prévalents (glaucome), sans intervalle de confiance
(James 1991) : effectifs incidents (de blessés de la route), sans intervalle de confiance, mais présentés comme minima
(Peabody, Schau et al. 2005) : effectifs prévalents (de COPD), sans intervalle de confiance

Références :

Ardilly, P. (1994). Les techniques de sondage. Paris, Editions Technip.
Foster, P. J. and G. J. Johnson (2001). "Glaucoma in China: how big is the problem?" British Journal of Ophthalmology **85**(11): 1277-82.
Frey, C. M., E. J. Feuer, et al. (1994). "Projection of incidence rates to a larger population using ecologic variables." Statistics in Medecine **13**(17): 1755-70.
Frischer, M., M. Hickman, et al. (2001). "A comparison of different methods for estimating the prevalence of problematic drug misuse in Great Britain." Addiction **96**(10): 1465-76.
Gossop, M., J. Strang, et al. (1994). "A ratio estimation method for determining the prevalence of cocaine use." Bristish Journal of Psychiatry **164**(5): 676-9.
James, H. F. (1991). "Under-reporting of road traffic accidents." Traffic Engineering and Control **32**: 573-583.
Jensen, O. M., J. Esteve, et al. (1990). "Cancer in the European Community and its member states." European Journal of Cancer **26**(11-12): 1167-1256.
Jones, I. E., R. Cannan, et al. (2000). "Distal forearm fractures in New Zealand children: annual rates in a geographically defined area." New Zealand Medicine Journal **113**(1120): 443-5.
Menegoz, F., R. J. Black, et al. (1997). "Cancer incidence and mortality in France in 1975-95." European Journal of Cancer Prevention **6**(5): 442-466.
Nachbaur, C., Z. Uhry, et al. (2004). Estimation du taux d'incidence annuel. d'accidents de la vie courante, en France, en 2001. Journées scientifiques de l'InVS., Paris.
Peabody, J. W., B. Schau, et al. (2005). "COPD: A prevalence estimation model." Respirology **10**(5): 594-602.
Stutts, J. C., J. E. Williamson, et al. (1990). "Bicycle accidents and injuries: A pilot study comparing hospital- and police-reported data." Accident Analysis and Prevention **22**(1): 67-78.
Takala, J. (1999). "Global estimates of fatal occupational accidents." Epidemiology **10**(5): 640-6.

Estimation du nombre de faux négatifs et de faux positifs sur le résultat du chaînage des données des années 1996 à 2004.

La différence par rapport à l'estimation des nombres de faux négatifs et de faux positifs sur les données de 2001 vient du changement de stratégie de chaînage à partir de 2002. En effet à partir de 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 ; l'idée est de récupérer une partie des informations contenues dans les données des forces de l'ordre pour les copier dans le Registre. Cela conduit à une décision de rapprochement qui est basée sur des critères plus restrictifs. Concrètement : seulement une discordance sur une seule variable est tolérée (et cela ne doit pas porter sur le mois de l'accident).

Pour estimer le nombre de faux négatifs, on n'utilise pas les proportions de concordance chez les positifs estimées sur 2002-2004, car elles sont trop dépendantes du changement de stratégie, mais on utilise la moyenne de celles estimées sur 1996-2001.

Pour estimer le nombre de faux positifs, on utilise bien les proportions de concordance chez les négatifs observées sur 2002-2004 car elles reflètent un réel changement de répartition de certaines variables (type d'usager et commune : plus de valeurs manquantes)

Faux positifs :

Un faux positif est une paire d'enregistrements de 2 victimes distinctes que l'on a reliée.

P(faux positifs)
=P(décider un lien alors qu'il s'agit d'enregistrements de 2 victimes distinctes)
=P(décider un lien / victimes distinctes)
=P(concordance sur toutes ou une partie des variables de chaînage / 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

Sur la période 1996-2001, la décision de relier la paire est la même que pour les données de 2001, c'est-à-dire, 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.

Les calculs de probabilité correspondant sont exactement les mêmes que ceux donnés dans l'annexe sur l'estimation des nombres de faux négatifs et faux positifs sur l'année 2001.

Sur la période 2002-2004, on décide un lien entre deux enregistrements s'il y a :
- concordance sur les sept variables, 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(décider un lien/ victimes distinctes)
$= P(7 \text{ var concordent} / \text{ victimes distinctes}) + P(\text{exact.6 var concordent (sf mois)} / \text{ victimes distinctes})$

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

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

La probabilité de chaque situation s'écrit, en supposant l'indépendance des variables de chaînage, comme le produit des probabilités de concordance pour chaque variable de chaînage.
Nous rappelons que la concordance des valeurs d'une variable sur des victimes distinctes correspond à une concordance par hasard. L'ensemble des paires d'enregistrements de victimes distinctes est approximé par l'ensemble des paires non-reliées.

Tableau 13 : proportions de concordance sur les variables de chaînage, observées sur l'ensemble des paires non reliées, registre et BAAC, Rhône,

variables de 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%

Les proportions de concordance (par hasard) sont liées à la répartition des modalités. Les valeurs différentes observées entre les années correspondent à une plus grande proportion de valeurs manquantes dans le type d'usager et dans la commune (lieu d'accident) dans les années récentes.

Concernant le lieu de l'accident, nous minorons la proportion obtenue car celle-ci est la concordance par hasard sur la commune, alors 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 à 1% est arbitrairement choisi.

Calcul du nombre de faux positifs :

Nous notons :
FP = faux positifs
FN = faux négatifs
VN =vrais négatifs
VP = vrais positifs
#= effectif

alors :
#FP = P(FP) × #VN

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

Tableau 14 : proportions de concordance sur U, de faux positifs et effectifs de faux positifs ; résultats du chaînage Registre et BAAC, Rhône, par année

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

Faux négatifs

Un faux négatif est une paire d'enregistrements de la même victime pour lesquels on a décidé un non-lien (à tort donc).

P(faux négatif)
=P(décider un non-lien alors qu'il s'agit de 2 enregistrements d'une même victime)
=P(décider un non-lien / 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.

Sur la période 1996-2001, la décision de relier la paire est la même que pour les données de 2001, c'est-à-dire, 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.
Les calculs de probabilité correspondant sont exactement les mêmes que ceux donnés dans l'annexe sur l'estimation des nombres de faux négatifs et faux positifs sur l'année 2001.

Sur la période 2002-2004, 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 chaque situation s'écrit comme le produit des probabilités de concordance pour chaque variable de chaînage (entre les deux enregistrements) :

P(concordance sur toutes ou une partie des variables de chaînage / même victime)
$= P(7 \text{ variables concordent} / \text{ même victime}) + P(6 \text{ variables concordent} / \textit{même} \text{ victime})$

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

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

Ne disposant pas du gold standard, nous approximons l'ensemble des vrais positifs par l'ensemble des positifs autrement dit par les paires rapprochées. Les paires rapprochées sur-estiment la probabilité de concordance. Ceci est d'autant plus vrai à partir de 2002, où la décision de relier est plus conservatrice. Pour que les proportions pour ces années-là ne soient pas dépendantes du changement

dans la stratégie de chaînage, nous utilisons les proportions calculées sur l'ensemble de la période 1996-2001. Ces proportions restent des proportions estimées sur les paires reliées, qui sur-estiment donc la probabilité de concordance sur les vrais positifs. De la même façon que pour les estimations sur les données de 2001, nous minorons arbitrairement de 5% les proportions observées. Cela nous donne des probabilités de concordance proches de celles trouvées dans la littérature (cf. annexe sur le même sujet sur les données de 2001).

Tableau 15 : proportions de concordance sur les variables de chaînage, observées sur les paires reliées, Registre et BAAC, Rhône, 1996-2001

variables de chaînage	1996 (n=2919)	1997 (n=2616)	1998 (n=3169)	1999 (n=3040)	2000 (n=2761)	2001 (n=2809)	moyenne
jour accident	97,7%	97,9%	98,8%	98,3%	98,5%	97,3%	98,1%
mois accident	99,9%	99,8%	99,9%	99,8%	99,8%	99,9%	99,9%
année accident	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
lieu de l'accident (commune)	85,4%	79,7%	82,4%	82,1%	80,6%	82,0%	82,1%
mois de naissance	88,5%	95,9%	94,5%	95,5%	95,8%	95,8%	94,3%
année de naissance	92,3%	97,0%	95,8%	96,2%	97,0%	96,5%	95,8%
sexe	96,9%	96,3%	96,9%	95,9%	96,0%	97,0%	96,5%
type d'usager	95,9%	94,3%	95,6%	95,7%	95,1%	96,1%	95,5%

Tableau 16 : 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%	93%
mois accident	95%	95%	95%	95%	95%	95%	95%
année accident	100%	100%	100%	100%	100%	100%	100%
lieu de l'accident (commune)	81%	76%	78%	78%	77%	78%	78%
mois de naissance	84%	91%	90%	91%	91%	91%	90%
année de naissance	88%	92%	91%	91%	92%	92%	91%
sexe	92%	91%	92%	91%	91%	92%	92%
type d'usager	91%	90%	91%	91%	90%	91%	91%

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 :

$\# FN = P(FN) \times \# VP$

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

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

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

Tableau 17 : proportions de concordance sur M, de faux négatifs, d'effectifs de faux négatifs ; résultats du chaînage Registre et BAAC, Rhône, par année

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

Résultats du modèle de capture-recapture (multinomial logit) ajusté sur cinq facteurs de biais, sur les données du Rhône 1996-2004

Tableau 18 : résultats du modèle de capture-recapture (multinomial logit) ajusté sur NISS, type d'usager, tiers, type de réseau et force de l'ordre et agrégé sur le NISS, sur les données du Rhône 1996-2004

Gravité lésionnelle	Effectif BAAC	Effectif Registre	Effectif intersection	Effectif union des deux	Effectif capture-recapture	Taux de couverture BAAC	Taux de couverture du registre	Taux de couverture de l'union des 2
NISS 1-3	2512	6788	1774	7526	9387	26,8	72,3	80,2
NISS 4-8	826	1615	593	1848	2131	38,8	75,8	86,7
NISS 9+	698	869	529	1038	1129	61,8	76,9	91,9
total	4037	9272	2896	10412	12647	31,9	73,3	82,3

Tableau 19 : résultats du modèle de capture-recapture (multinomial logit) ajusté sur NISS, type d'usager, tiers, type de réseau et force de l'ordre, sur les données du Rhône 1996-2004

Type d'usager et tiers	Gravité lésionnelle	Effectif BAAC	Effectif Registre	Effectif intersection	Effectif union des deux	Effectif capture-recapture	Taux de couverture BAAC	Taux de couverture du registre	Taux de couverture de l'union des 2
<u>piétons</u>									
Avec tiers	NISS 1-3	270	550	196	625	780	34,6	70,5	80,0
	NISS 4-8	153	197	111	239	271	56,3	72,7	88,1
	NISS 9+	157	167	119	205	218	72,1	76,6	94,0
<u>cyclistes</u>									
Avec tiers	NISS 1-3	66	200	48	218	276	23,8	72,5	78,9
	NISS 4-8	34	67	25	77	90	38,2	74,4	85,2
	NISS 9+	27	34	20	41	45	60,7	75,4	90,8
Sans tiers	NISS 1-3	5	550	4	551	643	0,8	85,5	85,6
	NISS 4-8	3	237	3	238	269	1,2	88,2	88,3
	NISS 9+	4	59	3	59	67	5,4	88,1	89,0
<u>usagers de 2RM</u>									
Avec tiers	NISS 1-3	271	575	193	654	803	33,8	71,7	81,4
	NISS 4-8	145	201	105	242	273	53,1	73,6	88,5
	NISS 9+	175	176	134	216	231	75,8	76,2	93,8
Sans tiers	NISS 1-3	46	560	38	569	682	6,8	82,1	83,4
	NISS 4-8	34	237	27	245	281	12,2	84,2	86,9
	NISS 9+	42	103	34	111	124	33,7	83,0	89,4
<u>automobilistes</u>									
Avec tiers	NISS 1-3	1456	3167	1023	3600	4471	32,6	70,8	80,5
	NISS 4-8	314	402	226	491	564	55,8	71,4	87,1
	NISS 9+	178	179	134	223	240	74,0	74,8	92,9
Sans tiers	NISS 1-3	294	917	204	1007	1334	22,0	68,8	75,5
	NISS 4-8	117	215	81	251	303	38,7	71,1	82,9
	NISS 9+	98	123	71	149	167	58,5	73,3	89,2
<u>autres</u>									
	NISS 1-3	72	96	43	125	164	44,0	58,3	76,0
	NISS 4-8	14	14	8	20	23	59,5	62,5	85,5
	NISS 9+	9	8	6	11	12	73,1	69,4	92,3
		33	172	26	179	234	14,1	73,6	76,7
	NISS 1-3	11	44	8	46	56	19,5	77,6	82,7
	NISS 4-8	9	19	7	22	25	36,5	78,0	87,2
	NISS 9+	72	96	43	125	164	44,0	58,3	76,0

Les tableaux ci-dessous donnent les résultats pour l'année 2001

Tableau 20 : résultats du modèle de capture-recapture (multinomial logit) ajusté sur NISS, type d'usager, tiers, type de réseau et force de l'ordre, agrégé sur le NISS, sur les données du Rhône 2001

Gravité lésionnelle	Effectif BAAC	Effectif Registre	Effectif intersection	Effectif union des deux	Effectif capture-recapture	Taux de couverture BAAC	Taux de couverture du registre	Taux de couverture de l' union des 2
NISS 1-3	2594	7651	1974	8271	10378	25,0	73,7	79,7
NISS 4-8	843	1711	640	1914	2225	37,9	76,9	86,0
NISS 9+	645	863	506	1001	1089	59,2	79,2	91,9
	4081	10225	3120	11186	13692	29,8	74,7	81,7

Tableau 21 : résultats du modèle de capture-recapture (multinomial logit) ajusté sur NISS, type d'usager, tiers, type de réseau et force de l'ordre, sur les données du Rhône 2001

Type d'usager et tiers	Gravité lésionnelle	Effectif BAAC	Effectif Registre	Effectif intersection	Effectif union des deux	Effectif capture-recapture	Taux de couverture BAAC	Taux de couverture du registre	Taux de couverture union des 2
piétons									
Avec tiers	NISS 1-3	272	587	212	646	811	33,5	72,4	79,7
	NISS 4-8	143	189	109	223	256	55,8	74,0	87,2
	NISS 9+	133	142	108	168	178	74,6	79,8	94,0
cyclistes									
	NISS 1-3	63	728	52	738	880	7,1	82,7	83,9
	NISS 4-8	23	283	18	288	332	6,9	85,3	86,8
	NISS 9+	19	67	17	70	77	25,0	87,5	90,9
usagers de 2RM									
Avec tiers	NISS 1-3	276	643	218	701	866	31,9	74,3	81,0
	NISS 4-8	137	205	106	237	271	50,8	75,8	87,6
	NISS 9+	160	175	127	208	222	71,8	78,9	93,8
Sans tiers	NISS 1-3	55	668	44	678	815	6,7	81,9	83,2
	NISS 4-8	45	261	37	269	311	14,4	83,9	86,4
	NISS 9+	52	121	44	128	143	36,1	84,4	89,6
automobilistes									
Avec tiers	NISS 1-3	1501	3571	1126	3946	4944	30,4	72,2	79,8
	NISS 4-8	340	448	254	533	616	55,1	72,7	86,5
	NISS 9+	167	179	123	223	240	69,7	74,5	92,9
Sans tiers	NISS 1-3	324	1081	249	1157	1532	21,2	70,6	75,5
	NISS 4-8	130	235	98	267	324	40,1	72,5	82,5
	NISS 9+	100	145	78	167	186	53,9	77,7	89,9
autres									
	NISS 1-3	103	373	72	404	530	19,4	70,4	76,3
	NISS 4-8	25	90	19	96	116	22,0	77,7	83,4
	NISS 9+	13	33	10	36	42	32,3	79,1	87,4

Extrait de programme SAS : modèle de capture-recapture

```
title 'capture recapture, données BAAC Rhone U registre,1996- 2004';

*sans format pour source (plus simple pour la suite);
proc logistic data=coll9604  descending ;
class source(param=ref) nisscodeC (param=ref) usertiersC(param=ref ref='carAvecT')
corpsC(param=ref ref='Gend')
cadmC(param=ref ref='VC') ;
freq nveffarr;
model source=  nisscodeC  usertiersC  corpsC cadmC anqt /link=glogit ;
ods output ParameterEstimates=est;
run;

title 'paramètres estimés du modèle capture-recapture';
proc print;
run;

data est;
set est (keep =variable classVal0 response estimate );
format nom $3. ;
nom='R'||response; * paramètres theta;
run;
proc sort data=est;
by variable classval0;
run;
proc transpose data=est out=estwide;
id nom;
by variable Classval0;
var estimate;
run;

*ajouter les lignes des categories de reference;
data reference;
input variable $10.  ClassVal0 $8. R3 f2.0 R2 f2.0;
cards;
NISScodeC NISS9P  0 0
usertiersCcarAvecT0 0
corpsC    Gend    0 0
cadmC     VC      0 0
;
run;

data estwide;
set estwide reference;
run;

proc sort data=estwide;
by variable classval0;
run;

*parametres k;
data estwide (drop=R3 R2 _name_);
set estwide ;
k2=R3;
k1=R3-R2;
run;
```

```
*créer un dataset par variable, stockant les estimations des paramètres;
%macro beta(nom, nomC);
data capt.beta_&nom ;
set estwide ;
if variable=&nomC;
run;
proc print ;run;
%mend;
%beta(Intercept, %str('Intercept'));
%beta(NISScodeC, %str('NISScodeC'));
%beta(usertiersC, %str('usertiersC'));
%beta(corpsC, %str('corpsC'));
%beta(cadmC, %str('cadmC'));
%beta(anqt, %str('anqt'));

*on travaille maintenant sur  BAAC france, ;
data collBF;
set class.collBF;
run;

*on récupère, pour chaque modalité, les parametres k associés;
%macro recupbeta(nom, nomC);
proc sql;
create table collBF as
select v.*, i.k1 as k1_&nom, i.k2 as k2_&nom
from collBF v
left join capt.beta_&nom  i on
v.&nomC=i.Classval0;

proc freq data=collBF;
tables &nomC*k1_&nom &nomC*k2_&nom;
run;
%mend;

%recupbeta(NISScodeC, NISSCodeC);
%recupbeta(usertiersC, usertiersC);
%recupbeta(corpsC, corpsC);
%recupbeta(cadmC, cadmC);
proc sql;
create table collBF as
select v.*, i.k1 as k1_anqt, i.k2 as k2_anqt
from collBF v, capt.beta_anqt i;
proc sql;
create table collBF as
select v.*, i.k1 as k1_intercept, i.k2 as k2_intercept
from collBF v, capt.beta_intercept i;

*on estime PI0, la proba d etre recensé par aucune des 2 sources;
*PI1, proba d'etre recensé dans les BAAC ;
* et coefficient de correction sous-enregistrement BAAC;
*proba PI, non conditionnelles aux observés ;
data collBF;
set collBF;
eta1=k1_intercept + k1_NISSCodeC + k1_usertiersC  + k1_corpsC +k1_cadmC + anqt * k1_anqt;
eta2=k2_intercept + k2_NISSCodeC + k2_usertiersC  + k2_corpsC +k2_cadmC + anqt * k2_anqt;
PI0 = 1/( ( (1+exp(eta1) ) * (1+exp(eta2)) ));
PI1= 1 / (1 + exp (-eta1));
corr_BAAC_m= 1+exp(-eta1); * inverse de PI1;
run;
```

Traitement des valeurs manquantes : imputation simple

Valeurs manquantes pour les variables âge ou sexe
Imputation de l'âge manquant en fonction de sexe et type d'usager : la classe d'âge est tirée au sort, d'après une distribution multinomiale, estimée par la répartition observée des classes d'âge en fonction des catégories de sexe et de type d'usager.

Nous procédons de la même façon pour le sexe, imputé en fonction de type d'usager et classe d'âge.

Valeurs manquantes pour le type de réseau et/ou la force de l'ordre
Nous rappelons que cela ne concerne pas les données des forces de l'ordre : la force de l'ordre y est toujours renseignée (c'est celle qui écrit le procès-verbal) ; le type de réseau aussi. Les situations où type de réseau et/ou force de l'ordre sont manquants ne se rencontrent que dans les données du Registre. Nous rappelons que la force de l'ordre en charge d'un secteur est déterminable en fonction de la commune et éventuellement en fonction du type de réseau (où a eu lieu l'accident).

Imputation du réseau et/ou de la force de l'ordre :
Le principe de base est le même que pour des valeurs manquantes sur âge ou sexe. Nous distinguons plusieurs situations :

1) seul le type de réseau est manquant :
Il est imputé en fonction du type d'usager et de la force de l'ordre
1a) pour les communes sans réseau autoroutier :
La valeur à imputer est tirée au sort parmi quatre catégories de type de réseau (toutes sauf autoroutes) à partir d'une distribution multinomiale ? estimée par la distribution observée du type de réseau en fonction du type d'usager et de la force de l'ordre.
1b) pour les communes avec réseau autoroutier :
La valeur à imputer est tirée au sort parmi cinq catégories de type de réseau (AR, RN, RD, VC=voies communales, autres)
2) seul la force de l'ordre est manquante :
Cela correspond à commune manquante. La force de l'ordre est imputée en fonction du type de réseau et du type d'usager
3) type de réseau et force de l'ordre sont manquants :
Cette situation correspond à la variable commune manquante ou à des communes renseignées mais avec réseau autoroutier.
Le type de réseau est d'abord imputé en fonction du type d'usager.
Pour les communes renseignées (et avec réseau autoroutier), la force de l'ordre est déterminée en fonction du type de réseau.
Pour les communes non renseignées, la force de l'ordre est imputée en fonction du type de réseau et du type d'usager.

Adéquation des modèles de prédiction du NISS en trois catégories : NISS 1-3, NISS 4-8, NISS 9+

Tableau 22 : test d'adéquation du modèle logit P(NISS 4+) aux données

Déciles de P(NISS4+)	P(NISS 4+)	effectif	observés	prédits	Résidus de Pearson (carré)
1	0,12	2271	277	276,4	0,001
2	0,15	2274	331	337,7	0,156
3	0,18	2285	403	403,1	0,000
4	0,23	2272	522	521,1	0,002
5	0,30	2270	648	670,4	1,062
6	0,35	2273	816	794,0	0,937
7	0,40	2272	901	899,4	0,005
8	0,49	2271	1122	1120,9	0,002
9	0,78	2272	1766	1768,4	0,015
10	0,94	2245	2106	2100,7	0,208
ensemble		22705	8892	8892,1	2,388

statistique de Hosmer-Lemeshow=2,4 (p=0,97)

Le modèle de prédiction du nombre de blessés NISS 4+ semble en parfaite adéquation avec les données.

Tableau 23 : test d'adéquation du modèle logit P(NISS 9+) aux données

Déciles de P(NISS 9+)	P(NISS 9+)	effectif	observés	prédits	Résidus de Pearson (carré)
1	0,03	2306	52	65,0	2,68
2	0,04	2282	46	80,9	15,61
3	0,04	2270	84	97,6	1,98
4	0,06	2271	118	131,1	1,39
5	0,08	2273	189	184,2	0,14
6	0,11	2271	253	240,8	0,69
7	0,13	2273	333	300,9	3,95
8	0,18	2271	432	400,6	2,99
9	0,45	2271	1066	1018,4	4,03
10	0,75	2217	1618	1671,6	6,99
ensemble		22705	4191	4191,1	40,44

statistique de Hosmer-Lemeshow=40,4 (p<0,001)

Le test indique une mauvaise adéquation du modèle aux données. Cependant, la puissance associée à ce test est sans doute proche de 1 étant donné le large effectif (22705 observations), et donc le test semble très sensible aux écarts entre effectifs observés et prédits. La comparaison visuelle, ligne par ligne, des effectifs observés et prédits, par décile de P(NISS 9+), montre une adéquation tout à fait acceptable.

Autres critères de santé étudiés : modèles de la procédure d'estimation

Modèles de prédiction de la présence/absence du critère de santé chez les blessés

Les critères de santé autres que le NISS sont tous binaires : hospitalisés (oui/non), victimes avec blessure(s) IIS3+ : (oui/non), avec blessure(s) IIS 3+ à la tête: (oui/non), à la colonne : (oui/non), aux membres inférieurs : (oui/non). Chaque critère de santé est donc modélisé par une régression logistique (binomiale), et en fonction du critère de gravité des BAAC et des mêmes autres variables. Nous gardons en effet les mêmes variables, pour qu'il y ait cohérence entre les modèles. De plus, certains modèles ont peu de puissance pour tester les paramètres à zéro.

Cependant, pour les modèles de prédiction des blessés avec lésions IIS3+, (quelle que soit la région corporelle ou pour une région corporelle précise), les petits effectifs de blessés avec de telles lésions ne permettent pas de garder dans le modèle la variable âge en neuf catégories, mais seulement en cinq, ni l'interaction entre le type d'usager et la présence/ absence de tiers.

Traitement des valeurs manquantes et exclusion d'observations

Pour le critère hospitalisation (oui/non), 770 blessés (3,5%) ont une valeur manquante. Celle-ci a été remplacé par hospitalisation=oui. Ce choix s'est basé sur l'observation de l'évolution du taux d'hospitalisés à gravité lésionnelle constante (sur l'ensemble des victimes du Registre). Si nous supposions que les valeurs manquantes étaient des non-hospitalisations, l'évolution du taux d'hospitalisés (à gravité lésionnelle donnée) était à la hausse, ce qui n'est pas cohérent avec la politique actuelle d'hospitalisation; si nous supposions que les valeurs manquantes étaient des hospitalisations, alors le taux d'hospitalisés était stable (les valeurs manquantes ne sont pas uniformément réparties dans le temps mais plus fréquentes au début du fonctionnement du Registre).
La construction du modèle de prédiction porte donc sur 22 969 blessés, dont 8038 hospitalisés

Pour les victimes avec séquelles majeures prévues, nous excluons les observations où le MIIS (maximum IIS) est non renseigné, soit 197 (0,8%). La construction du modèle de prédiction porte donc sur 22 772 blessés, dont 468 avec une lésion IIS3+.

Blessures IIS3+ à la tête, à la colonne ou aux membres inférieurs : la non-mention d'une telle blessure est considérée comme l'absence de cette blessure. Les modèles sont construits sur 22771 blessés, dont 285 présentent une lésion IIS3+ à la tête, et 105 présentent une lésion IIS3+ aux membres inférieurs.
Pour la prédiction des lésions IIS3+ à la colonne, nous excluons les usagers « autre » (véhicule utilitaires, poids lourds, autres). En effet, il n'est observé aucun blessé dans cette catégorie avec une lésion IIS3+ sur la période 1996-2004 dans le département du Rhône et il ne nous paraît pas pertinent d'en déduire que leur probabilité d'une telle atteinte est nulle. L'étude porte donc sur 21 963 blessés, dont 74 avec une atteinte IIS3+ à la colonne vertébrale.

Critère de santé= hospitalisés

Tableau 24 : paramètres du modèle logit P(hospitalisé=oui),
construit sur 22 969 blessés de l'intersection BAAC-Registre, Rhône, 1996-2004

variable	modalité	beta	OR	IC à 95 %	
ordonnée à l'origine		-1,50	0,22	0,19	0,26
force ordre × gravité	gendarmerie, blessé léger	0,00			
	gendarmerie, blessé grave	2,25	9,47	8,08	11,11
	police, blessé léger	-0,84	0,43	0,38	0,49
	police, blessé grave	2,42	11,23	9,28	13,58
	CRS, blessé léger	-1,44	0,24	0,20	0,28
	CRS, blessé grave	1,99	7,29	4,85	10,96
sexe	féminin	0,00			
	masculin	0,29	1,33	1,24	1,43
âge	0-4 ans	1,66	5,25	4,11	6,70
	5-9 ans	0,94	2,55	2,12	3,07
	10-14 ans	0,55	1,73	1,46	2,04
	15-19 ans	0,03	1,03	0,92	1,15
	20-29 ans	-0,07	0,93	0,86	1,02
	30-54 ans	0,00			
	55-64 ans	0,21	1,24	1,07	1,43
	65-74 ans	0,45	1,57	1,33	1,86
	75 ans et plus	0,63	1,89	1,56	2,28
type d'usager × tiers	piétons, avec tiers	1,21	3,37	3,03	3,75
	cyclistes, avec tiers	0,86	2,37	1,98	2,84
	cyclistes, sans tiers	0,46	1,58	0,64	3,88
	usagers de 2RM, avec tiers	0,84	2,31	2,08	2,56
	usagers de 2RM, sans tiers	1,61	5,02	4,03	6,24
	automobilistes, avec tiers	0,00			
	automobilistes, sans tiers	0,94	2,57	2,30	2,86
	autres, avec tiers	-0,15	0,86	0,68	1,09
	autres, sans tiers	0,86	2,37	1,71	3,29
type d'antagoniste	piéton ou cycliste	-0,13	0,88	0,62	1,26
	deux-roues motorisé	-0,07	0,93	0,74	1,17
	voiture, autre ou aucun	0,00			
	véhicule utilitaire	0,20	1,22	1,04	1,44
	véhicule lourd	0,38	1,46	1,28	1,67
réseau	autoroute	0,51	1,66	1,35	2,05
	route nationale	0,30	1,35	1,16	1,57
	route départementale	0,21	1,23	1,11	1,37
	voie communale	0,00			
	autre	-0,20	0,82	0,66	1,01
agglomération	oui	0,00			
	non	0,35	1,42	1,26	1,61
année	(quantitative)	0,01	1,01	1,00	1,03

N=22 969 blessés, nombre de blessés hospitalisés=8038, coefficient D de Somer=

Toutes tes les variables avec p< 0,0001, sauf année : p=0.108

Tableau 25 : test d'adéquation du modèle logit P(hospitalisés) aux données

Déciles de P(hospitalisé)	P(hospitalisé)	effectif	observés	prédits	Résidus de Pearson (carré)
1	0,09	2298	206	202,7	0,058
2	0,11	2307	208	253,7	9,249
3	0,13	2297	297	302,7	0,122
4	0,19	2291	428	428,0	0,000
5	0,23	2290	541	537,8	0,026
6	0,27	2299	634	618,6	0,523
7	0,33	2297	790	765,0	1,227
8	0,46	2297	1092	1055,4	2,355
9	0,76	2298	1746	1753,3	0,126
10	0,92	2295	2096	2121,0	3,877
ensemble		22969	8038	8038,0	17,563

statistique de Hosmer-Lemeshow = 17,6 (p=0,025)

Le test de Hosmer-Lemeshow indique une mauvaise adéquation du modèle aux données. Cependant le test semble très sensible (puissance vraisemblablement très élevée) et une comparaison attentive des effectifs observés et prédits, par décile de P(hospitalisé) montre une adéquation tout à fait acceptable.

Critère de santé= au moins une lésion IIS3+

Tableau 26 : paramètres du modèle de prédiction P(au moins une lésion IIS3+), construit sur 22 271 blessés de l'intersection BAAC-Registre, Rhône, 1996-2004

variable	modalité	beta	OR	IC à 95 %	
ordonnée à l'origine		-6,42	0,00	0,00	0,00
force ordre × gravité	gendarmerie, blessé léger	0,00			
	gendarmerie, blessé grave	2,73	15,31	9,68	24,21
	police, blessé léger	0,10	1,10	0,64	1,90
	police, blessé grave	2,62	13,69	8,04	23,29
	CRS, blessé léger	-0,86	0,42	0,20	0,91
	CRS, blessé grave	2,70	14,87	7,54	29,32
sexe	féminin	0,00			
	masculin	0,46	1,58	1,25	2,00
âge	0-14 ans	-0,03	0,97	0,66	1,43
	15-19	-0,01	0,99	0,73	1,33
	20-29	-0,06	0,94	0,74	1,20
	30-64	0,00			
	65 +	0,01	1,01	0,70	1,44
type d'usager	piéton	0,84	2,33	1,71	3,17
	cycliste	0,40	1,49	0,90	2,47
	usager de 2RM	0,32	1,37	1,06	1,78
	automobiliste	0,00			
	autre	-0,13	0,87	0,46	1,67
type d'antagoniste	piéton ou cycliste	0,63	1,88	0,67	5,31
	deux-roues motorisé	0,36	1,43	0,82	2,49
	voiture, autre	0,00			
	véhicule utilitaire	0,15	1,17	0,70	1,95
	véhicule lourd	0,95	2,58	1,88	3,55
	aucun	0,84	2,31	1,80	2,97
réseau	autoroute	0,13	1,13	0,62	2,07
	route nationale	0,03	1,04	0,69	1,55
	route départementale	0,03	1,03	0,76	1,39
	voie communale	0,00			
	autre	-0,06	0,94	0,53	1,67
agglomération	oui	0,00			
	non	0,43	1,53	1,15	2,04
année	(quantitative)	0,11	1,11	1,07	1,15

N=22271, nombre de blessés avec au moins une lésion ISS 3+=468, coefficient D de Somer=85,8

Tableau 27 : test d'adéquation du modèle logit P(lésion IIS3+) aux données

Déciles de P(IIS 3+)	P(IIS 3+)	effectif	observés	prédits	Résidus de Pearson (carré)
1 et 2	0,002	4555	7	10,14	0,97
3	0,003	2277	7	7,62	0,05
4	0,004	2227	4	9,09	2,86
5	0,005	2278	9	11,16	0,42
6	0,006	2285	12	13,63	0,20
7	0,007	2276	19	17,01	0,23
8	0,010	2280	32	23,71	2,93
9	0,039	2277	89	88,63	0,00
10	0,124	2317	289	287,01	0,02
ensemble		22772	468	468,00	7,7

statistique de Hosmer-Lemeshow = 7,7 (p≈0,4)

Les déciles 1 et 2 ont du être regroupés afin que les effectifs prédits soient supérieurs à 5 (condition d'application du test)

Critère de santé= au moins une lésion IIS3+, à la tête :

Tableau 28 : paramètres du modèle de prédiction P(au moins une lésion IIS3+ à la tête), construit sur 22 271 blessés de l'intersection BAAC-Registre, Rhône, 1996-2004

variable	modalité	beta	OR	IC à 95 %	
ordonnée à l'origine		-6,79	0,00	0,00	0,00
force ordre × gravité	gendarmerie, blessé léger	0,00	1,00		
	gendarmerie, blessé grave	2,83	16,88	9,26	30,78
	police, blessé léger	-0,28	0,75	0,36	1,56
	police, blessé grave	2,70	14,84	7,52	29,30
	CRS, blessé léger	-0,79	0,45	0,15	1,36
	CRS, blessé grave	3,01	20,27	8,22	50,01
sexe	féminin	0,00	1,00		
	masculin	0,40	1,49	1,12	1,99
âge	0-14 ans	0,26	1,29	0,82	2,03
	15-19	0,23	1,26	0,88	1,83
	20-29	-0,03	0,97	0,71	1,34
	30-64	0,00	1,00		
	65 +	0,16	1,17	0,76	1,79
type d'usager	piéton	0,81	2,25	1,54	3,30
	cycliste	0,55	1,74	0,98	3,08
	usager de 2RM	-0,05	0,95	0,67	1,34
	automobiliste	0,00			
	autre	0,22	1,24	0,59	2,64
type d'antagoniste	piéton ou cycliste	0,94	2,56	0,77	8,54
	deux-roues motorisé	0,60	1,83	0,98	3,42
	voiture, autre	0,00	1,00		
	véhicule utilitaire	0,21	1,23	0,66	2,32
	véhicule lourd	0,62	1,86	1,21	2,86
	aucun	0,84	2,31	1,68	3,17
réseau	autoroute	-0,25	0,78	0,35	1,74
	route nationale	-0,32	0,73	0,42	1,26
	route départementale	0,04	1,04	0,72	1,52
	voie communale	0,00	1,00		
	autre	0,04	1,04	0,54	2,03
agglomération	oui	0,00	1,00		
	non	0,31	1,36	0,96	1,95
année	(quantitative)	0,09	1,10	1,05	1,15

N=22271, nombre de blessés avec au moins une lésion ISS 3+ à la tête=285, coefficient D de Somer=86,5

Tableau 29 : test d'adéquation du modèle logit P(lésion IIS 3+ à la tête) aux données

Déciles de P(IIS 3+ tête)	P(IIS 3+ tête)	effectif	observés	prédits	résidus de Pearson (carré)
1, 2 et 3	0,001	6790	6	8,15	0,57
4 et 5	0,002	4585	7	9,2	0,53
6	0,003	2281	6	6,16	0,00
7	0,004	2291	8	8,02	0,00
8	0,005	2283	15	11,62	0,99
9	0,025	2274	56	57,47	0,04
10	0,081	2268	187	184,39	0,04
ensemble		22772	285	285,01	2,17

statistique de Hosmer-Lemeshow = 2,17 (p≈0,8)

Critère de santé= au moins une lésion IIS3+, aux membres inférieurs

Tableau 30 : paramètres du modèle de prédiction P(au moins une lésion IIS3+ aux membres inférieurs), construit sur 21 963 blessés de l'intersection BAAC-Registre, Rhône, 1996-2004

variable	modalité	beta	OR	IC à 95 %	
ordonnée à l'origine		-7,90	0,00	0,00	0,00
force ordre × gravité	gendarmerie, blessé léger	0,00	1,00		
	gendarmerie, blessé grave	1,85	6,35	2,77	14,59
	police, blessé léger	0,62	1,85	0,68	5,04
	police, blessé grave	2,34	10,35	3,70	28,93
	CRS, blessé léger	-0,55	0,58	0,17	2,00
	CRS, blessé grave	2,02	7,57	2,19	26,22
sexe	féminin	0,00	1,00		
	masculin	0,55	1,73	1,07	2,81
âge	0-14 ans	-1,14	0,32	0,10	1,06
	15-19	-0,42	0,66	0,33	1,30
	20-29	0,14	1,15	0,74	1,80
	30-64	0,00	1,00		
	65 +	-0,23	0,79	0,36	1,76
type d'usager	piéton	0,81	2,24	1,21	4,17
	cycliste	-0,97	0,38	0,05	2,81
	usager de 2RM	0,66	1,94	1,18	3,18
	automobiliste	0,00	1,00		
	autre	-0,57	0,57	0,13	2,39
type d'antagoniste	piéton, cycliste, ou 2RM	-1,19	0,31	0,04	2,23
	voiture, autre	0,00	1,00		
	véhicule utilitaire	0,24	1,27	0,50	3,19
	véhicule lourd	1,23	3,43	2,01	5,88
	aucun	0,06	1,07	0,60	1,88
réseau	autoroute	0,33	1,40	0,45	4,36
	route nationale	0,54	1,71	0,78	3,76
	route départementale	0,29	1,34	0,72	2,49
	voie communale	0,00	1,00		
	autre	-0,97	0,38	0,05	2,83
agglomération	oui	0,00	1,00		
	non	0,71	2,03	1,07	3,86
année	(quantitative)	0,12	1,13	1,04	1,22

N=22271, nombre de blessés avec au moins une lésion IIS 3+ aux membres inférieurs=105, D de Somer=76,8

Tableau 31 : test d'adéquation du modèle logit P(lésion IIS 3+ aux membres inférieurs) aux données

Déciles de P(IIS 3+)	P(IIS 3+ memb. inf.)	effectif	observés	prédits	Résidus de Pearson (carré)
1 à 3	0,001	6879	1	5,58	3,76
4 et 5	0,002	4548	7	7,22	0,01
6	0,002	2276	7	5,11	0,70
7	0,003	2283	8	6,71	0,25
8	0,004	2275	10	9,5	0,03
9	0,007	2278	18	16,92	0,07
10	0,024	2233	54	53,94	0,00
ensemble		22772	105	104,98	4,81

statistique de Hosmer-Lemeshow =4,81 (p≈0,45)

Critère de santé= au moins une lésion IIS3+, à la colonne vertébrale

Tableau 32 : paramètres du modèle de prédiction P(au moins une lésion IIS3+ à la colonne vertébrale), construit sur 22 272 blessés de l'intersection BAAC-Registre, Rhône, 1996-2004

variable	modalité	beta	OR	IC à 95 %	
ordonnée à l'origine		-8,65	0,00	0,00	0,00
force ordre × gravité	gendarmerie, blessé léger	0,00	1,00		
	gendarmerie, blessé grave	3,26	26,01	6,24	108,4
	police, blessé léger	0,05	1,05	0,20	5,56
	police, blessé grave	2,41	11,12	2,22	55,65
	CRS, blessé léger	-1,50	0,22	0,02	2,76
	CRS, blessé grave	3,18	24,03	4,25	136,0
sexe	féminin	0,00	1,00		
	masculin	0,66	1,94	1,01	3,72
âge	0-14 ans	-0,34	0,71	0,23	2,15
	15-19	-0,30	0,74	0,36	1,51
	20-29	-0,21	0,81	0,46	1,42
	30-64	0,00	1,00		
	65 +	-1,15	0,32	0,07	1,38
type d'usager	piéton	0,81	2,25	0,93	5,43
	cycliste	0,53	1,70	0,48	5,98
	usager de 2RM	0,82	2,28	1,30	3,99
	automobiliste	0,00	1,00		
	autre	exclus			
type d'antagoniste	piéton, cycliste, ou 2RM	0,54	1,71	0,51	5,77
	voiture, autre	0,00	1,00		
	véhicule utilitaire	-0,61	0,54	0,07	4,01
	véhicule lourd	0,77	2,16	0,92	5,07
	aucun	1,15	3,15	1,81	5,49
réseau	autoroute	0,42	1,52	0,42	5,55
	route nationale	0,13	1,14	0,46	2,83
	route départementale	-0,32	0,73	0,34	1,55
	voie communale	0,00	1,00		
	autre	0,13	1,14	0,31	4,12
agglomération	oui	0,00	1,00		
	non	0,39	1,47	0,78	2,77
année	(quantitative)	0,08	1,08	0,99	1,19

N=21963, nombre de blessés avec au moins une lésion ISS 3+ à la colonne vertébrale=74, D de Somer=81,5

Tableau 33 : test d'adéquation du modèle logit P(lésion IIS 3+ à la colonne vertébrale) aux données

Déciles de P(IIS 3+ col.)	P(IIS 3+ col.)	effectif	observés	prédits	Résidus de Pearson (carré)
1 à 8	0,000	15353	7	6,75	0,01
9	0,004	2197	12	8,88	1,10
10	0,026	2198	54	56,6	0,12
ensemble		21963	74	73,99	1,23

statistique de Hosmer-Lemeshow = 1,23 (p≈0,25)

Modèles de capture-recapture

Quel que soit le critère de santé étudié, les modèles de capture-recapture incluent les mêmes variables : gravité mesurée par le critère de santé en question, interaction type d'usager et tiers, type de réseau, force de l'ordre, et année (sous forme quantitative)

Traitement des valeurs manquantes et exclusion d'observations :
Le traitement des observations avec valeurs manquantes est en cohérence avec le traitement des valeurs manquantes pour la construction des modèles de prédiction du critère de santé étudié.
Pour le critère hospitalisation (oui/non), les valeurs manquantes ont été remplacées par hospitalisation=oui.
Pour les victimes avec séquelles majeures prévues, les observations où le MIIS (maximum IIS) est non renseigné, soit 197 (0,8%) ont été considérées comme ne présentant pas ces blessures.
Blessures IIS3+ à la tête, à la colonne ou aux membres inférieurs : la non-mention d'une telle blessure est considérée comme l'absence de cette blessure.
Pour la prédiction des lésions IIS3+ à la colonne, les observations avec type d'usagers= autre (véhicule utilitaires, poids lourds, autres) ont été exclues du modèle précédent de prédiction. Elles sont aussi exclues du modèle de capture-recapture, et donc les estimations nationales ne sont pas disponibles pour ce type d'usagers.

Critère de santé= hospitalisés

Tableau 34 : paramètres du modèle de capture-recapture (critère de santé= hospitalisés), construit sur environ 90 000 blessés, BAAC et/ou Registre, Rhône, 1996-2004

variable	source	beta	OR		IC à 95 %
ordonnée à l'origine	Registre seulement	3,21	24,81	22,82	26,98
ordonnée à l'origine	intersection	1,82	6,17	5,65	6,73
hospitalisé (p<0,0001)					
non	Registre seulement	0,00	1,00		
oui	Registre seulement	-1,04	0,35	0,33	0,37
non	intersection	0,00	1,00		
oui	intersection	0,26	1,30	1,23	1,37
type d'usager x tiers (p<0,0001)					
piétons, avec tiers	Registre seulement	-0,09	0,92	0,85	0,99
cyclistes, avec tiers	Registre seulement	0,49	1,64	1,43	1,87
cyclistes, sans tiers	Registre seulement	4,72	112,13	68,50	183,55
usagers de 2RM, avec tiers	Registre seulement	-0,29	0,75	0,70	0,81
usagers de 2RM, sans tiers	Registre seulement	2,19	8,90	7,67	10,32
automobilistes, avec tiers	Registre seulement	0,00	1,00		
automobilistes, sans tiers	Registre seulement	0,56	1,76	1,64	1,89
autres, avec tiers	Registre seulement	-0,81	0,45	0,39	0,52
autres, sans tiers	Registre seulement	1,64	5,16	4,22	6,31
piétons, avec tiers	intersection	0,01	1,01	0,94	1,09
cyclistes, avec tiers	intersection	0,11	1,12	0,97	1,29
cyclistes, sans tiers	intersection	0,76	2,15	1,25	3,67
usagers de 2RM, avec tiers	intersection	0,01	1,01	0,94	1,09
usagers de 2RM, sans tiers	intersection	0,53	1,69	1,44	1,99
automobilistes, avec tiers	intersection	0,00	1,00		
automobilistes, sans tiers	intersection	-0,01	0,99	0,92	1,07
autres, avec tiers	intersection	-0,42	0,66	0,57	0,76
autres, sans tiers	intersection	0,28	1,32	1,06	1,65
type de réseau (p<0,0001)					
autoroutes	Registre seulement	-0,93	0,39	0,34	0,46
routes nationales	Registre seulement	-0,78	0,46	0,41	0,51
routes départementales	Registre seulement	-2,23	0,11	0,10	0,12
voies communales	Registre seulement	0,00	1,00		
autre	Registre seulement	-0,76	0,47	0,41	0,54
autoroutes	intersection	-0,44	0,64	0,55	0,75
routes nationales	intersection	-0,57	0,56	0,51	0,63
routes départementales	intersection	-1,89	0,15	0,14	0,17
voies communales	intersection	0,00	1,00		
autre	intersection	-1,85	0,16	0,13	0,19
force de l'ordre (p<0,0001)					
gendarmerie	Registre seulement	0,00	1,00		
CRS	Registre seulement	-1,30	0,27	0,24	0,31
police	Registre seulement	-1,69	0,19	0,17	0,20
gendarmerie	intersection	0,00	1,00		
CRS	intersection	-0,43	0,65	0,57	0,75
police	intersection	-0,75	0,47	0,44	0,51
année (p<0,0001)					
quantitative	Registre seulement	0,02	1,02	1,01	1,03
quantitative	intersection	-0,02	0,98	0,97	0,99

Critère de santé= au moins une lésion IIS3+

Tableau 35 : paramètres du modèle de capture-recapture (critère de santé= lésion IIS 3+), construit sur environ 90 000 blessés, BAAC et/ou Registre, Rhône, 1996-2004

variable	source	beta	OR		IC à 95 %
ordonnée à l'origine	Registre seulement	2,93	18,81	17,39	20,35
ordonnée à l'origine	intersection	1,84	6,30	5,80	6,84
une lésion IIS3+					
non	Registre seulement	0,00	1,00		
oui	Registre seulement	1,03	2,81	1,77	4,44
non	intersection	0,00	1,00		
oui	intersection	2,33	10,30	6,56	16,18
type d'usager x tiers					
piétons, avec tiers	Registre seulement	-0,39	0,68	0,63	0,73
cyclistes, avec tiers	Registre seulement	0,15	1,16	1,02	1,31
cyclistes, sans tiers	Registre seulement	4,61	100,76	60,63	167,44
usagers de 2RM, avec tiers	Registre seulement	-0,43	0,65	0,61	0,70
usagers de 2RM, sans tiers	Registre seulement	1,88	6,54	5,71	7,50
automobilistes, avec tiers	Registre seulement	0,00	1,00		
automobilistes, sans tiers	Registre seulement	0,43	1,53	1,43	1,64
autres, avec tiers	Registre seulement	-0,90	0,41	0,35	0,47
autres, sans tiers	Registre seulement	1,39	4,00	3,31	4,84
piétons, avec tiers	intersection	0,03	1,03	0,96	1,11
cyclistes, avec tiers	intersection	-0,01	0,99	0,86	1,13
cyclistes, sans tiers	intersection	0,90	2,46	1,41	4,26
usagers de 2RM, avec tiers	intersection	0,05	1,05	0,98	1,13
usagers de 2RM, sans tiers	intersection	0,43	1,53	1,32	1,78
automobilistes, avec tiers	intersection	0,00	1,00		
automobilistes, sans tiers	intersection	0,02	1,02	0,94	1,09
autres, avec tiers	intersection	-0,52	0,59	0,52	0,69
autres, sans tiers	intersection	0,24	1,27	1,03	1,57
type de réseau					
autoroutes	Registre seulement	-1,23	0,29	0,25	0,34
routes nationales	Registre seulement	-0,99	0,37	0,34	0,41
routes départementales	Registre seulement	-2,31	0,10	0,09	0,11
voies communales	Registre seulement	0,00	1,00		
autre	Registre seulement	-0,94	0,39	0,34	0,44
autoroutes	intersection	-0,58	0,56	0,48	0,65
routes nationales	intersection	-0,69	0,50	0,46	0,56
routes départementales	intersection	-1,87	0,15	0,14	0,17
voies communales	intersection	0,00	1,00		
autre	intersection	-2,05	0,13	0,11	0,15
force de l'ordre					
gendarmerie	Registre seulement	0,00	1,00		
CRS	Registre seulement	-0,89	0,41	0,36	0,47
police	Registre seulement	-1,49	0,23	0,21	0,24
gendarmerie	intersection	0,00	1,00		
CRS	intersection	-0,27	0,76	0,67	0,87
police	intersection	-0,70	0,50	0,46	0,53
année					
quantitative	Registre seulement	0,02	1,02	1,01	1,02
quantitative	intersection	-0,02	0,98	0,97	0,99

Critère de santé= au moins une lésion IIS3+ à la tête

Tableau 36 : paramètres du modèle de capture-recapture (critère de santé : lésion IIS 3+ à la tête), construit sur environ 90 000 blessés, BAAC et/ou Registre, Rhône, 1996-2004

variable	source	beta	OR		IC à 95 %
ordonnée à l'origine	Registre seulement	2,93	18,68	17,28	20,20
ordonnée à l'origine	intersection	1,83	6,25	5,76	6,79
lésion IIS 3+ tête					
non	Registre seulement	0,00	1,00		
oui	Registre seulement	2,11	8,21	2,60	25,90
non	intersection	0,00	1,00		
oui	intersection	3,70	40,55	12,97	126,73
type d'usager x tiers					
piétons, avec tiers	Registre seulement	-0,40	0,67	0,63	0,72
cyclistes, avec tiers	Registre seulement	0,14	1,15	1,02	1,30
cyclistes, sans tiers	Registre seulement	4,61	100,81	60,66	167,52
usagers de 2RM, avec tiers	Registre seulement	-0,43	0,65	0,61	0,70
usagers de 2RM, sans tiers	Registre seulement	1,86	6,44	5,62	7,37
automobilistes, avec tiers	Registre seulement	0,00	1,00		
automobilistes, sans tiers	Registre seulement	0,42	1,53	1,43	1,63
autres, avec tiers	Registre seulement	-0,90	0,41	0,35	0,47
autres, sans tiers	Registre seulement	1,38	3,96	3,28	4,77
piétons, avec tiers	intersection	0,03	1,03	0,96	1,11
cyclistes, avec tiers	intersection	-0,02	0,98	0,86	1,13
cyclistes, sans tiers	intersection	0,90	2,46	1,42	4,26
usagers de 2RM, avec tiers	intersection	0,05	1,05	0,98	1,13
usagers de 2RM, sans tiers	intersection	0,42	1,53	1,32	1,78
automobilistes, avec tiers	intersection	0,00	1,00		
automobilistes, sans tiers	intersection	0,02	1,02	0,95	1,09
autres, avec tiers	intersection	-0,52	0,59	0,52	0,69
autres, sans tiers	intersection	0,24	1,27	1,03	1,57
type de réseau					
autoroutes	Registre seulement	-1,26	0,28	0,25	0,33
routes nationales	Registre seulement	-1,00	0,37	0,34	0,41
routes départementales	Registre seulement	-2,30	0,10	0,09	0,11
voies communales	Registre seulement	0,00	1,00		
autre	Registre seulement	-0,93	0,39	0,35	0,45
autoroutes	intersection	-0,59	0,55	0,48	0,64
routes nationales	intersection	-0,69	0,50	0,45	0,55
routes départementales	intersection	-1,86	0,16	0,14	0,17
voies communales	intersection	0,00	1,00		
autre	intersection	-2,04	0,13	0,11	0,16
force de l'ordre					
gendarmerie	Registre seulement	0,00	1,00		
CRS	Registre seulement	-0,87	0,42	0,37	0,48
police	Registre seulement	-1,48	0,23	0,21	0,24
gendarmerie	intersection	0,00	1,00		
CRS	intersection	-0,70	0,50	0,46	0,53
police	intersection	-0,26	0,77	0,68	0,88
année					
quantitative	Registre seulement	0,02	1,02	1,01	1,02
quantitative	intersection	-0,02	0,98	0,97	0,99

Critère de santé= au moins une lésion IIS3+ aux membres inférieurs

Tableau 37 : paramètres du modèle de capture-recapture (critère de santé : lésion IIS3+ aux membres inférieurs), construit sur environ 90 000 blessés, BAAC et/ou Registre, Rhône, 1996-2004

variable	source	beta	OR		IC à 95 %
ordonnée à l'origine	Registre seulement	2,92	18,59	17,19	20,09
ordonnée à l'origine	intersection	1,84	6,27	5,78	6,81
lésion IIS 3+ membres inférieurs					
non	Registre seulement	0,00	1,00		
oui	Registre seulement	10,83	50477,9	0,00	5,27E42
non	intersection	0,00	1,00		
oui	intersection	11,46	94927,1	0,00	9,91E42
type d'usager x tiers					
piétons, avec tiers	Registre seulement	-0,41	0,67	0,62	0,71
cyclistes, avec tiers	Registre seulement	0,13	1,14	1,01	1,29
cyclistes, sans tiers	Registre seulement	4,56	95,38	58,26	156,13
usagers de 2RM, avec tiers	Registre seulement	-0,43	0,65	0,61	0,70
usagers de 2RM, sans tiers	Registre seulement	1,85	6,37	5,57	7,28
automobilistes, avec tiers	Registre seulement	0,00	1,00		
automobilistes, sans tiers	Registre seulement	0,42	1,52	1,42	1,62
autres, avec tiers	Registre seulement	-0,90	0,41	0,35	0,47
autres, sans tiers	Registre seulement	1,35	3,84	3,19	4,62
piétons, avec tiers	intersection	0,04	1,04	0,97	1,12
cyclistes, avec tiers	intersection	-0,02	0,98	0,86	1,12
cyclistes, sans tiers	intersection	0,85	2,34	1,36	4,00
usagers de 2RM, avec tiers	intersection	0,05	1,05	0,98	1,13
usagers de 2RM, sans tiers	intersection	0,42	1,52	1,31	1,77
automobilistes, avec tiers	intersection	0,00	1,00		
automobilistes, sans tiers	intersection	0,02	1,02	0,95	1,10
autres, avec tiers	intersection	-0,52	0,60	0,52	0,69
autres, sans tiers	intersection	0,21	1,23	1,00	1,52
type de réseau					
autoroutes	Registre seulement	-1,26	0,28	0,25	0,33
routes nationales	Registre seulement	-1,00	0,37	0,33	0,40
routes départementales	Registre seulement	-2,30	0,10	0,09	0,11
voies communales	Registre seulement	0,00	1,00		
autre	Registre seulement	-0,94	0,39	0,34	0,44
autoroutes	intersection	-0,60	0,55	0,47	0,64
routes nationales	intersection	-0,69	0,50	0,45	0,55
routes départementales	intersection	-1,86	0,16	0,14	0,17
voies communales	intersection	0,00	1,00		
autre	intersection	-2,06	0,13	0,11	0,15
force de l'ordre					
gendarmerie	Registre seulement	0,00	1,00		
CRS	Registre seulement	-0,85	0,43	0,38	0,48
police	Registre seulement	-1,48	0,23	0,21	0,24
gendarmerie	intersection	0,00	1,00		
CRS	intersection	-0,25	0,78	0,68	0,89
police	intersection	-0,70	0,50	0,46	0,53
année					
quantitative	Registre seulement	0,01	1,01	1,01	1,02
quantitative	intersection	-0,02	0,98	0,97	0,99

Critère de santé= au moins une lésion IIS3+ à la colonne vertébrale

Tableau 38 : paramètres du modèle de capture-recapture (critère de santé : lésion IIS3+ à la colonne vertébrale), construit sur environ 90 000 blessés, BAAC et/ou Registre, Rhône, 1996-2004

variable	source	beta	OR	IC à 95 %	
ordonnée à l'origine	Registre seulement	2,96	19,24	17,76	20,84
ordonnée à l'origine	intersection	1,87	6,46	5,94	7,03
lésion IIS 3+ colonne vertébrale					
non	Registre seulement	0,00	1,00		
oui	Registre seulement	9,36	11609,6	0,00	3,83E43
non	intersection	0,00	1,00		
oui	intersection	11,08	65068,4	0,00	2,15E44
type d'usager x tiers					
piétons, avec tiers	Registre seulement	-0,41	0,67	0,62	0,72
cyclistes, avec tiers	Registre seulement	0,14	1,15	1,01	1,30
cyclistes, sans tiers	Registre seulement	4,55	94,71	57,85	155,06
usagers de 2RM, avec tiers	Registre seulement	-0,43	0,65	0,61	0,70
usagers de 2RM, sans tiers	Registre seulement	1,85	6,38	5,58	7,31
automobilistes, avec tiers	Registre seulement	0,00	1,00		
automobilistes, sans tiers	Registre seulement	0,43	1,53	1,43	1,64
autres, avec tiers	Registre seulement	exclus			
autres, sans tiers	Registre seulement	exclus			
piétons, avec tiers	intersection	0,05	1,05	0,98	1,13
cyclistes, avec tiers	intersection	0,42	1,52	1,31	1,76
cyclistes, sans tiers	intersection	-0,02	0,98	0,86	1,13
usagers de 2RM, avec tiers	intersection	0,84	2,31	1,35	3,95
usagers de 2RM, sans tiers	intersection	0,00	1,00		
automobilistes, avec tiers	intersection	0,04	1,04	0,97	1,12
automobilistes, sans tiers	intersection	0,03	1,03	0,96	1,11
autres, avec tiers	intersection	exclus			
autres, sans tiers	intersection	exclus			
type de réseau					
autoroutes	Registre seulement	-1,25	0,29	0,25	0,33
routes nationales	Registre seulement	-0,99	0,37	0,34	0,41
routes départementales	Registre seulement	-2,31	0,10	0,09	0,11
voies communales	Registre seulement	0,00	1,00		
autre	Registre seulement	-0,97	0,38	0,33	0,43
autoroutes	intersection	-0,62	0,54	0,46	0,63
routes nationales	intersection	-0,70	0,50	0,45	0,55
routes départementales	intersection	-1,87	0,15	0,14	0,17
voies communales	intersection	0,00	1,00		
autre	intersection	-2,12	0,12	0,10	0,14
force de l'ordre					
gendarmerie	Registre seulement	0,00	1,00		
CRS	Registre seulement	-0,87	0,42	0,37	0,48
police	Registre seulement	-1,49	0,23	0,21	0,24
gendarmerie	intersection	0,00	1,00		
CRS	intersection	-0,26	0,77	0,67	0,88
police	intersection	-0,70	0,50	0,46	0,53
année					
quantitative	Registre seulement	0,01	1,01	1,00	1,02
quantitative	intersection	-0,02	0,98	0,97	0,98

Détail du calcul sur la région Aquitaine de nombre de traumatisés crâniens graves d'après la littérature

D'après des résultats publiés (Masson, Thicoipe et al. 1998; Masson, Thicoipe et al. 2001)

Nombre de traumatisés crâniens graves en 1996, région Aquitaine, avec une lésion AIS 4 ou 5 à la tête, toutes causes:
497 résidents de la région
145 non-résidents (les garder car notre critère géographique n'est pas le lieu d'habitation mais le lieu d'accident)
642, au total

Ne garder que les accidents de la route :
240 parmi les résidents, car concerne 48,3% des résidents
83 parmi les non-résidents, car concerne 57,2% des non-résidents
323, au total :

Enlever les décédés à 6 jours : pas disponible, mais :
- 18% de décédés à 6 heures, à enlever,
- les décédés immédiats sont déjà exclus,
- les victimes MAIS 6 sont aussi déjà exclues,
265, effectif restant

Ce sont des blessés avec lésions de gravité AIS 4-5, donc il faudrait estimer combien ont une lésion IIS3+ ; d'après le Registre du Rhône : 67,7% des MAIS 4-5 ont des lésions IIS 3+, mais d'après l'article de Masson, le MAIS est sous-estimé car il provient d'une conversion du ICD 9. Nous avons décidé de nous en tenir là, c'est-à-dire à l'estimation d'un effectif de 265.

Références :

Masson, F., M. Thicoipe, et al. (1998). "Epidémiologie des traumatisés crâniens graves (Aquitaine 1996) [résumé]." Annales françaises d'anesthésie et de réanimation **17**: 921 (R218).

Masson, F., M. Thicoipe, et al. (2001). "Epidemiology of Severe Brain Injuries: A Prospective Population-Based Study." The Journal of Trauma, Injury, Infection and Critical Care **51**(3): 481-489.

Résultats supplémentaires du bilan France métropolitaine

Blessés toutes gravités, par type d'usager

Blessés toutes gravités, piétons

Tableau 39 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, piétons

année	effectif	incidence	évolution
1996	19 789	34,1	
1997	19 152	32,9	-3,5 %
1998	18 493	31,7	-3,8 %
1999	18 636	31,8	0,3 %
2000	18 398	31,2	-1,9 %
2001	17 458	29,4	-5,7 %
2002	16 072	26,9	-8,5 %
2003	14 068	23,4	-13,0 %
2004	13 841	22,9	-2,2 %
moyenne	17 323	29,4	-4,8 %

Tableau 40 : Effectif et incidence d'après la projection, blessés toutes gravités, piétons

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	44 289	42 244	45 871	76,3	72,8	79,1	
1997	43 023	41 128	44 391	73,9	70,7	76,3	-3,2 %
1998	42 164	40 356	43 449	72,2	69,1	74,4	-2,3 %
1999	43 020	41 230	44 252	73,3	70,3	75,4	1,6 %
2000	43 231	41 431	44 415	73,3	70,2	75,3	-0,1 %
2001	41 942	40 131	43 131	70,6	67,6	72,6	-3,6 %
2002	39 376	37 601	40 569	65,9	62,9	67,9	-6,7 %
2003	35 500	33 800	36 672	59,0	56,2	61,0	-10,4 %
2004	35 438	33 648	36 655	58,6	55,6	60,6	-0,8 %
moyenne	40 887	39 063	42 156	69,2	66,2	71,4	-3,2 %

Blessés toutes gravités cyclistes

Tableau 41 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, cyclistes

année	effectif	incidence	évolution
1996	6 771	11,7	
1997	7 190	12,4	5,9 %
1998	6 366	10,9	-11,7 %
1999	6 260	10,7	-2,1 %
2000	5 927	10,0	-5,9 %
2001	5 259	8,9	-11,8 %
2002	4 571	7,6	-13,6 %
2003	4 806	8,0	4,5 %
2004	4 427	7,3	-8,4 %
moyenne	5 731	9,7	-5,4 %

Tableau 42 : Effectif et incidence d'après la projection, blessés toutes gravités, cyclistes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	66 000	53 848	79 588	113,7	92,8	137,2	
1997	68 890	56 406	82 690	118,4	96,9	142,1	4,1 %
1998	63 614	52 034	76 345	108,9	89,1	130,7	-8,0 %
1999	59 460	49 184	70 596	101,4	83,8	120,3	-6,9 %
2000	54 179	45 249	63 843	91,8	76,7	108,2	-9,4 %
2001	46 551	39 023	54 585	78,4	65,7	91,9	-14,6 %
2002	48 621	39 862	58 138	81,3	66,7	97,3	3,8 %
2003	50 465	41 648	59 912	83,9	69,2	99,6	3,1 %
2004	42 678	35 648	50 205	70,5	58,9	83,0	-15,9 %
moyenne	55 606	45 878	66 211	94,3	77,8	112,3	-5,5 %

Blessés toutes gravités, usagers de deux-roues motorisés

Tableau 43 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, usagers de 2RM

année	effectif	incidence	évolution
1996	36 512	62,9	
1997	38 592	66,3	5,4 %
1998	38 059	65,2	-1,7 %
1999	39 232	66,9	2,6 %
2000	39 267	66,5	-0,5 %
2001	38 041	64,0	-3,7 %
2002	34 330	57,4	-10,3 %
2003	31 847	52,9	-7,8 %
2004	31 500	52,0	-1,7 %
moyenne	36 376	61,6	-2,2 %

Tableau 44 : Effectif et incidence d'après la projection, blessés toutes gravités, usagers de 2RM

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	113 969	107 259	118 788	196,4	184,8	204,7	
1997	120 849	114 155	125 378	207,6	196,1	215,4	5,7 %
1998	122 775	115 910	127 187	210,2	198,5	217,8	1,3 %
1999	126 356	119 644	130 549	215,4	204,0	222,5	2,5 %
2000	128 361	121 789	132 319	217,5	206,4	224,2	1,0 %
2001	125 429	118 934	129 114	211,2	200,2	217,4	-2,9 %
2002	116 134	109 890	119 475	194,3	183,8	199,9	-8,0 %
2003	114 004	107 292	117 513	189,5	178,4	195,4	-2,4 %
2004	114 586	107 581	118 283	189,3	177,8	195,4	-0,1 %
moyenne	120 274	113 606	124 290	203,5	192,2	210,3	-0,4 %

Blessés toutes gravités, automobilistes

Tableau 45 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, automobilistes

année	effectif	incidence	évolution
1996	100 570	173,3	
1997	98 552	169,3	-2,3 %
1998	99 341	170,1	0,5 %
1999	97 245	165,8	-2,5 %
2000	93 236	158,0	-4,7 %
2001	88 272	148,6	-5,9 %
2002	78 383	131,1	-11,8 %
2003	60 597	100,7	-23,2 %
2004	55 496	91,7	-9,0 %
moyenne	85 744	145,4	-7,4 %

Tableau 46 : Effectif et incidence d'après la projection, blessés toutes gravités, automobilistes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	306 644	293 262	315 310	528,5	505,4	543,4	
1997	305 599	292 883	313 134	525,0	503,2	538,0	-0,7 %
1998	313 573	300 907	320 825	537,0	515,3	549,4	2,3 %
1999	309 328	297 559	315 714	527,3	507,3	538,2	-1,8 %
2000	302 175	291 289	307 841	512,1	493,6	521,7	-2,9 %
2001	290 577	279 662	296 115	489,2	470,9	498,6	-4,5 %
2002	262 045	251 576	267 081	438,4	420,9	446,8	-10,4 %
2003	208 178	199 119	212 602	346,1	331,0	353,4	-21,1 %
2004	192 357	183 415	196 741	317,8	303,1	325,1	-8,2 %
moyenne	276 720	265 519	282 818	469,0	450,1	479,4	-5,9 %

Blessés toutes gravités, usagers autres

Tableau 47 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, usagers autres

année	effectif	incidence	évolution
1996	6 475	11,2	
1997	6 091	10,5	-6,2 %
1998	6 276	10,7	2,7 %
1999	6 199	10,6	-1,7 %
2000	5 289	9,0	-15,2 %
2001	4 915	8,3	-7,7 %
2002	4 483	7,5	-9,4 %
2003	3 852	6,4	-14,6 %
2004	3 463	5,7	-10,6 %
moyenne	5 227	8,9	-7,8 %

Tableau 48 : Effectif et incidence d'après la projection, blessés toutes gravités, usagers autres

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	23 786	21 327	25 301	41,0	36,8	43,6	
1997	23 229	20 759	24 657	39,9	35,7	42,4	-2,6 %
1998	24 193	21 652	25 632	41,4	37,1	43,9	3,8 %
1999	24 765	22 103	26 287	42,2	37,7	44,8	1,9 %
2000	20 953	18 764	22 223	35,5	31,8	37,7	-15,9 %
2001	19 536	17 483	20 683	32,9	29,4	34,8	-7,4 %
2002	18 676	16 633	19 776	31,2	27,8	33,1	-5,0 %
2003	17 293	15 305	18 408	28,7	25,4	30,6	-8,0 %
2004	15 177	13 462	16 145	25,1	22,2	26,7	-12,8 %
moyenne	20 845	18 610	22 124	35,3	31,5	37,5	-5,7 %

Blessés toutes gravités, par sexe

Hommes

Tableau 49 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, hommes

année	effectif	incidence	évolution
1996	106 034	376,1	
1997	105 601	373,5	-0,7 %
1998	104 551	368,6	-1,3 %
1999	104 876	368,2	-0,1 %
2000	104 028	363,0	-1,4 %
2001	99 754	345,8	-4,7 %
2002	89 761	309,1	-10,6 %
2003	74 896	256,3	-17,1 %
2004	70 686	240,4	-6,2 %
moyenne	95 576	332,8	-5,3 %

Tableau 50 : Effectif et incidence d'après la projection, blessés toutes gravités, hommes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	352 218	332 082	366 816	1 249,2	1 177,8	1 301,0	
1997	355 535	336 336	368 553	1 257,4	1 189,5	1 303,4	0,7 %
1998	358 263	339 473	370 571	1 263,2	1 197,0	1 306,6	0,5 %
1999	358 667	341 246	369 144	1 259,1	1 197,9	1 295,8	-0,3 %
2000	359 277	342 991	368 254	1 253,6	1 196,8	1 284,9	-0,4 %
2001	343 948	328 954	351 196	1 192,3	1 140,3	1 217,4	-4,9 %
2002	321 828	306 019	330 081	1 108,3	1 053,9	1 136,7	-7,0 %
2003	283 020	267 845	290 976	968,4	916,5	995,6	-12,6 %
2004	266 114	251 562	273 546	905,0	855,5	930,3	-6,5 %
moyenne	333 208	317 203	342 545	1160,1	1104,4	1192,6	-3,8 %

Femmes

Tableau 51 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, femmes

année	effectif	incidence	évolution
1996	64 083	214,8	
1997	63 976	213,7	-0,5 %
1998	63 984	213,0	-0,3 %
1999	62 696	207,8	-2,5 %
2000	58 089	191,4	-7,9 %
2001	54 191	177,4	-7,3 %
2002	48 078	156,4	-11,8 %
2003	40 274	130,2	-16,7 %
2004	38 041	122,3	-6,1 %
moyenne	54 824	180,3	-6,6 %

Tableau 52 : Effectif et incidence d'après la projection, blessés toutes gravités, femmes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	202 470	192 935	209 314	678,7	646,8	701,7	
1997	206 055	196 657	212 570	688,4	657,0	710,2	1,4 %
1998	208 056	199 120	213 677	692,7	662,9	711,4	0,6 %
1999	204 261	196 256	209 095	676,9	650,4	693,0	-2,3 %
2000	189 623	182 848	193 333	624,7	602,4	637,0	-7,7 %
2001	180 087	173 225	183 818	589,6	567,1	601,8	-5,6 %
2002	163 024	156 544	166 624	530,3	509,3	542,0	-10,0 %
2003	142 421	136 048	145 963	460,5	439,9	471,9	-13,2 %
2004	134 122	127 891	137 121	431,0	411,0	440,7	-6,4 %
moyenne	181 124	173 929	185 277	595,7	572,0	609,3	-5,4 %

Blessés toutes gravités, par classe d'âge

Tableau 53 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, 0-4 ans

année	effectif	incidence	évolution
1996	2729	76,6	
1997	2657	75,0	-2,1 %
1998	2577	72,6	-3,1 %
1999	2668	74,4	2,5 %
2000	2541	69,8	-6,2 %
2001	2373	64,2	-8,0 %
2002	2132	56,9	-11,4 %
2003	1759	46,4	-18,5 %
2004	1671	43,7	-5,7 %
moyenne	2345	64,4	-6,6 %

Tableau 54 : Effectif et incidence d'après la projection, blessés toutes gravités, 0-4 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	8247	7776	8613	231,4	218,2	241,7	
1997	8335	7849	8720	235,2	221,5	246,1	1,6 %
1998	8864	8290	9318	249,9	233,7	262,7	6,2 %
1999	8982	8462	9391	250,6	236,1	262,0	0,3 %
2000	8280	7837	8591	227,4	215,3	236,0	-9,2 %
2001	7886	7448	8188	213,3	201,4	221,5	-6,2 %
2002	7397	6921	7710	197,3	184,6	205,7	-7,5 %
2003	6147	5759	6399	162,0	151,8	168,6	-17,9 %
2004	5676	5329	5913	148,4	139,4	154,6	-8,4 %
moyenne	7757	7297	8094	212,8	200,2	222,1	-5,1 %

Tableau 55 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, 5-9 ans

année	effectif	incidence	évolution
1996	5415	142,7	
1997	4977	132,1	-7,4 %
1998	4681	125,8	-4,8 %
1999	4544	124,0	-1,4 %
2000	4148	114,5	-7,7 %
2001	3809	105,8	-7,6 %
2002	3311	92,3	-12,8 %
2003	2738	76,1	-17,6 %
2004	2456	67,5	-11,3 %
moyenne	4009	109,0	-8,8 %

Tableau 56 : Effectif et incidence d'après la projection, blessés toutes gravités, 5-9 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	17112	16077	17919	450,8	423,6	472,1	
1997	16271	15260	17008	431,9	405,1	451,5	-4,2 %
1998	16039	15002	16830	431,1	403,3	452,4	-0,2 %
1999	15521	14602	16203	423,6	398,5	442,2	-1,7 %
2000	13876	13155	14350	382,9	363,0	395,9	-9,6 %
2001	12435	11819	12812	345,3	328,2	355,8	-9,8 %
2002	11509	10858	11938	320,8	302,6	332,7	-7,1 %
2003	9725	9153	10072	270,2	254,3	279,8	-15,8 %
2004	8560	8057	8844	235,2	221,3	243,0	-13,0 %
moyenne	13450	12665	13997	365,7	344,4	380,6	-7,7 %

Tableau 57 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, 10-14 ans

année	effectif	incidence	évolution
1996	8018	208,4	
1997	7677	200,9	-3,6 %
1998	7350	192,4	-4,2 %
1999	7330	190,9	-0,8 %
2000	6998	181,8	-4,8 %
2001	6224	161,9	-10,9 %
2002	5497	143,7	-11,2 %
2003	4673	123,4	-14,1 %
2004	4397	117,8	-4,6 %
moyenne	6463	169,0	-6,8 %

Tableau 58 : Effectif et incidence d'après la projection, blessés toutes gravités, 10-14 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	32464	29470	35278	843,6	765,8	916,8	
1997	32253	29221	35098	844,1	764,7	918,5	0,1 %
1998	30025	27503	32273	786,1	720,0	844,9	-6,9 %
1999	30624	28179	32769	797,6	734,0	853,5	1,5 %
2000	28601	26483	30388	743,0	688,0	789,4	-6,8 %
2001	24578	22906	25922	639,5	596,0	674,4	-13,9 %
2002	22047	20511	23248	576,5	536,3	607,9	-9,9 %
2003	21146	19385	22599	558,6	512,1	597,0	-3,1 %
2004	19802	18197	21062	530,6	487,6	564,4	-5,0 %
moyenne	26838	24651	28737	702,2	644,9	751,9	-5,5 %

Tableau 59 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, 15-19 ans

année	effectif	incidence	évolution
1996	25623	665,6	
1997	26140	668,6	0,4 %
1998	26165	665,0	-0,5 %
1999	27003	685,4	3,1 %
2000	26215	667,9	-2,6 %
2001	24092	617,8	-7,5 %
2002	20967	540,0	-12,6 %
2003	18216	468,2	-13,3 %
2004	17535	448,1	-4,3 %
moyenne	23551	602,9	-4,7 %

Tableau 60 : Effectif et incidence d'après la projection, blessés toutes gravités, 15-19 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	89507	83894	93765	2325,0	2179,2	2435,6	
1997	90467	85423	93894	2313,8	2184,8	2401,5	-0,5 %
1998	93035	87712	96583	2364,4	2229,1	2454,5	2,2 %
1999	94516	89817	97359	2399,1	2279,8	2471,3	1,5 %
2000	92688	88410	94972	2361,4	2252,4	2419,6	-1,6 %
2001	86408	82211	88681	2215,8	2108,1	2274,1	-6,2 %
2002	78119	73867	80508	2011,8	1902,3	2073,3	-9,2 %
2003	68827	65012	70640	1769,0	1670,9	1815,6	-12,1 %
2004	67325	63433	69205	1720,6	1621,2	1768,7	-2,7 %
moyenne	84544	79975	87290	2164,5	2047,5	2234,9	-3,6 %

Tableau 61 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, 20-29 ans

année	effectif	incidence	évolution
1996	48456	587,7	
1997	47464	585,0	-0,5 %
1998	46902	585,5	0,1 %
1999	46237	583,2	-0,4 %
2000	44964	570,3	-2,2 %
2001	42572	542,0	-5,0 %
2002	37318	477,6	-11,9 %
2003	30318	390,8	-18,2 %
2004	28176	365,1	-6,6 %
moyenne	41379	520,8	-5,6 %

Tableau 62 : Effectif et incidence d'après la projection, blessés toutes gravités, 20-29 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	160394	152562	165747	1945,4	1850,4	2010,3	
1997	159517	152008	164414	1966,2	1873,7	2026,6	1,1 %
1998	158930	151815	163080	1984,0	1895,2	2035,8	0,9 %
1999	155882	149536	159224	1966,2	1886,2	2008,4	-0,9 %
2000	153137	147218	156137	1942,2	1867,1	1980,2	-1,2 %
2001	146540	140870	149240	1865,5	1793,3	1899,9	-3,9 %
2002	131681	126277	134163	1685,1	1616,0	1716,9	-9,7 %
2003	112388	107221	114895	1448,6	1382,0	1480,9	-14,0 %
2004	104537	99470	106934	1354,4	1288,8	1385,5	-6,5 %
moyenne	142556	136331	145982	1795,3	1717,0	1838,3	-4,3 %

Tableau 63 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, 30-54 ans

année	effectif	incidence	évolution
1996	56174	277,9	
1997	57118	278,8	0,3 %
1998	57429	277,5	-0,5 %
1999	57032	272,8	-1,7 %
2000	55451	262,4	-3,8 %
2001	53950	253,4	-3,4 %
2002	49098	229,9	-9,3 %
2003	40808	190,8	-17,0 %
2004	38339	179,4	-6,0 %
moyenne	51711	247,0	-5,2 %

Tableau 64 : Effectif et incidence d'après la projection, blessés toutes gravités, 30-54 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	178407	169447	184807	882,5	838,1	914,1	
1997	184095	175189	189966	898,6	855,1	927,3	1,8 %
1998	187526	179116	192769	906,1	865,5	931,5	0,8 %
1999	188508	180223	193499	901,8	862,2	925,7	-0,5 %
2000	185677	178134	189968	878,5	842,8	898,8	-2,6 %
2001	181971	174640	185778	854,6	820,2	872,5	-2,7 %
2002	171110	163540	175178	801,3	765,8	820,3	-6,2 %
2003	149018	141690	152729	696,8	662,5	714,1	-13,0 %
2004	140968	133808	144523	659,5	626,0	676,1	-5,4 %
moyenne	174142	166199	178802	831,1	793,1	853,4	-3,5 %

Tableau 65 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, 55-64 ans

année	effectif	incidence	évolution
1996	9990	179,3	
1997	9663	176,4	-1,6 %
1998	9490	174,3	-1,2 %
1999	9383	172,2	-1,2 %
2000	8974	163,9	-4,8 %
2001	8604	153,5	-6,4 %
2002	8132	138,9	-9,5 %
2003	7145	116,5	-16,1 %
2004	6993	109,1	-6,4 %
moyenne	8708	153,8	-5,9 %

Tableau 66 : Effectif et incidence d'après la projection, blessés toutes gravités, 55-64 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	30987	29152	32358	556,1	523,2	580,7	
1997	31353	29365	32957	572,2	536,0	601,5	2,9 %
1998	30581	28833	31865	561,6	529,5	585,2	-1,9 %
1999	30269	28644	31405	555,4	525,5	576,2	-1,1 %
2000	29747	28101	30897	543,3	513,3	564,3	-2,2 %
2001	27171	25960	27861	484,6	463,0	496,9	-10,8 %
2002	27592	26131	28537	471,2	446,2	487,3	-2,8 %
2003	27062	25018	28615	441,3	408,0	466,6	-6,3 %
2004	24671	23209	25575	384,9	362,1	399,0	-12,8 %
moyenne	28826	27157	30008	507,9	478,5	528,6	-4,4 %

Tableau 67 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, 65-74 ans

année	effectif	incidence	évolution
1996	8240	159,8	
1997	8117	157,2	-1,6 %
1998	8160	157,9	0,4 %
1999	7633	147,7	-6,5 %
2000	7198	139,3	-5,7 %
2001	6751	131,0	-5,9 %
2002	6075	118,2	-9,8 %
2003	5028	98,3	-16,9 %
2004	4728	92,8	-5,5 %
moyenne	6881	133,6	-6,4 %

Tableau 68 : Effectif et incidence d'après la projection, blessés toutes gravités, 65-74 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	23176	21976	24033	449,6	426,3	466,2	
1997	23646	22343	24586	458,0	432,8	476,2	1,9 %
1998	25500	23942	26680	493,5	463,3	516,3	7,7 %
1999	22970	21717	23804	444,3	420,1	460,5	-10,0 %
2000	21609	20525	22280	418,2	397,2	431,2	-5,9 %
2001	21152	19975	21942	410,5	387,7	425,8	-1,8 %
2002	20070	18851	20888	390,6	366,9	406,6	-4,8 %
2003	17011	15948	17750	332,5	311,7	346,9	-14,9 %
2004	15637	14668	16216	307,0	288,0	318,4	-7,7 %
moyenne	21197	19994	22020	411,6	388,2	427,6	-4,4 %

Tableau 69 : Effectif et incidence d'après les BAAC France, blessés toutes gravités, 75 ans et plus

année	effectif	incidence	évolution
1996	5472	144,8	
1997	5764	146,9	1,4 %
1998	5781	142,5	-3,0 %
1999	5742	137,3	-3,7 %
2000	5628	130,6	-4,9 %
2001	5570	125,4	-4,0 %
2002	5309	116,1	-7,4 %
2003	4485	95,6	-17,7 %
2004	4432	92,0	-3,8 %
moyenne	5354	125,7	-5,4 %

Tableau 70 : Effectif et incidence d'après la projection, blessés toutes gravités, 75 ans et plus

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	14394	13629	14997	380,8	360,6	396,8	
1997	15652	14824	16245	398,8	377,7	413,9	4,7 %
1998	15819	15035	16363	389,9	370,6	403,3	-2,2 %
1999	15658	14905	16141	374,4	356,4	385,9	-4,0 %
2000	15284	14586	15707	354,6	338,4	364,4	-5,3 %
2001	15895	15119	16381	357,8	340,3	368,7	0,9 %
2002	15328	14529	15826	335,3	317,8	346,2	-6,3 %
2003	14116	13226	14752	301,0	282,0	314,5	-10,2 %
2004	13060	12331	13507	271,2	256,0	280,5	-9,9 %
moyenne	15023	14243	15547	351,5	333,3	363,8	-4,0 %

Hospitalisés, par type d'usager

Tableau 71 : Effectif et incidence d'après la projection, blessés hospitalisés, piétons

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	16 492	15 700	17 193	28,4	27,1	29,6	
1997	16 004	15 284	16 637	27,5	26,3	28,6	-3,3 %
1998	15 409	14 731	15 990	26,4	25,2	27,4	-4,0 %
1999	15 443	14 770	15 981	26,3	25,2	27,2	-0,2 %
2000	14 971	14 306	15 481	25,4	24,2	26,2	-3,6 %
2001	14 448	13 749	14 962	24,3	23,1	25,2	-4,1 %
2002	13 456	12 776	13 962	22,5	21,4	23,4	-7,5 %
2003	11 952	11 289	12 447	19,9	18,8	20,7	-11,7 %
2004	11 850	11 138	12 386	19,6	18,4	20,5	-1,5 %
moyenne	14 447	13 786	14 957	24,4	23,3	25,3	-4,5 %

Tableau 72 : Effectif et incidence d'après la projection, blessés hospitalisés, cyclistes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	15 671	12 604	18 955	27,0	21,7	32,7	
1997	16 256	13 136	19 623	27,9	22,6	33,7	3,4 %
1998	14 093	11 316	17 180	24,1	19,4	29,4	-13,6 %
1999	13 135	10 679	15 798	22,4	18,2	26,9	-7,2 %
2000	11 425	9 311	13 713	19,4	15,8	23,2	-13,5 %
2001	9 924	8 117	11 854	16,7	13,7	20,0	-13,7 %
2002	10 852	8 647	13 372	18,2	14,5	22,4	8,6 %
2003	11 902	9 517	14 566	19,8	15,8	24,2	9,0 %
2004	8 608	6 972	10 415	14,2	11,5	17,2	-28,1 %
moyenne	12 430	10 038	15 030	21,0	17,0	25,4	-6,9 %

Tableau 73 : Effectif et incidence d'après la projection, blessés hospitalisés, usagers de 2RM

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	37 240	34 792	39 112	64,2	60,0	67,4	
1997	40 114	37 582	41 918	68,9	64,6	72,0	7,4 %
1998	39 802	37 331	41 529	68,2	63,9	71,1	-1,1 %
1999	39 797	37 426	41 401	67,8	63,8	70,6	-0,5 %
2000	39 840	37 464	41 428	67,5	63,5	70,2	-0,5 %
2001	39 239	36 876	40 799	66,1	62,1	68,7	-2,1 %
2002	37 274	34 943	38 760	62,4	58,5	64,8	-5,6 %
2003	35 754	33 323	37 401	59,4	55,4	62,2	-4,7 %
2004	35 491	32 933	37 184	58,6	54,4	61,4	-1,3 %
moyenne	38 283	35 964	39 821	64,7	60,8	67,3	-1,1 %

Tableau 74 : Effectif et incidence d'après la projection, blessés hospitalisés, automobilistes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	75 796	72 125	78 508	130,6	124,3	135,3	
1997	75 870	72 393	78 353	130,3	124,4	134,6	-0,2 %
1998	76 422	72 992	78 682	130,9	125,0	134,7	0,4 %
1999	74 285	71 187	76 390	126,6	121,4	130,2	-3,2 %
2000	70 495	67 499	72 419	119,5	114,4	122,7	-5,7 %
2001	68 541	65 441	70 441	115,4	110,2	118,6	-3,4 %
2002	62 817	59 714	64 694	105,1	99,9	108,2	-8,9 %
2003	49 312	46 667	50 978	82,0	77,6	84,7	-22,0 %
2004	44 973	42 347	46 652	74,3	70,0	77,1	-9,4 %
moyenne	66 501	63 708	68 324	112,5	107,7	115,6	-6,5 %

Hospitalisés, autres usagers

Tableau 75 : Effectif et incidence d'après la projection, blessés hospitalisés, autres usagers

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	6 229	5 409	6 875	10,7	9,3	11,8	
1997	6 167	5 341	6 789	10,6	9,2	11,7	-1,3 %
1998	6 404	5 546	7 033	11,0	9,5	12,0	3,5 %
1999	6 285	5 425	6 908	10,7	9,2	11,8	-2,3 %
2000	5 069	4 351	5 585	8,6	7,4	9,5	-19,8 %
2001	4 707	4 041	5 185	7,9	6,8	8,7	-7,7 %
2002	4 581	3 922	5 048	7,7	6,6	8,4	-3,3 %
2003	4 636	3 953	5 105	7,7	6,6	8,5	0,6 %
2004	3 646	3 106	4 035	6,0	5,1	6,7	-21,8 %
moyenne	5 303	4 573	5 824	9,0	7,7	9,8	-6,5 %

Hospitalisés, par sexe

Tableau 76 : Effectif et incidence d'après la projection, blessés hospitalisés, hommes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	105385	99452	109465	373,8	352,7	388,2	
1997	107457	101816	111191	380,0	360,1	393,2	1,7 %
1998	105829	100547	109133	373,1	354,5	384,8	-1,8 %
1999	104062	99248	106825	365,3	348,4	375,0	-2,1 %
2000	102091	97449	104556	356,2	340,0	364,8	-2,5 %
2001	99107	94657	101358	343,6	328,1	351,4	-3,6 %
2002	94541	89822	97093	325,6	309,3	334,4	-5,2 %
2003	83362	78673	86030	285,2	269,2	294,4	-12,4 %
2004	76054	71744	78660	258,6	244,0	267,5	-9,3 %
moyenne	97543	92601	100479	340,2	322,9	350,4	-4,4 %

Tableau 77 : Effectif et incidence d'après la projection, blessés hospitalisés, femmes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	46042	43729	47817	154,3	146,6	160,3	
1997	46954	44746	48581	156,9	149,5	162,3	1,6 %
1998	46300	44219	47725	154,1	147,2	158,9	-1,7 %
1999	44882	42907	46207	148,7	142,2	153,1	-3,5 %
2000	39710	38017	40783	130,8	125,3	134,4	-12,0 %
2001	37752	36077	38848	123,6	118,1	127,2	-5,5 %
2002	34440	32781	35501	112,0	106,6	115,5	-9,3 %
2003	30195	28558	31223	97,6	92,3	101,0	-12,9 %
2004	28514	26847	29576	91,6	86,3	95,1	-6,1 %
moyenne	39421	37542	40696	130,0	123,8	134,2	-6,2 %

Hospitalisés, par classe d'âge

Tableau 78 : Effectif et incidence d'après la projection, blessés hospitalisés, 0-4 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	3248	2996	3466	91,1	84,1	97,3	
1997	3184	2932	3401	89,9	82,8	96,0	-1,4 %
1998	3286	3017	3501	92,6	85,0	98,7	3,1 %
1999	3301	3031	3524	92,1	84,6	98,3	-0,6 %
2000	3106	2863	3312	85,3	78,6	91,0	-7,4 %
2001	2929	2699	3124	79,2	73,0	84,5	-7,1 %
2002	2817	2594	2994	75,2	69,2	79,9	-5,1 %
2003	2294	2106	2443	60,5	55,5	64,4	-19,6 %
2004	2096	1918	2245	54,8	50,2	58,7	-9,3 %
moyenne	2918	2684	3112	80,1	73,7	85,4	-5,9 %

Tableau 79 : Effectif et incidence d'après la projection, blessés hospitalisés, 5-9 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	5975	5549	6341	157,4	146,2	167,1	
1997	5581	5188	5920	148,1	137,7	157,1	-5,9 %
1998	5145	4778	5472	138,3	128,4	147,1	-6,6 %
1999	4966	4610	5270	135,5	125,8	143,8	-2,0 %
2000	4489	4172	4744	123,9	115,1	130,9	-8,6 %
2001	4119	3835	4347	114,4	106,5	120,7	-7,7 %
2002	3756	3479	3967	104,7	97,0	110,6	-8,5 %
2003	3234	2984	3417	89,8	82,9	94,9	-14,2 %
2004	2752	2531	2920	75,6	69,5	80,2	-15,9 %
moyenne	4446	4125	4711	120,9	112,1	128,0	-8,7 %

Tableau 80 : Effectifs d'après les BAAC France, blessés hospitalisés, 10-14 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	9849	8909	10710	255,9	231,5	278,3	
1997	9598	8700	10426	251,2	227,7	272,9	-1,9 %
1998	8915	8141	9598	233,4	213,1	251,3	-7,1 %
1999	9214	8414	9872	240,0	219,2	257,1	2,8 %
2000	8360	7663	8920	217,2	199,1	231,7	-9,5 %
2001	7291	6712	7733	189,7	174,6	201,2	-12,7 %
2002	6537	5998	6942	170,9	156,8	181,5	-9,9 %
2003	6071	5537	6509	160,4	146,3	171,9	-6,2 %
2004	5659	5142	6070	151,6	137,8	162,6	-5,4 %
moyenne	7944	7246	8531	207,8	189,6	223,2	-6,2 %

Tableau 81 : Effectif et incidence d'après la projection, blessés hospitalisés, 15-19 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	24387	22826	25638	633,5	592,9	666,0	
1997	25528	23987	26657	652,9	613,5	681,8	3,1 %
1998	25799	24223	26921	655,6	615,6	684,2	0,4 %
1999	25411	23987	26418	645,0	608,9	670,6	-1,6 %
2000	24150	22788	25048	615,3	580,6	638,1	-4,6 %
2001	22745	21437	23610	583,3	549,7	605,4	-5,2 %
2002	20915	19654	21735	538,6	506,2	559,7	-7,7 %
2003	18812	17593	19587	483,5	452,2	503,4	-10,2 %
2004	18146	16895	18956	463,8	431,8	484,5	-4,1 %
moyenne	22877	21488	23841	585,7	550,1	610,4	-3,7 %

Tableau 82 : Effectif et incidence d'après la projection, blessés hospitalisés, 20-29 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	39793	37664	41260	482,6	456,8	500,4	
1997	39725	37715	41053	489,7	464,9	506,0	1,5 %
1998	38135	36294	39264	476,1	453,1	490,2	-2,8 %
1999	37161	35425	38156	468,7	446,8	481,3	-1,5 %
2000	35139	33506	36060	445,7	424,9	457,3	-4,9 %
2001	34089	32428	35022	434,0	412,8	445,8	-2,6 %
2002	31220	29590	32151	399,5	378,7	411,4	-7,9 %
2003	26365	24839	27243	339,8	320,2	351,1	-14,9 %
2004	24458	22932	25384	316,9	297,1	328,9	-6,7 %
moyenne	34009	32266	35066	428,1	406,1	441,4	-5,0 %

Tableau 83 : Effectif et incidence d'après la projection, blessés hospitalisés, 30-54 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	45107	42562	46903	223,1	210,5	232,0	
1997	47098	44572	48685	229,9	217,6	237,6	3,0 %
1998	47407	45013	48896	229,1	217,5	236,3	-0,4 %
1999	46287	44012	47688	221,4	210,6	228,1	-3,3 %
2000	45023	42836	46322	213,0	202,7	219,2	-3,8 %
2001	45027	42853	46295	211,5	201,2	217,4	-0,7 %
2002	43039	40754	44431	201,5	190,8	208,1	-4,7 %
2003	37545	35375	38841	175,6	165,4	181,6	-12,9 %
2004	34250	32208	35548	160,2	150,7	166,3	-8,7 %
moyenne	43420	41132	44845	207,3	196,3	214,1	-3,9 %

Tableau 84 : Effectifs d'après les BAAC France, blessés hospitalisés, 55-64 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	9155	8457	9711	164,3	151,8	174,3	
1997	9153	8435	9725	167,1	154,0	177,5	1,7 %
1998	8787	8125	9293	161,4	149,2	170,7	-3,4 %
1999	8610	7966	9082	158,0	146,2	166,6	-2,1 %
2000	8210	7591	8667	150,0	138,7	158,3	-5,1 %
2001	7525	6993	7899	134,2	124,7	140,9	-10,5 %
2002	8166	7538	8620	139,5	128,7	147,2	3,9 %
2003	7578	6859	8147	123,6	111,8	132,9	-11,4 %
2004	7024	6454	7436	109,6	100,7	116,0	-11,3 %
moyenne	8245	7602	8731	145,3	134,0	153,8	-4,8 %

Tableau 85 : Effectif et incidence d'après la projection, blessés hospitalisés, 65-74 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	8286	7723	8737	160,7	149,8	169,5	
1997	8479	7895	8940	164,2	152,9	173,2	2,2 %
1998	8443	7867	8912	163,4	152,2	172,5	-0,5 %
1999	7786	7270	8201	150,6	140,6	158,6	-7,8 %
2000	7370	6874	7759	142,6	133,0	150,2	-5,3 %
2001	7173	6649	7583	139,2	129,0	147,2	-2,4 %
2002	6617	6118	7021	128,8	119,1	136,7	-7,5 %
2003	6030	5560	6397	117,9	108,7	125,0	-8,5 %
2004	5095	4690	5402	100,0	92,1	106,1	-15,1 %
moyenne	7253	6738	7661	140,8	130,8	148,8	-5,6 %

Tableau 86 : Effectif et incidence d'après la projection, blessés hospitalisés, 75 ans et plus

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	5628	5238	5974	148,9	138,6	158,1	
1997	6065	5649	6420	154,5	143,9	163,6	3,8 %
1998	6211	5787	6547	153,1	142,6	161,4	-0,9 %
1999	6207	5789	6543	148,4	138,4	156,4	-3,1 %
2000	5954	5545	6273	138,1	128,6	145,5	-6,9 %
2001	5961	5531	6292	134,2	124,5	141,6	-2,9 %
2002	5913	5480	6232	129,3	119,9	136,3	-3,6 %
2003	5627	5168	5947	120,0	110,2	126,8	-7,2 %
2004	5088	4691	5369	105,6	97,4	111,5	-11,9 %
moyenne	5850	5431	6177	136,9	127,1	144,6	-4,1 %

Blessés graves, par type d'usager

Blessés graves, piétons

Tableau 87 : Effectifs d'après les BAAC France, blessés graves (hosp.>6 jours), piétons

année	effectif	incidence	évolution
1996	4 769	8,2	
1997	4 566	7,8	-4,6 %
1998	4 224	7,2	-7,8 %
1999	3 981	6,8	-6,2 %
2000	3 532	6,0	-11,8 %
2001	3 170	5,3	-10,8 %
2002	2 939	4,9	-7,9 %
2003	2 478	4,1	-16,2 %
2004	2 310	3,8	-7,3 %
moyenne	3 552	6,0	-9,1 %

Tableau 88 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), piétons

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	7 454	6 933	7 906	12,8	11,9	13,6	
1997	7 348	6 867	7 759	12,6	11,8	13,3	-1,7 %
1998	7 169	6 721	7 550	12,3	11,5	12,9	-2,8 %
1999	7 139	6 700	7 497	12,2	11,4	12,8	-0,9 %
2000	6 856	6 432	7 221	11,6	10,9	12,2	-4,5 %
2001	6 656	6 222	7 030	11,2	10,5	11,8	-3,5 %
2002	6 314	5 890	6 683	10,6	9,9	11,2	-5,7 %
2003	5 656	5 238	6 010	9,4	8,7	10,0	-11,0 %
2004	5 630	5 178	6 005	9,3	8,6	9,9	-1,1 %
moyenne	6691	6 242	7 073	11,3	10,6	12,0	-3,9 %

Blessés graves, cyclistes

Tableau 89 : Effectifs d'après les BAAC France, blessés graves (hosp.> 6 jours), cyclistes

année	effectif	incidence	évolution
1996	1 643	2,8	
1997	1 666	2,9	1,1 %
1998	1 419	2,4	-15,1 %
1999	1 281	2,2	-10,1 %
2000	1 039	1,8	-19,4 %
2001	925	1,6	-11,5 %
2002	850	1,4	-8,7 %
2003	841	1,4	-1,7 %
2004	669	1,1	-20,9 %
moyenne	1 148	2,0	-10,8 %

Tableau 90 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), cyclistes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	6 838	5 177	8 574	11,8	8,9	14,8	
1997	7 199	5 495	8 978	12,4	9,4	15,4	5,0 %
1998	6 165	4 707	7 674	10,6	8,1	13,1	-14,6 %
1999	5 747	4 449	7 078	9,8	7,6	12,1	-7,2 %
2000	5 041	3 924	6 187	8,5	6,6	10,5	-12,8 %
2001	4 477	3 501	5 458	7,5	5,9	9,2	-11,8 %
2002	5 167	3 879	6 437	8,6	6,5	10,8	14,7 %
2003	5 721	4 313	7 112	9,5	7,2	11,8	10,0 %
2004	4 062	3 133	5 012	6,7	5,2	8,3	-29,4 %
moyenne	5602	4 286	6 946	9,5	7,3	11,8	-5,8 %

Blessés graves, usagers de deux-roues motorisés

Tableau 91 : Effectifs d'après les BAAC France, blessés graves (hosp.>6 jours), usagers de 2RM

année	effectif	incidence	évolution
1996	8 896	15,3	
1997	9 383	16,1	5,1 %
1998	8 756	15,0	-7,0 %
1999	8 313	14,2	-5,5 %
2000	7 604	12,9	-9,1 %
2001	7 353	12,4	-3,9 %
2002	6 924	11,6	-6,4 %
2003	5 829	9,7	-16,3 %
2004	5 659	9,4	-3,5 %
moyenne	7 635	13,0	-5,8 %

Tableau 92 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), usagers de 2RM

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	20 236	18 622	21 469	34,9	32,1	37,0	
1997	22 157	20 524	23 388	38,1	35,3	40,2	9,2 %
1998	22 065	20 487	23 150	37,8	35,1	39,6	-0,7 %
1999	22 077	20 546	23 101	37,6	35,0	39,4	-0,4 %
2000	22 035	20 501	23 005	37,3	34,7	39,0	-0,8 %
2001	22 109	20 536	23 070	37,2	34,6	38,8	-0,3 %
2002	21 407	19 854	22 330	35,8	33,2	37,4	-3,8 %
2003	20 444	18 782	21 416	34,0	31,2	35,6	-5,1 %
2004	20 539	18 772	21 577	33,9	31,0	35,7	-0,1 %
moyenne	21452	19 847	22 501	36,3	33,6	38,1	-0,3 %

Blessés graves, automobilistes

Tableau 93 : Effectifs d'après les BAAC France, blessés graves (hosp.> 6 jours), automobilistes

année	effectif	incidence	évolution
1996	19 761	34,1	
1997	19 003	32,6	-4,1 %
1998	18 522	31,7	-2,8 %
1999	17 264	29,4	-7,2 %
2000	14 569	24,7	-16,1 %
2001	14 121	23,8	-3,7 %
2002	12 767	21,4	-10,2 %
2003	9 357	15,6	-27,2 %
2004	8 355	13,8	-11,2 %
moyenne	14 858	25,2	-10,3 %

Tableau 94 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), automobilistes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	27 669	25 760	29 184	47,7	44,4	50,3	
1997	28 022	26 237	29 356	48,1	45,1	50,4	1,0 %
1998	28 530	26 886	29 731	48,9	46,0	50,9	1,5 %
1999	27 908	26 337	28 986	47,6	44,9	49,4	-2,6 %
2000	26 112	24 631	27 117	44,2	41,7	46,0	-7,0 %
2001	26 071	24 550	27 089	43,9	41,3	45,6	-0,8 %
2002	24 353	22 832	25 367	40,7	38,2	42,4	-7,2 %
2003	19 271	17 964	20 187	32,0	29,9	33,6	-21,4 %
2004	17 782	16 442	18 732	29,4	27,2	31,0	-8,3 %
moyenne	25080	23 515	26 194	42,5	39,9	44,4	-5,6 %

Blessés graves, usagers autres

Tableau 95 : Effectifs d'après les BAAC France, blessés graves (hosp. > 6 jours), usager autres

année	effectif	incidence	évolution
1996	1 135	2,0	
1997	1 098	1,9	-3,6 %
1998	1 056	1,8	-4,1 %
1999	1 012	1,7	-4,6 %
2000	663	1,1	-34,9 %
2001	623	1,0	-6,6 %
2002	611	1,0	-2,6 %
2003	605	1,0	-1,6 %
2004	442	0,7	-27,4 %
moyenne	805	1,4	-10,7 %

Tableau 96 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), usagers autres

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	2 287	1 841	2 654	3,9	3,2	4,6	
1997	2 310	1 870	2 671	4,0	3,2	4,6	0,7 %
1998	2 430	1 968	2 801	4,2	3,4	4,8	4,9 %
1999	2 373	1 919	2 736	4,0	3,3	4,7	-2,8 %
2000	1 861	1 491	2 166	3,2	2,5	3,7	-22,0 %
2001	1 783	1 431	2 068	3,0	2,4	3,5	-4,8 %
2002	1 777	1 428	2 054	3,0	2,4	3,4	-1,0 %
2003	1 885	1 509	2 172	3,1	2,5	3,6	5,4 %
2004	1 452	1 162	1 682	2,4	1,9	2,8	-23,4 %
moyenne	2018	1 624	2 334	3,4	2,8	4,0	-5,4 %

Blessés graves, par sexe

Hommes

Tableau 97 : Effectif et incidence d'après les BAAC France, blessés graves (hosp.>6 jours), hommes

année	effectif	incidence	évolution
1996	24 397	86,5	
1997	24 102	85,2	-1,5 %
1998	22 968	81,0	-5,0 %
1999	21 669	76,1	-6,1 %
2000	19 106	66,7	-12,4 %
2001	18 508	64,2	-3,8 %
2002	17 253	59,4	-7,4 %
2003	13 576	46,5	-21,8 %
2004	12 398	42,2	-9,2 %
moyenne	19 331	67,5	-8,4 %

Tableau 98 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), hommes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	46 927	43 733	49 339	166,4	155,1	175,0	
1997	48 775	45 686	50 978	172,5	161,6	180,3	3,6 %
1998	48 331	45 514	50 159	170,4	160,5	176,9	-1,2 %
1999	47 656	45 122	49 148	167,3	158,4	172,5	-1,8 %
2000	46 397	43 968	47 781	161,9	153,4	166,7	-3,2 %
2001	46 009	43 590	47 262	159,5	151,1	163,8	-1,5 %
2002	44 991	42 307	46 511	154,9	145,7	160,2	-2,9 %
2003	40 389	37 583	42 011	138,2	128,6	143,7	-10,8 %
2004	37 462	34 834	39 095	127,4	118,5	133,0	-7,8 %
moyenne	45 215	42 738	46 633	157,4	148,8	162,4	-3,2 %

Femmes

Tableau 99 : Effectif et incidence d'après les BAAC France, blessés graves (hosp.>6 jours), femmes

année	effectif	incidence	évolution
1996	11 807	39,6	
1997	11 614	38,8	-2,0 %
1998	11 009	36,7	-5,5 %
1999	10 182	33,7	-7,9 %
2000	8 301	27,3	-19,0 %
2001	7 684	25,2	-8,0 %
2002	6 838	22,2	-11,6 %
2003	5 534	17,9	-19,6 %
2004	5 037	16,2	-9,5 %
moyenne	8 667	28,6	-10,4 %

Tableau 100 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), femmes

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	17 558	16 263	18 573	58,9	54,5	62,3	
1997	18 261	16 999	19 200	61,0	56,8	64,1	3,7 %
1998	18 027	16 898	18 842	60,0	56,3	62,7	-1,6 %
1999	17 588	16 490	18 330	58,3	54,6	60,7	-2,9 %
2000	15 508	14 554	16 144	51,1	47,9	53,2	-12,3 %
2001	15 086	14 110	15 726	49,4	46,2	51,5	-3,3 %
2002	14 026	13 074	14 637	45,6	42,5	47,6	-7,6 %
2003	12 588	11 648	13 216	40,7	37,7	42,7	-10,8 %
2004	12 003	11 042	12 673	38,6	35,5	40,7	-5,2 %
moyenne	15 627	14 654	16 290	51,4	48,2	53,6	-5,0 %

Blessés graves, par classe d'âge

Tableau 101 : Effectif et incidence d'après les BAAC France, blessés graves (hosp. >6 jours), 0-4 ans

année	effectif	incidence	évolution
1996	411	11,5	
1997	378	10,7	-7,5 %
1998	362	10,2	-4,4 %
1999	363	10,1	-0,8 %
2000	298	8,2	-19,2 %
2001	271	7,3	-10,5 %
2002	284	7,6	3,4 %
2003	202	5,3	-29,7 %
2004	160	4,2	-21,4 %
moyenne	303	8,3	-11,3 %

Tableau 102 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), 0-4 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	621	467	765	17,4	13,1	21,5	
1997	594	447	736	16,8	12,6	20,8	-3,8 %
1998	615	460	761	17,3	13,0	21,5	3,4 %
1999	637	481	783	17,8	13,4	21,8	2,5 %
2000	565	420	704	15,5	11,5	19,3	-12,7 %
2001	551	413	682	14,9	11,2	18,4	-4,0 %
2002	569	430	700	15,2	11,5	18,7	1,9 %
2003	452	339	561	11,9	8,9	14,8	-21,5 %
2004	400	298	499	10,5	7,8	13,0	-12,2 %
moyenne	556	417	688	15,3	11,4	18,9	-5,8 %

Tableau 103 : Effectif et incidence d'après les BAAC France, blessés graves (hosp. >6 jours), 5-9 ans

année	effectif	incidence	évolution
1996	1043	27,5	
1997	889	23,6	-14,1 %
1998	745	20,0	-15,1 %
1999	696	19,0	-5,1 %
2000	562	15,5	-18,4 %
2001	522	14,5	-6,5 %
2002	477	13,3	-8,3 %
2003	335	9,3	-30,0 %
2004	297	8,2	-12,3 %
moyenne	618	16,8	-13,7 %

Tableau 104 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), 5-9 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	1764	1493	2016	46,5	39,3	53,1	
1997	1635	1391	1867	43,4	36,9	49,6	-6,6 %
1998	1472	1253	1692	39,6	33,7	45,5	-8,8 %
1999	1428	1214	1636	39,0	33,1	44,6	-1,5 %
2000	1270	1079	1454	35,0	29,8	40,1	-10,1 %
2001	1194	1018	1362	33,2	28,3	37,8	-5,4 %
2002	1118	953	1272	31,2	26,6	35,5	-6,0 %
2003	965	811	1104	26,8	22,5	30,7	-14,0 %
2004	806	679	924	22,1	18,7	25,4	-17,4 %
moyenne	1295	1099	1481	35,2	29,9	40,2	-8,7 %

Tableau 105 : Effectif et incidence d'après les BAAC France, blessés graves (hosp. >6 jours), 10-14 ans

année	effectif	incidence	évolution
1996	1673	43,5	
1997	1575	41,2	-5,2 %
1998	1372	35,9	-12,9 %
1999	1370	35,7	-0,7 %
2000	1092	28,4	-20,5 %
2001	950	24,7	-12,9 %
2002	845	22,1	-10,6 %
2003	772	20,4	-7,7 %
2004	656	17,6	-13,8 %
moyenne	1145	29,9	-10,5 %

Tableau 106 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), 10-14 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	3548	3052	4028	92,2	79,3	104,7	
1997	3474	3004	3924	90,9	78,6	102,7	-1,4 %
1998	3156	2735	3552	82,6	71,6	93,0	-9,1 %
1999	3385	2934	3786	88,2	76,4	98,6	6,7 %
2000	2964	2583	3292	77,0	67,1	85,5	-12,7 %
2001	2665	2335	2941	69,3	60,8	76,5	-9,9 %
2002	2371	2067	2628	62,0	54,0	68,7	-10,6 %
2003	2269	1976	2531	59,9	52,2	66,9	-3,3 %
2004	2136	1846	2390	57,2	49,5	64,0	-4,5 %
moyenne	2885	2504	3230	75,5	65,5	84,5	-5,6 %

Tableau 107 : Effectif et incidence d'après les BAAC France, blessés graves (hosp. >6 jours), 15-19 ans

année	effectif	incidence	évolution
1996	5609	145,7	
1997	5635	144,1	-1,1 %
1998	5440	138,3	-4,1 %
1999	5262	133,6	-3,4 %
2000	4547	115,8	-13,3 %
2001	4196	107,6	-7,1 %
2002	3849	99,1	-7,9 %
2003	3070	78,9	-20,4 %
2004	2944	75,2	-4,6 %
moyenne	4506	115,4	-7,7 %

Tableau 108 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), 15-19 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	11126	10224	11845	289,0	265,6	307,7	
1997	11830	10915	12534	302,6	279,2	320,6	4,7 %
1998	12058	11124	12765	306,4	282,7	324,4	1,3 %
1999	11766	10951	12370	298,7	278,0	314,0	-2,5 %
2000	11128	10368	11701	283,5	264,1	298,1	-5,1 %
2001	10746	9983	11294	275,6	256,0	289,6	-2,8 %
2002	10182	9411	10725	262,2	242,4	276,2	-4,8 %
2003	9304	8536	9812	239,1	219,4	252,2	-8,8 %
2004	9153	8354	9680	233,9	213,5	247,4	-2,2 %
moyenne	10810	9985	11414	276,8	255,6	292,2	-2,5 %

Tableau 109 : Effectif et incidence d'après les BAAC France, blessés graves (hosp. >6 jours), 20-29 ans

année	effectif	incidence	évolution
1996	9733	118,1	
1997	9498	117,1	-0,8 %
1998	8788	109,7	-6,3 %
1999	8214	103,6	-5,6 %
2000	6974	88,4	-14,6 %
2001	6672	84,9	-4,0 %
2002	5947	76,1	-10,4 %
2003	4646	59,9	-21,3 %
2004	4149	53,8	-10,2 %
moyenne	7180	90,2	-9,2 %

Tableau 110 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), 20-29 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	16873	15598	17803	204,7	189,2	215,9	
1997	17210	15998	18043	212,1	197,2	222,4	3,7 %
1998	16486	15376	17188	205,8	191,9	214,6	-3,0 %
1999	16170	15150	16809	204,0	191,1	212,0	-0,9 %
2000	15158	14213	15758	192,2	180,3	199,9	-5,7 %
2001	15011	14066	15599	191,1	179,1	198,6	-0,6 %
2002	14059	13103	14626	179,9	167,7	187,2	-5,9 %
2003	12062	11171	12607	155,5	144,0	162,5	-13,6 %
2004	11420	10511	11993	148,0	136,2	155,4	-4,8 %
moyenne	14939	13910	15603	188,1	175,2	196,5	-3,9 %

Tableau 111 : Effectif et incidence d'après les BAAC France, blessés graves (hosp. >6 jours), 30-54 ans

année	effectif	incidence	évolution
1996	11465	56,7	
1997	11502	56,1	-1,0 %
1998	11253	54,4	-3,2 %
1999	10406	49,8	-8,4 %
2000	9099	43,0	-13,5 %
2001	9037	42,4	-1,4 %
2002	8419	39,4	-7,1 %
2003	6457	30,2	-23,4 %
2004	5821	27,2	-9,8 %
moyenne	9273	44,4	-8,5 %

Tableau 112 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), 30-54 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	19981	18547	21061	98,8	91,7	104,2	
1997	21313	19913	22326	104,0	97,2	109,0	5,3 %
1998	21682	20336	22609	104,8	98,3	109,2	0,7 %
1999	21336	20067	22169	102,1	96,0	106,1	-2,6 %
2000	20800	19549	21597	98,4	92,5	102,2	-3,6 %
2001	21239	19977	22037	99,7	93,8	103,5	1,4 %
2002	20831	19462	21679	97,5	91,1	101,5	-2,2 %
2003	18468	17135	19301	86,4	80,1	90,2	-11,5 %
2004	17044	15777	17923	79,7	73,8	83,8	-7,7 %
moyenne	20299	18974	21189	96,8	90,5	101,1	-2,5 %

Tableau 113 : Effectif et incidence d'après les BAAC France, blessés graves (hosp. >6 jours), 55-64 ans

année	effectif	incidence	évolution
1996	2329	41,8	
1997	2219	40,5	-3,1 %
1998	2097	38,5	-4,9 %
1999	1950	35,8	-7,1 %
2000	1666	30,4	-14,9 %
2001	1569	28,0	-8,0 %
2002	1516	25,9	-7,5 %
2003	1290	21,0	-18,7 %
2004	1251	19,5	-7,2 %
moyenne	1765	31,3	-8,9 %

Tableau 114 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), 55-64 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	4031	3570	4424	72,3	64,1	79,4	
1997	4107	3634	4503	75,0	66,3	82,2	3,6 %
1998	3936	3507	4290	72,3	64,4	78,8	-3,6 %
1999	3867	3443	4206	70,9	63,2	77,2	-1,8 %
2000	3671	3271	3997	67,1	59,7	73,0	-5,5 %
2001	3389	3028	3671	60,4	54,0	65,5	-9,9 %
2002	3829	3394	4169	65,4	58,0	71,2	8,2 %
2003	3629	3159	4042	59,2	51,5	65,9	-9,5 %
2004	3415	3012	3722	53,3	47,0	58,1	-10,0 %
moyenne	3764	3335	4114	66,2	58,7	72,4	-3,6 %

Tableau 115 : Effectif et incidence d'après les BAAC France, blessés graves (hosp. >6 jours), 65-74 ans

année	effectif	incidence	évolution
1996	2285	44,3	
1997	2276	44,1	-0,5 %
1998	2173	42,1	-4,6 %
1999	1914	37,0	-12,0 %
2000	1655	32,0	-13,5 %
2001	1539	29,9	-6,7 %
2002	1332	25,9	-13,2 %
2003	1154	22,6	-13,0 %
2004	982	19,3	-14,5 %
moyenne	1701	33,0	-9,8 %

Tableau 116 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), 65-74 ans

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	3880	3470	4224	75,3	67,3	81,9	
1997	3989	3567	4342	77,3	69,1	84,1	2,7 %
1998	3957	3555	4303	76,6	68,8	83,3	-0,9 %
1999	3655	3285	3963	70,7	63,5	76,7	-7,7 %
2000	3487	3138	3769	67,5	60,7	72,9	-4,6 %
2001	3447	3086	3732	66,9	59,9	72,4	-0,9 %
2002	3169	2821	3441	61,7	54,9	67,0	-7,8 %
2003	3011	2671	3264	58,8	52,2	63,8	-4,6 %
2004	2510	2236	2732	49,3	43,9	53,6	-16,3 %
moyenne	3456	3092	3752	67,1	60,0	72,9	-5,0 %

Tableau 117 : Effectif et incidence d'après les BAAC France, blessés graves (hosp. >6 jours), 75 ans et plus

année	effectif	incidence	évolution
1996	1656	43,8	
1997	1744	44,4	1,4 %
1998	1747	43,1	-3,1 %
1999	1676	40,1	-6,9 %
2000	1514	35,1	-12,3 %
2001	1436	32,3	-8,0 %
2002	1422	31,1	-3,8 %
2003	1184	25,2	-18,8 %
2004	1175	24,4	-3,4 %
moyenne	1506	35,5	-6,9 %

Tableau 118 : Effectif et incidence d'après la projection, blessés graves (NISS 9+), 75 ans et plus

année	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%		évolution
1996	2662	2353	2940	70,4	62,3	77,8	
1997	2884	2562	3168	73,5	65,3	80,7	4,3 %
1998	2999	2675	3281	73,9	65,9	80,9	0,6 %
1999	3002	2675	3287	71,8	64,0	78,6	-2,9 %
2000	2863	2548	3136	66,4	59,1	72,8	-7,5 %
2001	2854	2530	3136	64,2	56,9	70,6	-3,3 %
2002	2887	2564	3163	63,2	56,1	69,2	-1,7 %
2003	2815	2474	3097	60,0	52,8	66,0	-5,0 %
2004	2581	2282	2828	53,6	47,4	58,7	-10,7 %
moyenne	2839	2518	3115	66,3	58,9	72,8	-3,3 %

Blessés avec lésion IIS3+, quelle que soit la région atteinte

Tableau 119 : effectifs et incidences, moyennes annuelles sur 1996-2004, d'après la projection, blessés avec lésion IIS3+

	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%	
ensemble	7479	6554	8158	12,6	11,1	13,8
par type d'usager :						
piétons	726	578	875	1,2	1,0	1,5
cyclistes	857	504	1207	1,4	0,9	2,0
usagers de 2RM	2249	1838	2568	3,8	3,1	4,3
automobilistes	3393	2869	3811	5,7	4,9	6,4
autres	255	115	391	0,4	0,2	0,7
par sexe						
hommes	5925	5141	6511	20,6	17,9	22,7
femmes	1554	1242	1825	5,1	4,1	6,0
par classe d'âge :						
0-4 ans	74	50	97	2,0	1,4	2,6
5-9 ans	147	101	194	4,0	2,8	5,3
10-14 ans	335	223	451	8,8	5,8	11,8
15-19 ans	1242	943	1517	31,8	24,1	38,9
20-24 ans	1917	1546	2215	24,2	19,5	27,9
25-29 ans	2613	2193	2945	12,5	10,5	14,0
30-54 ans	476	393	544	8,3	6,9	9,5
55-64 ans	378	271	481	7,3	5,3	9,3
65-74 ans	298	215	376	6,9	5,0	8,7
75 ans et +	74	50	97	2,0	1,4	2,6

Blessés avec lésion IIS3+, à la tête

Tableau 120 : effectifs et incidences, moyennes annuelles sur 1996-2004, d'après la projection, blessés avec lésion IIS3+ à la tête

	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%	
ensemble	3993	3390	4487	6,8	5,7	7,6
par type d'usager :						
piétons	474	360	592	0,8	0,6	1,0
cyclistes	561	290	819	0,9	0,5	1,4
usagers de 2RM	946	701	1165	1,6	1,2	2,0
automobilistes	1846	1510	2146	3,1	2,6	3,6
autres	166	60	285	0,3	0,1	0,5
par sexe :						
hommes	3105	2591	3536	10,8	9,0	12,3
femmes	888	670	1104	2,9	2,2	3,6
par classe d'âge :						
0-4 ans	50	33	69	1,4	0,9	1,9
5-9 ans	108	72	146	2,9	2,0	4,0
10-14 ans	229	145	317	6,0	3,8	8,3
15-19 ans	733	515	928	18,8	13,2	23,8
20-24 ans	935	695	1139	11,8	8,8	14,4
25-29 ans	1261	995	1498	6,0	4,7	7,1
30-54 ans	250	191	303	4,4	3,3	5,3
55-64 ans	234	153	318	4,6	3,0	6,2
65-74 ans	192	129	254	4,5	3,0	5,9
75 ans et +	50	33	69	1,4	0,9	1,9

Blessés avec lésion IIS3+, à la colonne vertébrale

Tableau 121 : effectifs et incidences, moyennes annuelles sur 1996-2004, d'après la projection, blessés avec lésion IIS3+ à la colonne vertébrale

	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%	
ensemble	1145	853	1438	1,9	1,4	2,4
par type d'usager :						
piétons	60	24	114	0,1	0,0	0,2
cyclistes	126	0	279	0,2	0,0	0,5
usagers de 2RM	497	323	686	0,8	0,5	1,2
automobilistes	461	292	649	0,8	0,5	1,1
autres	NR					
par sexe :						
hommes	978	711	1242	3,4	2,5	4,3
femmes	167	82	272	0,5	0,3	0,9
par classe d'âge :						
0-4 ans	8	1	18	0,2	0,0	0,5
5-9 ans	15	3	33	0,4	0,1	0,9
10-14 ans	45	9	100	1,2	0,2	2,6
15-19 ans	190	90	308	4,9	2,3	7,9
20-24 ans	311	180	454	3,9	2,3	5,7
25-29 ans	468	309	631	2,2	1,5	3,0
30-54 ans	76	44	111	1,3	0,8	1,9
55-64 ans	18	0	50	0,4	0,0	1,0
65-74 ans	13	0	36	0,3	0,0	0,8
75 ans et +	8	1	18	0,2	0,0	0,5

Blessés avec lésion IIS3+, aux membres inférieurs

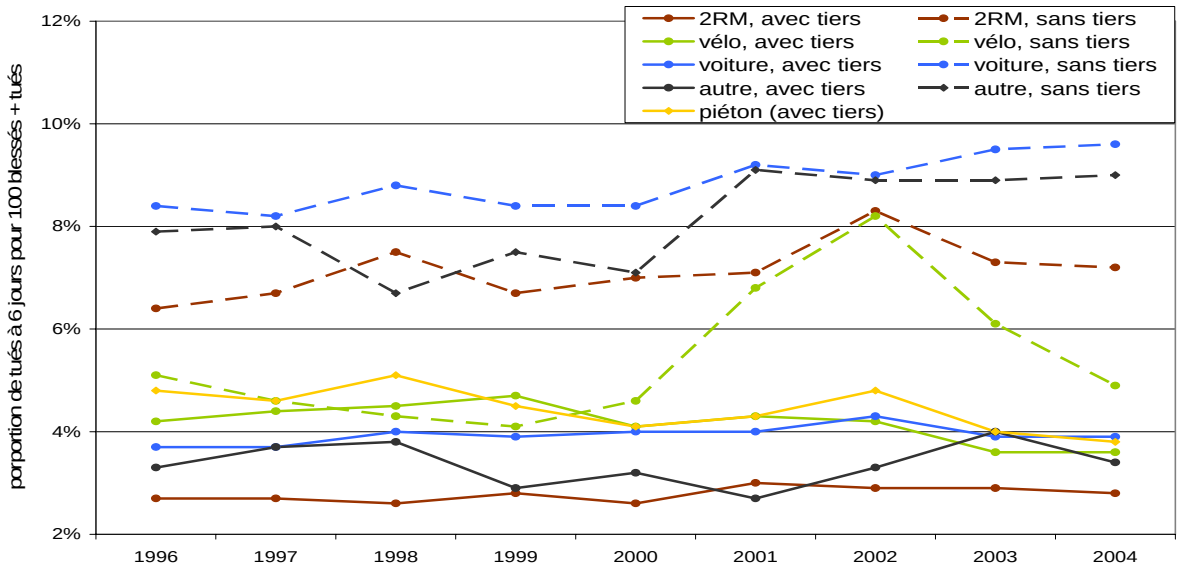
Tableau 122 : effectifs et incidences, moyennes annuelles sur 1996-2004, d'après la projection, blessés avec lésion IIS3+ aux membres inférieurs

	effectif	int. de confiance à 95%		incidence	int. de confiance à 95%	
ensemble	1821	1411	2173	3,1	2,4	3,7
par type d'usager :						
piétons	156	90	228	0,3	0,2	0,4
cyclistes	41	0	124	0,1	0,0	0,2
usagers de 2RM	694	459	920	1,2	0,8	1,6
automobilistes	884	610	1175	1,5	1,0	2,0
autres	47	0	118	0,1	0,0	0,2
par sexe :						
hommes	1462	1122	1766	5,1	3,9	6,1
femmes	359	213	527	1,2	0,7	1,7
par classe d'âge :						
0-4 ans	7	0	16	0,2	0,0	0,4
5-9 ans	11	0	25	0,3	0,0	0,7
10-14 ans	23	0	51	0,6	0,0	1,3
15-19 ans	215	96	346	5,5	2,4	8,9
20-24 ans	629	415	836	7,9	5,2	10,5
25-29 ans	701	483	897	3,3	2,3	4,3
30-54 ans	109	74	143	1,9	1,3	2,5
55-64 ans	69	25	126	1,3	0,5	2,4
65-74 ans	56	20	102	1,3	0,5	2,4
75 ans et +	7	0	16	0,2	0,0	0,4

Létalité selon type d'usager et tiers

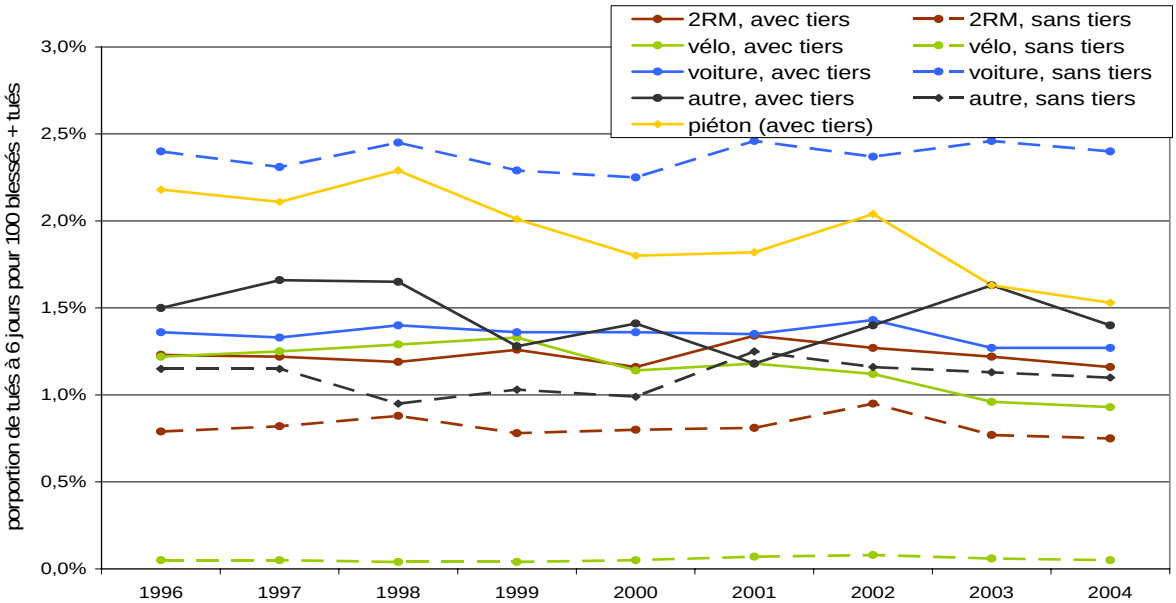
D'après les données des forces de l'ordre, la létalité la plus forte est celles des automobilistes, des usagers de 2RM et usagers « autre » dans les accidents sans tiers. A type d'usager donné, les accidents sans tiers seraient plus létaux que les accidents avec tiers. Parmi ceux-ci, les piétons ont une forte létalité et les usagers de 2RM la plus faible.
La létalité des accidents sans tiers semble augmenter. La létalité des cyclistes blessés dans les accidents sans tiers semble connaitre une forte hausse et rejoindre le taux de létalité des usagers de 2RM, dans des accidents sans tiers.

Figure 1 : Létalité selon type d'usager et tiers, source=données des forces de l'ordre



D'après les résultats de la projection, la létalité la plus élevée est celle des automobilistes dans un accident sans tiers et celle des piétons. La létalité la plus faible est celle des cyclistes dans un accident sans tiers.
La létalité des piétons semble diminuer. La létalité des autres usagers semble stable.

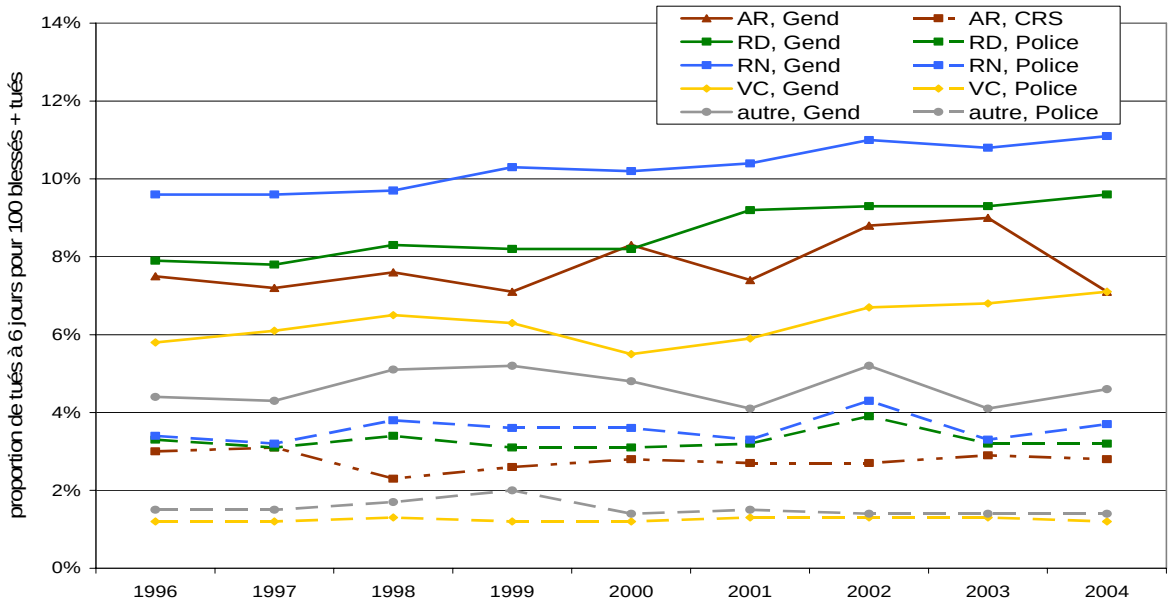
Figure 2 : Létalité selon type d'usager et tiers, source=résultats de la projection



Létalité selon le réseau

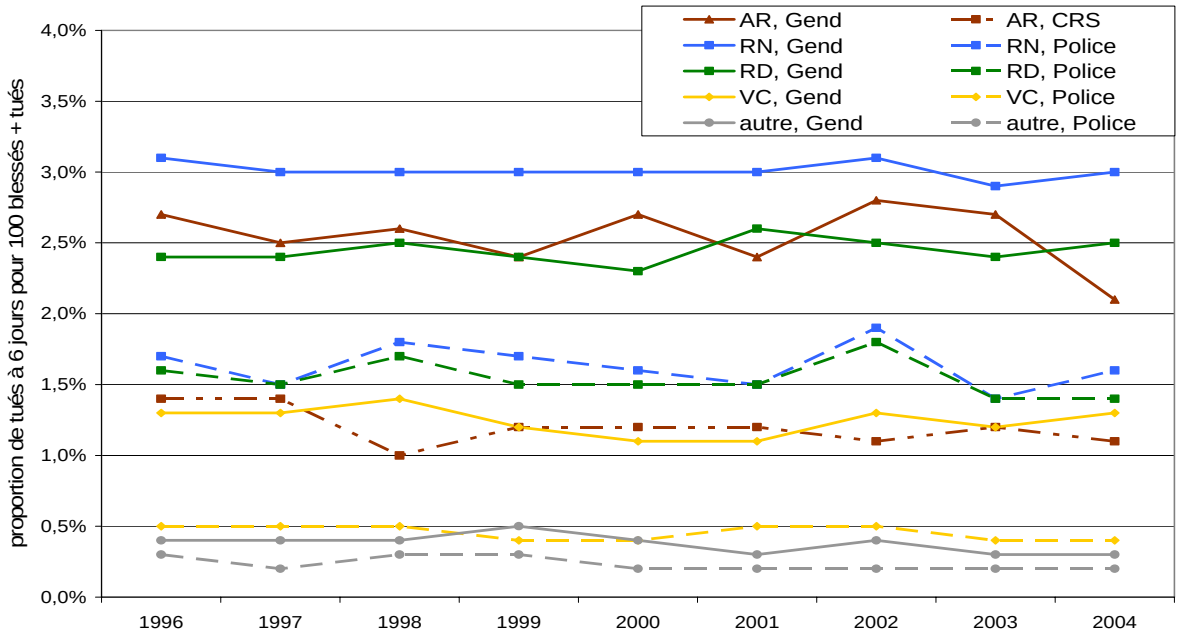
Rappelons que le type de force de l'ordre est surtout intéressant ici comme indicateur du milieu urbain ou rural. Grossièrement, les zones contrôlées par la gendarmerie sont des zones rurales, et les zones contrôlées par les CRS ou la police sont des zones urbaines.
D'après les données des forces de l'ordre, la létalité est plus élevée en rural qu'en urbain, et dans l'ordre de létalité décroissante : routes nationales, départementales, autoroutes. Il semblerait que la létalité augmente en rural.

Figure 3 : Létalité, selon le réseau, source=données des forces de l'ordre



D'après les résultats de la projection, la létalité des accidents sur voies communales, en rural est abaissée au niveau des réseaux majeurs en urbain, et la létalité sur réseau « autre » (parking, hors réseau) en rural est abaissée au niveau du réseau « autre » en urbain. La létalité n'augmente pas, quel que soit le réseau.

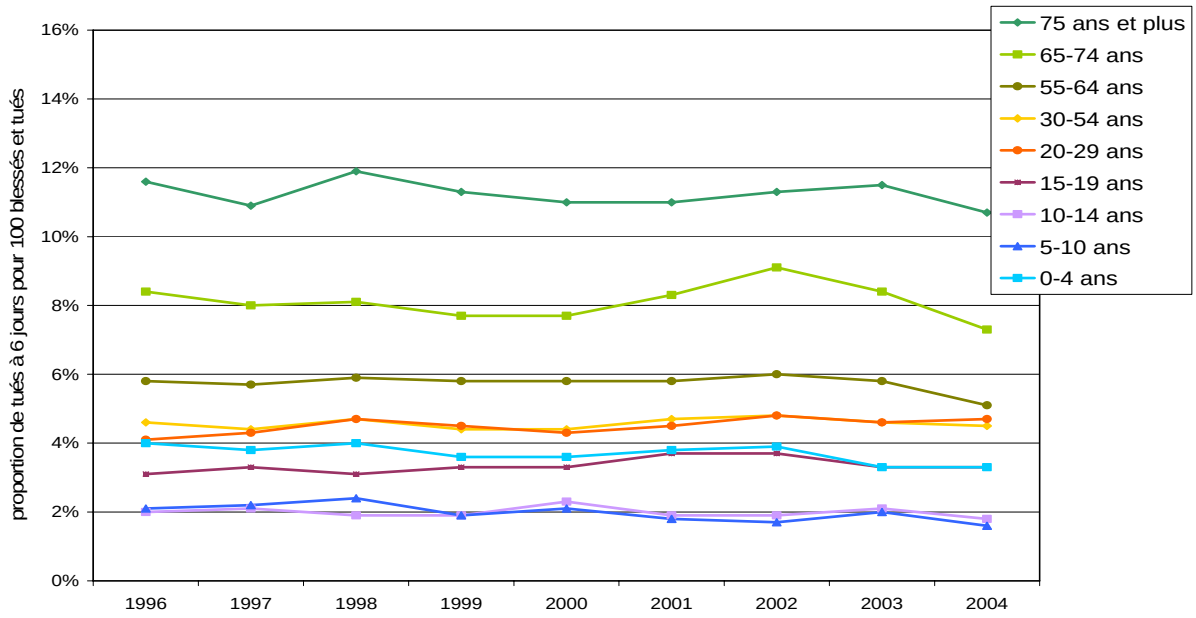
Figure 4 : Létalité, selon le réseau, source=résultats de la projection



Létalité selon l'âge

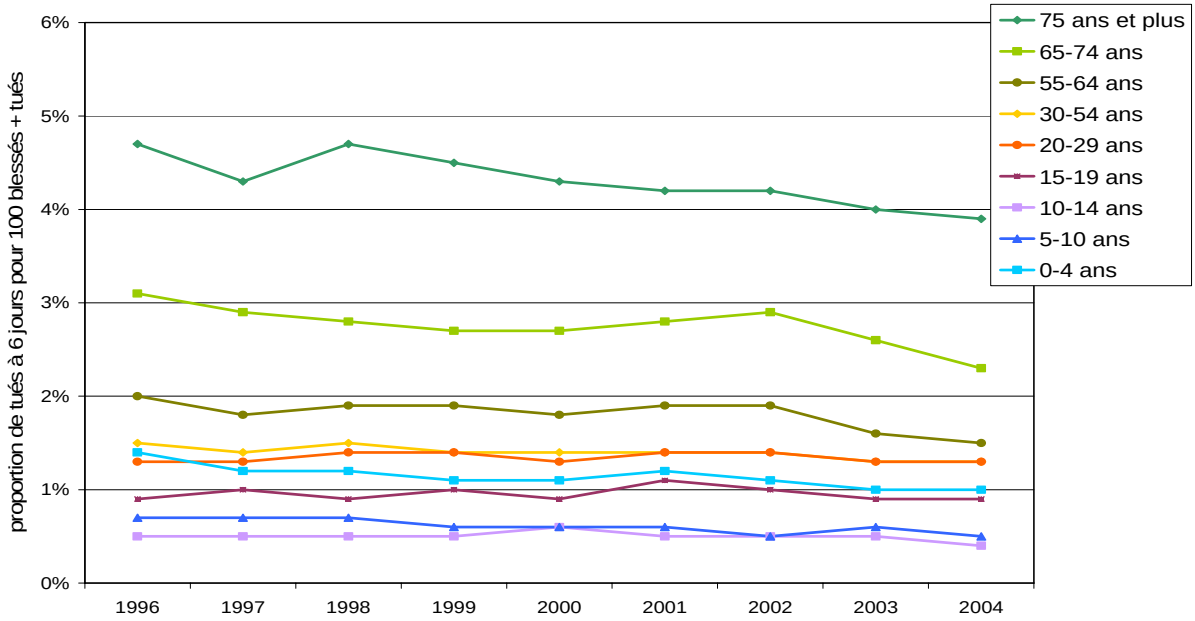
D'après les données des forces de l'ordre, la létalité en fonction de l'âge montre un gradient à la hausse avec l'âge. La seule exception est la classe d'âge des 0-4 ans dont la létalité est aussi élevée que celle des 15-19 ans.

Figure 5 : Létalité, selon l'âge, source=données des forces de l'ordre



La létalité basée sur les résultats de la projection donne la même image. La seule différence serait une baisse de la létalité chez les personnes de 75 ans et plus.

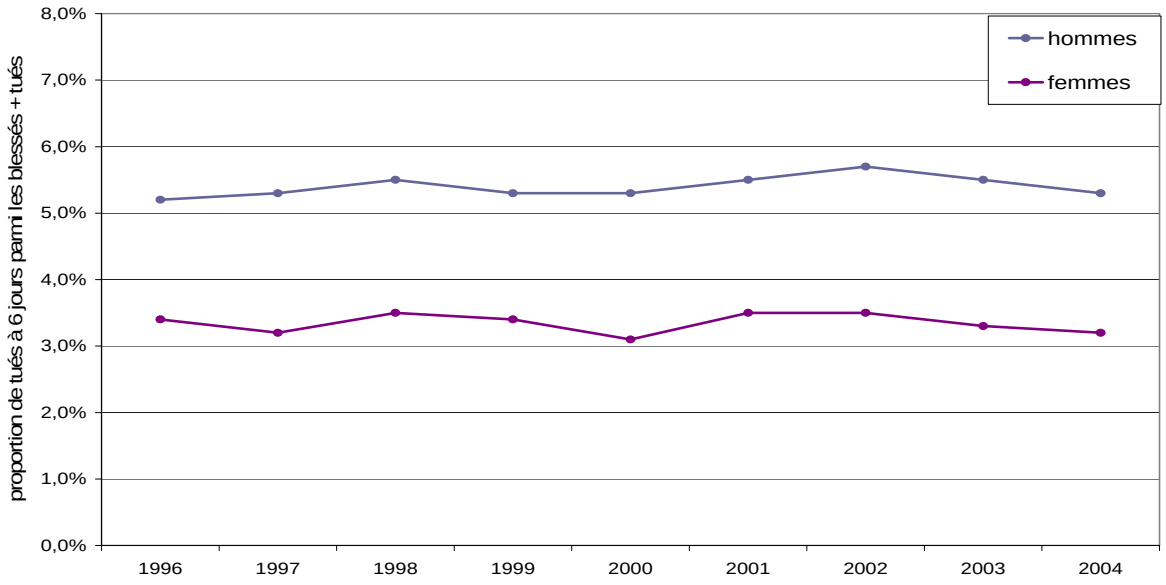
Figure 6 : Létalité, selon l'âge, source=résultats de la projection



Létalité selon le sexe

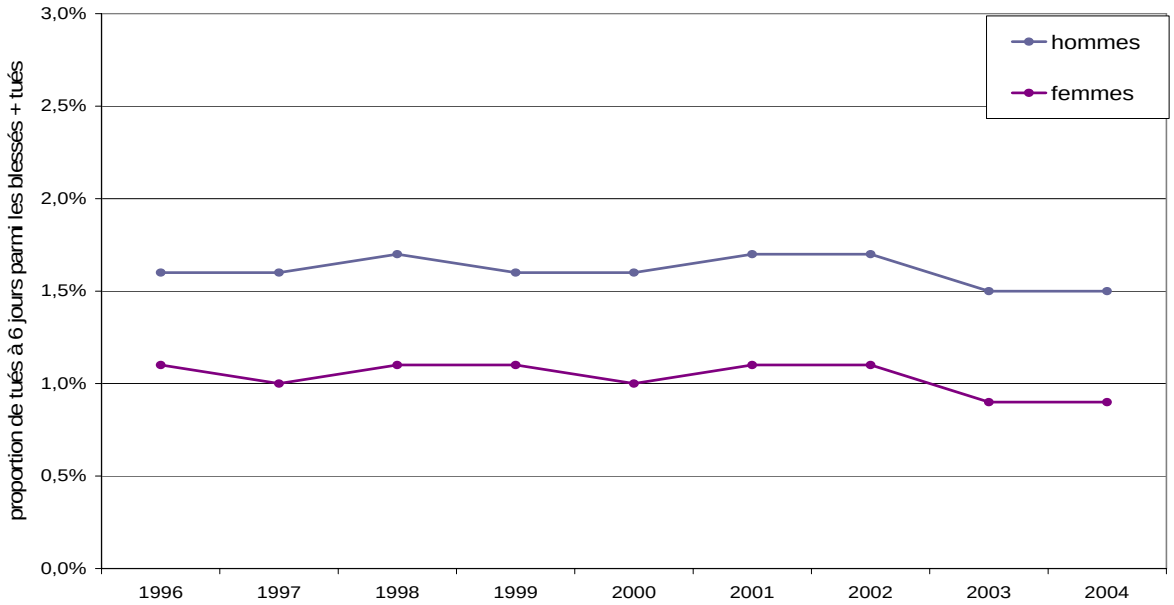
D'après les données des forces de l'ordre, la létalité chez les hommes est plus élevée que chez les femmes et le ratio entre les deux (1,62 en moyenne) semble stable sur la période.

Figure 7 : Létalité, selon le sexe, source=données des forces de l'ordre



La létalité basée sur les résultats de la projection indique la même chose : une sur-létalité des hommes par rapport aux femmes, avec un ratio qui semble constant (1,56 en moyenne)

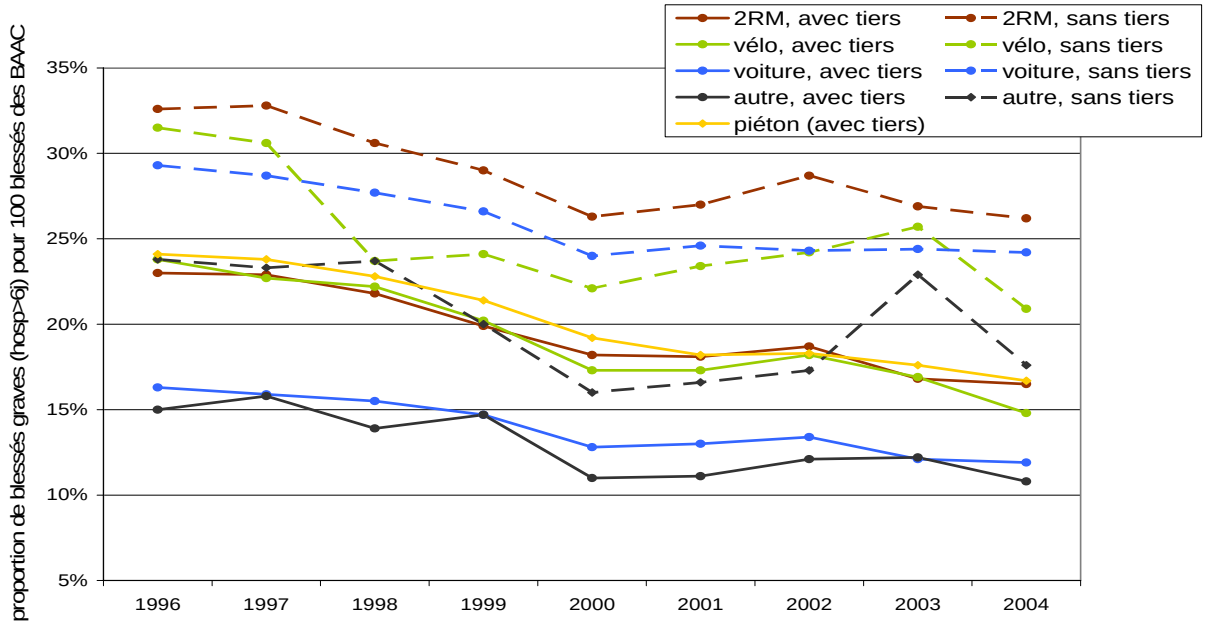
Figure 8 : Létalité, selon le sexe, source=résultats de la projection



Taux de gravité selon le type d'usager et tiers

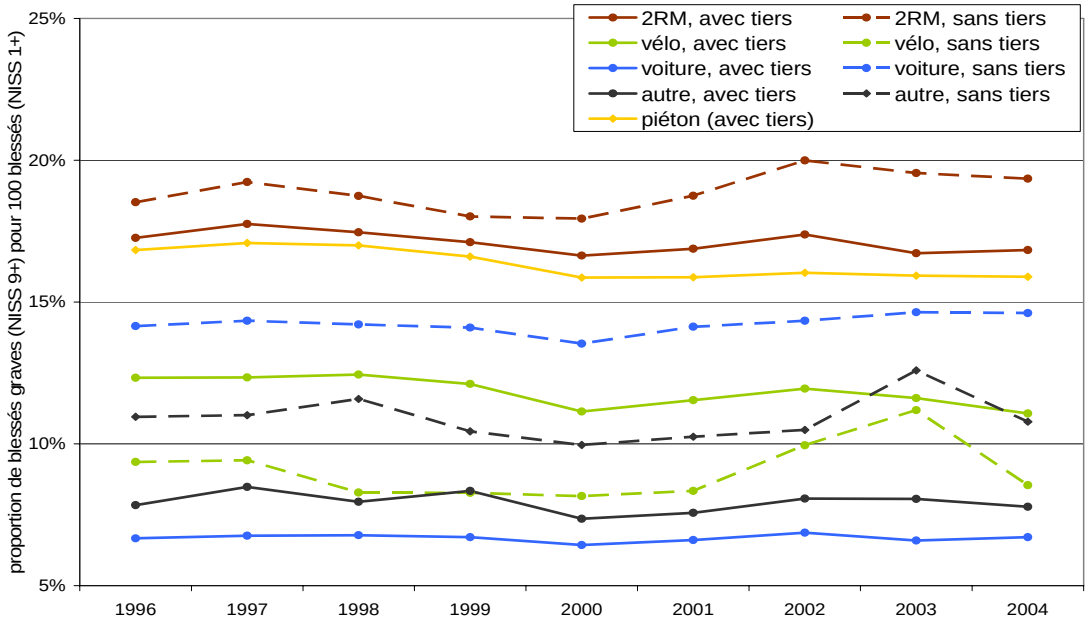
D'après les données des forces de l'ordre, la gravité baisserait pour tous les types d'usager, qu'il y ait collision avec un tiers ou non. A type d'usager donné, les configurations sans tiers seraient toujours plus graves que les configurations avec tiers.

Figure 9 : Taux de gravité selon type d'usager et tiers, source=données des forces de l'ordre



Les taux de gravité basés sur les résultats de la projection sont plutôt stables. A type d'usager donné, les configurations sans tiers sont plus graves que les configurations avec tiers, sauf pour les cyclistes. Les cyclistes blessés dans un accident sans tiers sont le plus souvent des chutes de vélo, peu graves car la vitesse propre du cycliste n'est pas très élevée. Les usagers de 2RM, dans un accident avec ou sans tiers, et les piétons apparaissent nettement comme ayant les plus forts taux de gravité.

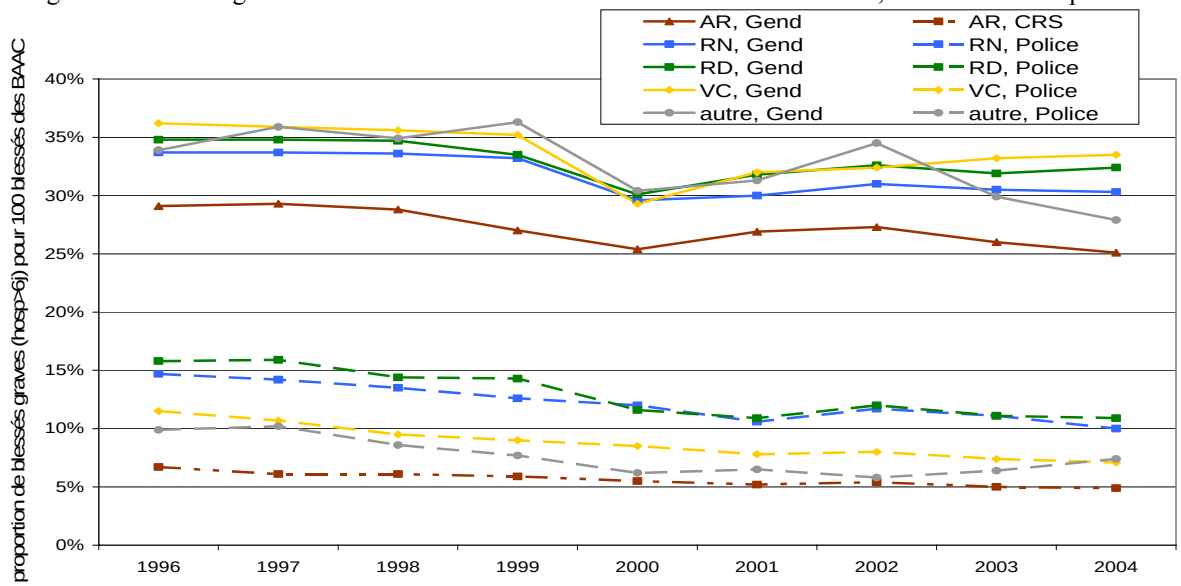
Figure 10 : Taux de gravité selon type d'usager et tiers, source=résultats de la projection



Taux de gravité par réseau

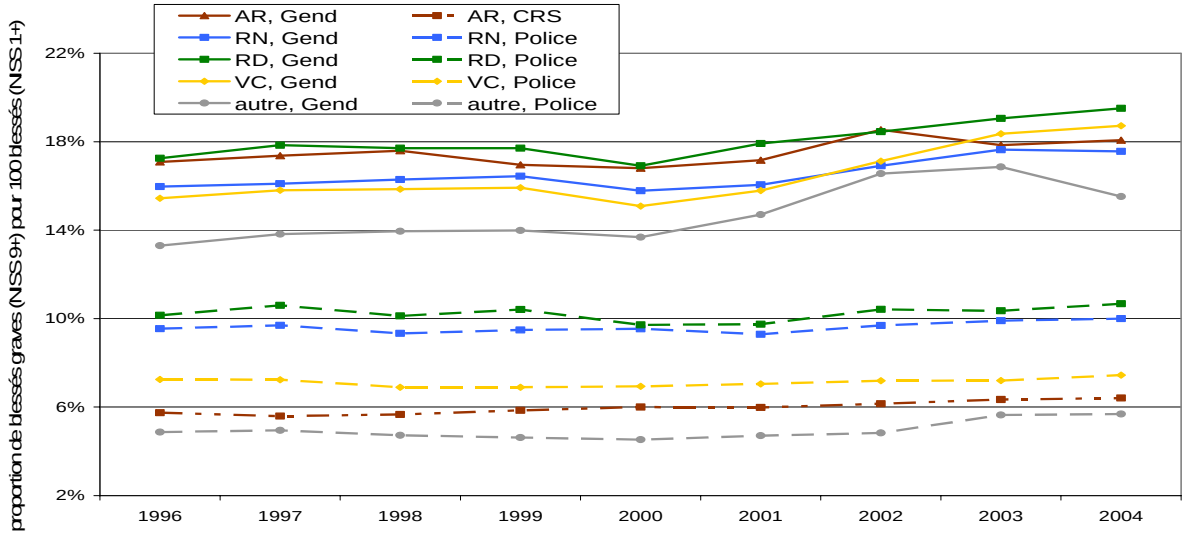
D'après les données des forces de l'ordre, les accidents en zones rurales apparaissent comme plus graves que ceux en zone urbaine ; en zone rurale, les accidents sur autoroute seraient moins graves que sur les autres réseaux. En zone urbaine, les accidents sur routes nationales et départementales seraient plus graves que ceux sur voie communales, réseau autre et autoroute.

Figure 11 : taux de gravité selon le réseau et la force de l'ordre couvrant la zone, source=données policières



Le taux de gravité basé sur les résultats de la projection confirme que les accidents sont plus graves en zone rurale qu'en zone urbaine.
En zone rurale, le taux de gravité du réseau « autre » se distingue par une moindre gravité que les autres réseaux. En zone urbaine, les accidents sur routes départementales et nationales apparaissent plus nettement comme plus graves que ceux sur les autres réseaux.
Le taux de gravité en zone urbaine semble stable. En revanche, en zone rurale il semblerait augmenter sur la période 2001-2004.

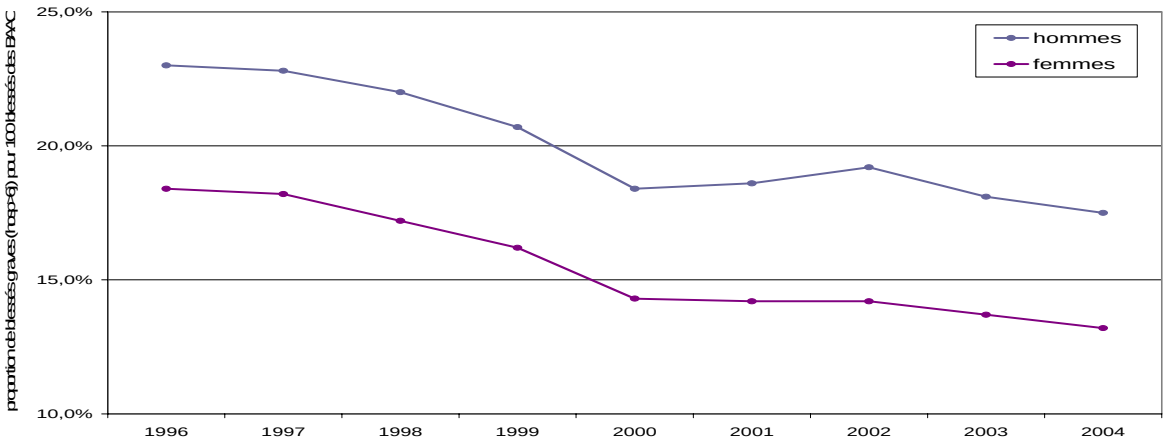
Figure 12 : Taux de gravité, par réseau, source=résultats de la projection



Taux de gravité selon le sexe

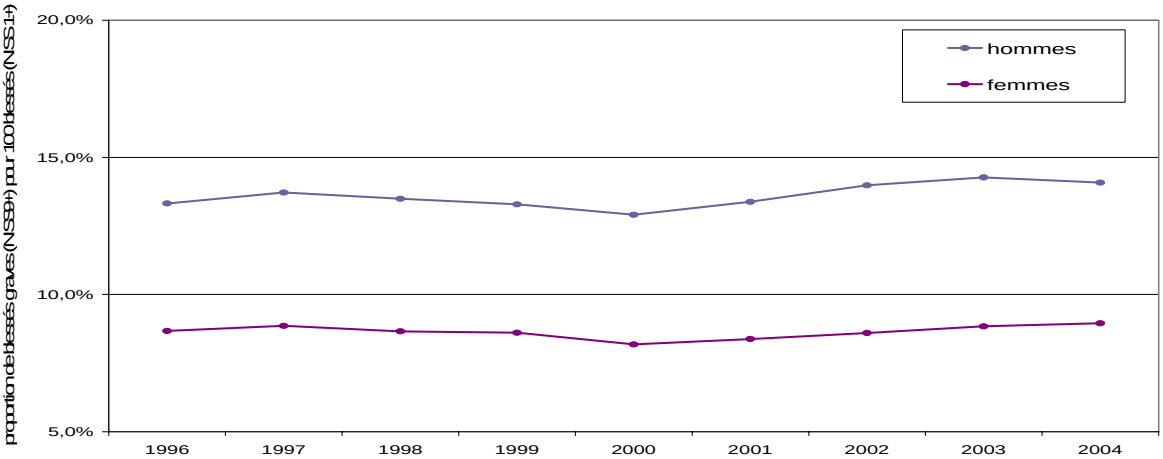
D'après les données des forces de l'ordre, le taux de gravité chez les hommes est plus élevé que chez les femmes, le sex-ratio est de 1,3 ; les évolutions sont à la baisse:

Figure 13 : taux de gravité selon le sexe, source=données des forces de l'ordre



D'après les résultats de la projection, le taux de gravité chez les hommes est aussi plus élevé que chez les femmes (sex-ratio=1,6). En revanche, ils sont plutôt stables.

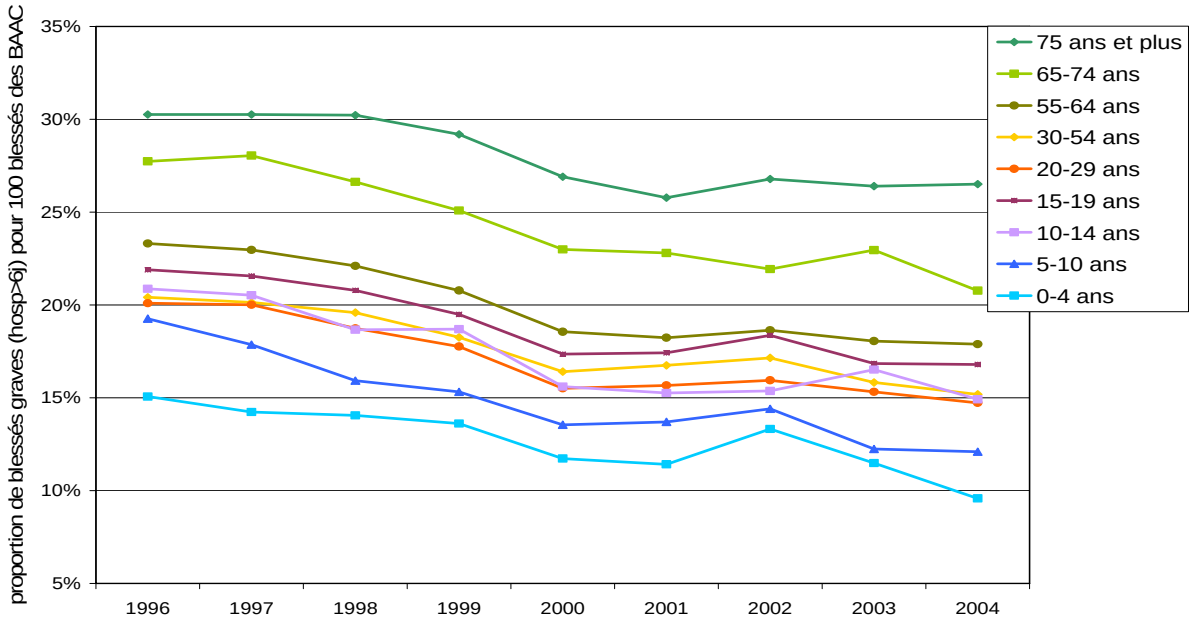
Figure 14 : taux de gravité selon le sexe, source= résultats de la projection



Taux de gravité selon l'âge

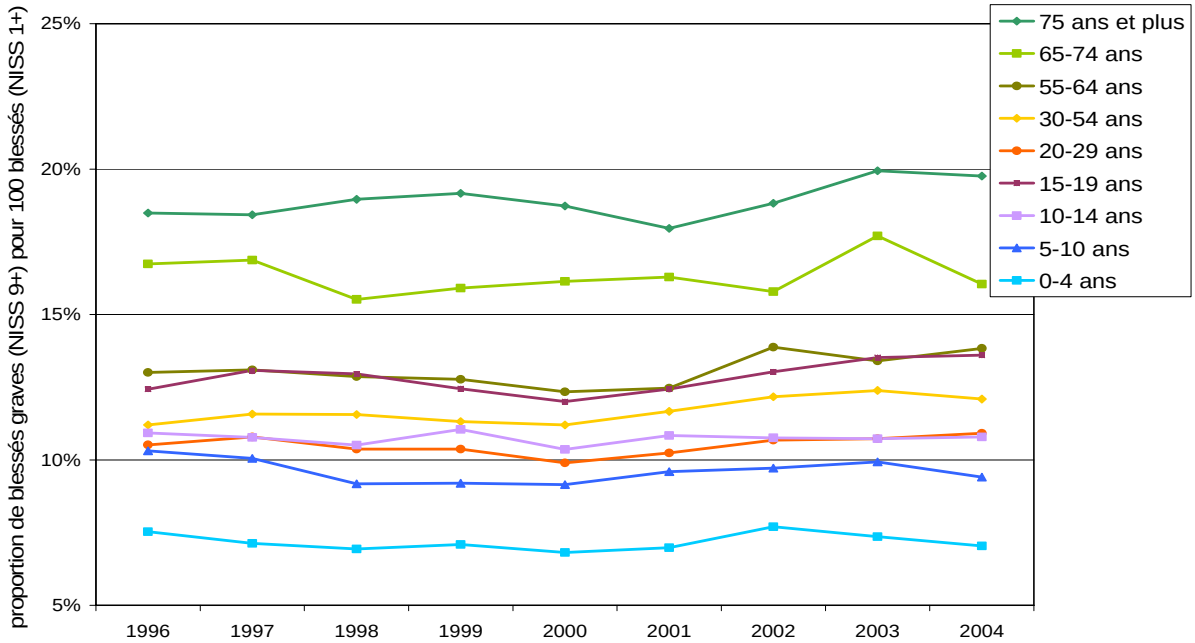
D'après les données des forces de l'ordre, le taux de gravité baisse dans toutes les classes d'âge. Il y a un gradient : plus l'âge des blessés est élevé, plus la proportion de blessés graves est élevée. Seule exception : les 15-19 ans ont un taux de gravité aussi élevé que les 30-64 ans.

Figure 15 : Taux de gravité, selon l'âge, source=données des forces de l'ordre



D'après les résultats de la projection, le taux de gravité ne diminue pas ; il semblerait augmenter dans certaines classes d'âge, notamment chez les 20-64 ans. L'ordre des classes d'âge en fonction du taux de gravité est le même que celui basé sur les données des forces de l'ordre.

Figure 16 : Taux de gravité, selon l'âge, source=résultats de la projection



Article soumis pour publication (estimation du bilan national)

Actual incidences of road casualties, and their injury severity, modelled from police and hospital data, France

Emmanuelle Amoros, Jean-Louis Martin, Sylviane Lafont, Bernard Laumon
Transport, Occupational and Environmental Epidemiology Research and Surveillance Unit (UMRESTTE), joint unit between three institutes:
French National Institute for Transport and Safety Research (INRETS), Bron, F-69500, France
University of Lyon, University of Lyon 1 (UCBL), Lyon, F-69008, France
French Institute for Public Health Surveillance (InVS), Lyon, F-69008, France

ABSTRACT
(245 words)
Background: Nationwide road casualty figures usually come from police data. In France, as in many countries, fatalities are well reported but non-fatal casualties are not. Their under-reporting is rather large and it is biased. Valid estimates are needed.
Methods: Using the capture-recapture method on police data and on a road trauma registry covering a large county of 1.6 million inhabitants, we estimate police under-reporting correction factors that account for unregistered casualties. These correction factors are then applied to the nationwide police data, with standardisation on under-reporting bias factors.
Results: In 2004, whereas the police report 108 727 non-fatally injured, the estimation yields 400 200. Over the 1996-2004 study period, the average annual estimated incidence is 871/100 000 for all injured (3.4 times the police incidence), 232/100 000 for hospitalised, 103/100 000 for seriously injured (2.2 times the police incidence), and 12.6/100 000 for casualties with long-term major impairment. The incidence of seriously injured is 11.3/100 000 for pedestrians, 9.5/100 000 for cyclists, 36.3/100 000 for motorised two-wheel users and 42.5/100 000 for car users.
Conclusions: The estimated incidences are much higher than the police-based ones. This changes the scale of the road injuries issue. The risk of suffering a major impairment from a road crash is equal to the risk of being killed. Motorised two-wheel users experience a large burden of traffic casualties, much larger than indicated by police data. The approach used can be reproduced in other countries, if an additional medical registration exists.

Keywords: traffic accidents, injuries, impairments, incidences, capture-recapture

CONFIDENTIEL

INTRODUCTION

In most countries, data on road injuries are provided by the police. While fatalities are usually well reported, non-fatal casualties are not. There is under-reporting, both to and from the police. It can be quite large, and it is biased, mostly on injury severity, mode of transport and number of vehicles involved (single vs. multiple)[1-4]. Another source of registration, namely hospital records, can sometimes be used. When both hospital and police data are available, it is possible to obtain a good estimation of the actual number of non-fatal road casualties and of their injury pattern. This is needed to better define prevention programs[5].

In France, there is both police data at the nationwide level and a hospital based registry in a large county. At the county level, we confront the two registrations, model the probability of registration by accounting for those not reported at all, and estimate police under-reporting correction factors. Applying these correction factors to the nationwide police data, standardising on the relevant covariates, we obtain a global picture of the health burden of road accidents at national scale. The underlying assumption is the homogeneity throughout France of police practices of reporting and classifying road casualties, within specific groups.

MATERIALS AND METHODS

The projection covers all non-fatal road casualties (fatalities are not corrected as they are well reported). At the nationwide level, only the police data is available, but at the Rhône county level, both the police data and the road trauma registry are available. This county has 1.6 million inhabitants, with a large city (Lyon), its suburbs and a rural area. Over the 1996-2004 study period, 83 447 non-fatally injured are reported by the registry, 36 329 by the Rhône police file and 1 353 600 by the nationwide police file.

Police road crash data

The French police have to write a report for every road crash causing at least one casualty. A road crash is officially defined as a crash involving at least one vehicle (motorised ones and bicycles). The police report should include everyone involved in the crash, and classify them as non-injured, slightly injured, seriously injured or killed. Slightly injured are outpatients and inpatients with a hospital stay of at most six days. Seriously injured are those requiring a longer hospital stay. Killed are those who died within six days of the crash.

The Rhône road trauma registry

A road trauma registry[6] has been in operation since 1995 and has been certified by the relevant French authority. It covers all casualties from road crashes in the Rhône county who seek medical care in health facilities. All health care facilities (from public and private hospitals) in the county and its surrounding area which may receive crash victims participate: about 220 health units ranging from pre-hospital emergency care, emergency departments, intensive care units, surgery units... to rehabilitation departments. Injury assessment is based on all diagnoses established in the different health services a casualty may have visited. Diagnoses are coded with the Abbreviated Injury Scale (AIS)[7], which includes a severity score, ranging from 1 (minor) to 6 (beyond treatment).

Matched casualties between the two registrations

At the Rhône level, casualties recorded in both the police and the registry were identified thanks to a semi-automated probabilistic record linkage[8]. Linking variables were date and time of the crash, crash location (town/village and possibly street/road), mode of transport, year of birth, month of birth and gender of the casualty.

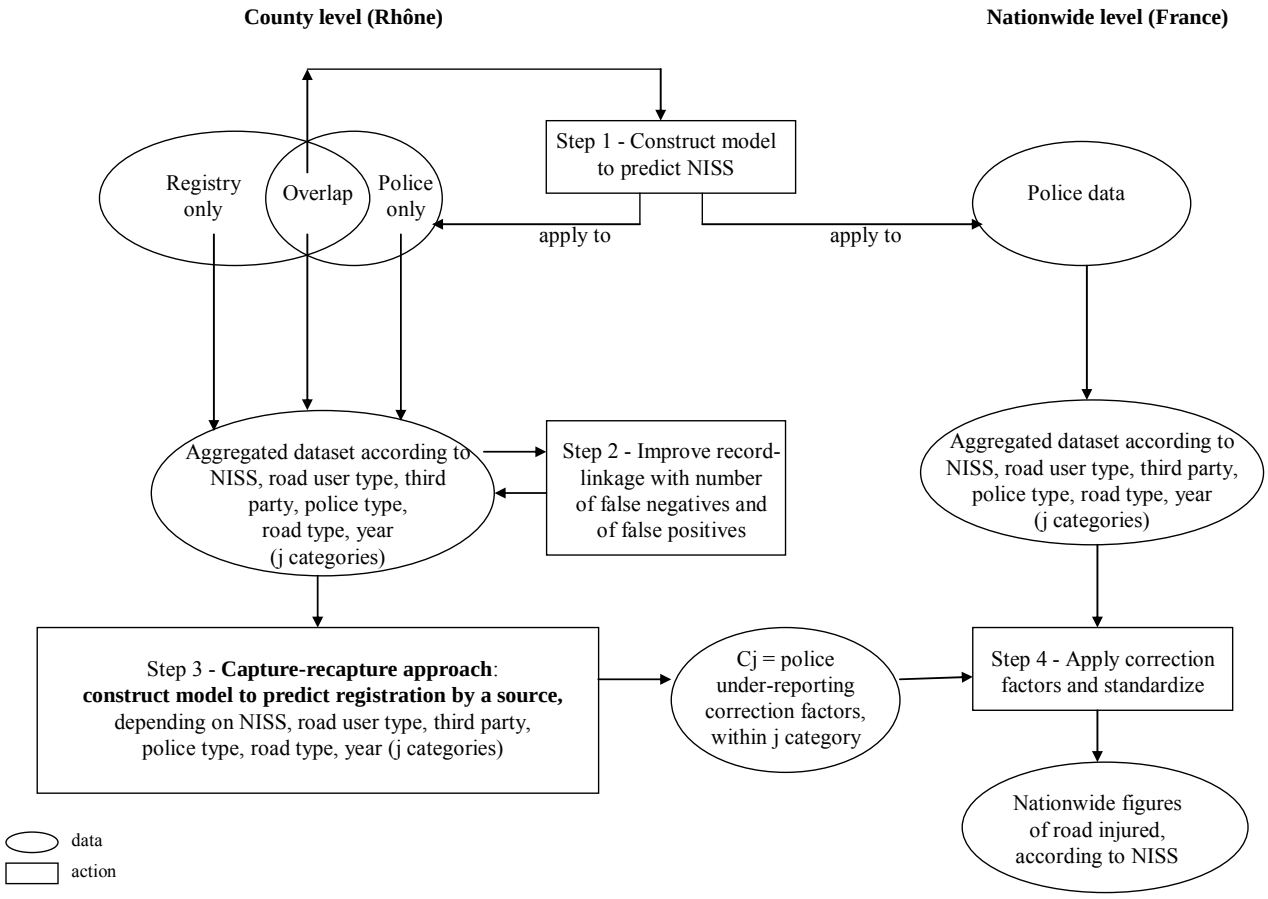
CONFIDENTIEL

Health outcomes

The outcomes of interest are all injured, hospitalised, seriously injured, and injured with major impairment, based on medical classifications (except for the hospitalised). The casualty's injury severity is defined with the New Injury Severity Score (NISS)[9]. It is defined as the sum of squares of the AIS severity scores of a casualty's three most severe injuries, regardless of body region. The NISS scores are grouped into three categories: NISS 1-3 (minor severity), NISS 4-8 (moderate), and NISS 9+ (serious). Casualties with long-term major impairment(s) are defined with the Injury Impairment Scale (IIS)[10], restricted to those with any injury of IIS 3+. These correspond mostly to serious head injuries (ex: intraventricular haemorrhage of cerebrum), spine injuries (contusion, laceration) and lower extremities injuries (ex: knee dislocation). Information on these health outcomes is available on registry casualties; for police-reported casualties they are estimated by prediction models (step 1 below).

The whole estimation procedure is presented for the NISS (here and in figure 1); it is repeated for: hospitalised, any IIS3+ injury, and any head, spine or lower extremities IIS3+ injury separately.

figure 1: flow chart of the whole estimation procedure



Predicting the injury severity (NISS) of police-reported casualties (step 1)

The police severity classification is not satisfactory but it is correlated with the NISS[11], and hence useful for its prediction. The prediction model is constructed on non-fatal casualties identified in both the police and registry files, for which both the NISS and police severity are available (22 704 matched casualties). The three-level NISS variable is ordinal, and we model it with a cumulative logit model. The proportional odds assumption being rejected ($p<0.01$) this corresponds to two binomial models, which are fitted with SAS software, logistic procedure. The variables retained in the predictive model are: interaction between police severity classification and type of police, road user type, interaction between road user type and third party (yes/no), road user type of third party, age, gender, road type,

urban/rural and calendar year. For the prediction of the other health outcomes, which are binary (yes/no), we merely use a logistic regression, with the same covariates.
The three-level NISS prediction model is applied to the Rhône non-fatal casualties reported only by the police (figure 1), so that the under-reporting model (which must be based on all observations from the two registrations in the Rhône county) can include the NISS. The three-level NISS prediction model is applied to the nationwide police data (figure 1), so that in step 4 the under-reporting correction factors, which are defined according to NISS-level (among others), can be applied to these data.

Improving the record linkage (step 2)
Before modeling under-reporting, we need to have a highly reliable picture of how many non-fatal casualties are recorded by the police only, by the registry only, and by both. As a result of the record-linkage, there might be false positives: linked pairs corresponding to two distinct casualties, and false negatives: non-linked pairs corresponding to the same casualty. We estimate their numbers[12-14] and use them to improve the result of the record-linkage.

Estimating. police under-reporting correction factors, with capture-recapture (step 3)
The capture-recapture approach has been used in epidemiology[15-18], including in road crash injuries[19-24]. The major feature of capture-recapture is to quantify those unobserved i.e. not recorded by any registration. The validity of the method is based on four key assumptions[15, 16] ; they are discussed later.
A capture-recapture model is fitted to the aggregated Rhône police and registry data (96 238 casualties). A standard result is the total number of road casualties in the Rhône county[12]. Here we aim to estimate under-reporting correction factors, and these can be directly obtained from the fitted model. The model formulates the probabilities of being reported by each source, conditional on the observed data. It enables the estimation of these probabilities, unconditional on the observed data. The inverse of these provides the under-reporting correction factors. The dependent variable is categorical nominal (police only, registry only, both), modelled by a generalised logit[25], using SAS logistic procedure.
One assumption of capture-recapture is "homogeneity of capture": all subjects should have the same probability of being recorded by a given source[15, 16]. If this is true only within groups (e.g. by injury severity level) one should stratify on or include the corresponding covariate. The probability of being reported by the French police mostly depends on injury severity, road user type, third party (yes/no) and their interaction, road type, urban/rural area and type of police, mostly[26]. The probability of being reported by the registry mostly depends on injury severity[12]. In other words, the probability of being reported by a source is homogenous within every j category defined by the combination of these aforementioned characteristics. We hence include the corresponding covariates in the model[25], except for urban/rural, not available in the registry. Year is also included (as a quantitative variable) to allow for a possible time trend in under-reporting. Under-reporting correction factors (Cj) are estimated within every j category.

Projection to the nationwide level (step 4)
It is a projection of county-level estimates to the nationwide level, standardising on the relevant covariates. It is similar to the indirect standardisation of rates[27], usually over age and sex. The nationwide estimated number of non-fatal casualties is obtained as $E = \sum_{J=1}^{J} Cj \times O_j$, where j indexes the categories defined by the combination of the under-reporting bias factors, instead of the age and sex categories; Cj are the j-specific reference correction factors, instead of the age and sex-specific reference incidence rates, and Oj are the observed police counts of non-fatal casualties nationwide, instead of the size or person-years of the study population.

Study outcomes
Counts and incidences per 100 000 inhabitants are displayed. Population figures are provided by INSEE, the French national institute for statistics and economics studies. Police-based figures, including fatalities, are provided for comparison. Trends are the proportion of change between 1996 and 2004. Confidence intervals (95%) are 2.5 and 97.5 percentiles from a non–parametric bootstrap[28] over the whole projection process, with 2000 iterations.

CONFIDENTIEL

RESULTS

Over the study period 1996-2004, for the whole country, the police-based average annual incidence for all injured is 255/100 000 whereas the estimated incidence is 871/100 000 (95% CI: 828, 894), i.e. 3.4 times higher. The police-based incidence of seriously injured is 47/100 000 (those requiring more than six days of hospital stay) whereas the estimated incidence of NISS 9+ casualties is 103/100 000 (97, 106), i.e. 2.2 times higher. As for the hospitalised, the police-based estimated incidence is 104/100 000 (101, 107) and the projected one is 232/100 000 (221, 239). Table 1 displays annual incidences. The estimated incidences display a lesser decrease than the police-based incidences, especially for seriously injured casualties: a decrease by less than a third instead of by half.

The average annual incidence of all injured is 69/100 000 (66, 71) for pedestrians, 94/100 000 (78, 112) for cyclists, 203/100 000 (192, 210) for motorised two-wheel users and 469/100 000 (450, 479) for car occupants. The number of injured cyclists is slightly higher than the number of injured pedestrians. The average annual incidence of seriously injured is 11.3/100 000 (10.5, 12.0) for pedestrians, 9.5/100 000 (7.2, 11.8) for cyclists, 36.3/100 000 (33.5, 38.1) for motorised two-wheel users and 42.5/100 000 (39.8, 44.4) for car occupants. The number of seriously injured is in the same order of magnitude for cyclists and pedestrians on the one hand, and for motorised two-wheel users and car occupants on the other hand. Table 2 displays annual estimates. Regarding time trends, motorised two-wheel users experience a much smaller decrease of their number of casualties than the other road users.

Casualties with major impairments (IIS 3+) are quantified in table 3. The IIS 3+ injured body region is principally the head (53.3%), followed by the lower extremities (24.3%), and the spine (15.8%). Cyclists and pedestrians are particularly prone to head IIS 3+ injuries. Motorised two-wheel users have a high frequency of spine and lower extremities IIS3+ injuries.

CONFIDENTIEL

Table 123. Counts and incidences of road casualties per 100 000 inhabitants, police-based and estimated, France

	all (non-fatally) injured					seriously injured					hospitalised					killed	
	police-based		estimated			police-based [a]		estimated [b]			police-based [c]		estimated			police-based	
year	count	incidence	count	incidence	95% CI	count	incidence	count	incidence	95% CI	count	incidence	count	incidence	95% C	count	incider
1996	170 117	293.2	554 700	955.9	905.2, 993.1	36 204	62.4	64 500	111.1	104.0, 116.6	70 878	122.1	151 400	261.0	247.5, 270.8	8 080	13.9
1997	169 577	291.3	561 600	964.8	915.9, 997.8	35 716	61.4	67 000	115.2	108.3, 119.9	71 055	122.1	154 400	265.3	252.3, 274.2	7 990	13.7
1998	168 535	288.6	566 300	969.8	923.2, 1000.4	33 977	58.2	66 400	113.6	107.5, 117.7	69 613	119.2	152 100	260.5	248.7, 268.1	8 437	14.4
1999	167 572	285.7	562 900	959.6	917.0, 985.8	31 851	54.3	65 200	111.2	105.7, 114.5	68 216	116.3	148 900	253.9	243.0, 260.1	8 029	13.7
2000	162 117	274.7	548 900	930.1	891.7, 950.5	27 407	46.4	61 900	104.9	99.8, 107.8	64 723	109.7	141 800	240.3	230.4, 245.6	7 643	13.0
2001	153 945	259.2	524 000	882.3	846.4, 900.3	26 192	44.1	61 100	102.9	97.8, 105.7	61 719	103.9	136 900	230.4	220.8, 235.5	7 720	13.0
2002	137 839	230.6	484 900	811.1	774.5, 829.5	24 091	40.3	59 000	98.7	93.3, 101.9	56 066	93.8	129 000	215.8	205.5, 221.4	7 242	12.1
2003	115 170	191.5	425 400	707.2	671.6, 725.6	19 110	31.8	53 000	88.1	82.2, 91.4	46 804	77.8	113 600	188.8	178.8, 194.5	5 724	9.5
2004	108 727	179.7	400 200	661.3	627.7, 677.5	17 435	28.8	49 500	81.7	76.2, 85.3	43 779	72.3	104 600	172.8	163.3, 178.4	5 232	8.6
trend		-38.7%		-30.8%			-53.8%		-26.5%			-40.8%		-33.8%			-37.9%

a: based also on police definition (requiring more then 6 days of hospital stay)

b: defined as NISS 9+

c: estimated. it corresponds to step 1 in figure 1 applied to the nationwide police data

CONFIDENTIEL

Table 124. Estimated incidences of road casualties per 100 000 inhabitants, according to road user type, France

year	all injured (NISS 1+)				seriously injured (NISS 9+)			
	pedestrians	cyclists	M2W users[a]	car occupants	pedestrians	cyclists	M2W users[a]	car occupants
1996	76.3	113.7	196.4	528.5	12.8	11.8	34.9	47.7
1997	73.9	118.4	207.6	525.0	12.6	12.4	38.1	48.1
1998	72.2	108.9	210.2	537.0	12.3	10.6	37.8	48.9
1999	73.3	101.4	215.4	527.3	12.2	9.8	37.6	47.6
2000	73.3	91.8	217.5	512.1	11.6	8.5	37.3	44.2
2001	70.6	78.4	211.2	489.2	11.2	7.5	37.2	43.9
2002	65.9	81.3	194.3	438.4	10.6	8.6	35.8	40.7
2003	59.0	83.9	189.5	346.1	9.4	9.5	34.0	32.0
2004	58.6	70.5	189.3	317.8	9.3	6.7	33.9	29.4
trend	-32.8%	-37.6%	-17.3%	-47.1%	-27.3%	-43.2%	-2.9%	-38.4%

a) motorised two-wheel users: include moped and motorbike users

Table 125. Road fatalities and casualties with major impairment (IIS 3+ injury), average annual estimated incidences per 100 000 inhabitants, counts and proportions, 1996-2004, France

	all road users incidence	all road users count	pedestrians count (row %)	cyclists count (row %)	M2W users[a] count (row %)	car occupants count (row %)
killed (police based) casualties	12.4	7 344	811 (11%)	256 (4%)	1 291 (18%)	4 722 (64%)
with any IIS3+ injury at body region:	12.6	7 479	726 (10%)	857 (12%)	2 249 (30%)	3 393 (45%)
head	6.8	3 993	474 (12%)	561 (14%)	946 (24%)	1 846 (46%)
spine	2.0	1 188	61 (5%)	134 (11%)	518 (44%)	475 (40%)
lower extremities	3.1	1 821	156 (9%)	41 (2%)	694 (38%)	884 (48%)

a) motorised two-wheel users: include moped and motorbike users

Incidences by age group indicate that the 20-29 and especially the 15-19 year old have very high incidences, up to 2164/100 000 for all injured, 277/100 000 for the seriously injured and 31.8/100 000 for the ones with major impairment. The ratio of incidences between the 15-19 year old and the 0-4 year old (who have the lowest incidences) is 10 for all injured, 18 for the seriously injured and 16 for the ones with major impairment.

The ratio between the male and female average incidences is 2.0 for all injured, 3.1 for the seriously injured and 4.0 for the ones with major impairment. The male average incidence reaches 1162/100 000 for all injured, 158/100 000 for the seriously injured, and 20.6/100 000 for the ones with major impairment.

CONFIDENTIEL

DISCUSSION

Limitations and strengths of the method

Capture-recapture is based on four key assumptions[15, 16]. The "homogeneity of capture" has been dealt with: covariates that are correlated with the registration probabilities are included in the model. A second assumption is the perfect identification of subjects common to the different registrations: the record-linkage is based on many informative linking variables and it has furthermore been corrected for false positives and false negatives. A third assumption is a closed population i.e. no entries or losses between two recordings. Most casualties injured in crashes in the Rhône county live in this county (89.8%) or a county next to it (6.7%)[29] so that even if slightly injured, most casualties will go to a hospital in the Rhône county or its close surroundings i.e. covered by the registry. The fourth assumption is independence between the two registrations. We think there is some positive dependence for the serious casualties[12] (the police alerting the emergency care and reciprocally). In such a case, it has been established[18] that the capture-recapture estimator (of the total number of subjects of interest) is biased downwards. The projected counts and incidences must hence be interpreted as lower limits.

There have been other attempts to estimate the actual number of road casualties. This can be done when additional registration of road casualties is available (hospital, insurance companies or self-reporting). Some have used the capture-recapture method to account for road casualties missed by all two or three registrations, but they have focused on sub-populations (children[23, 30], young adults[31], pedestrians and cyclists[20], heavy vehicles[21], a large town[22, 24] or an island[19]). Other attempts[3, 4, 32] have estimated under-reporting correction factors and sometimes nationwide incidences by further projection, but they have not considered casualties missed by all registrations, which can be far from negligible. The approach presented here which combines the two can be reproduced in other countries, if some additional registration exists, preferably a medical one to obtain injury severity, and provided that some conditions are fulfilled.

The underlying assumption of the projection is the homogeneity of police practices across France, within defined groups. In other words, the police practices of reporting and severity classifying are the same across France for any given group (e.g. seriously injured cyclists in a crash with a third party, on a county road, rural police area). This assumption is based on the centralised structure and management of each French police type (three of them in France). The groups are defined by the police under-reporting and severity classification bias factors.

The bootstrap confidence intervals of the incidences are very narrow. The problem is that they are based on observed data, even though re-sampled (by the use of the non-parametric bootstrap method), while the capture-recapture approach is based on the existence of unobserved[25] i.e. non-reported casualties.

Some of the results can be validated by comparing them with other existing estimations. For traumatic spine injuries, 2000 cases per year have been estimated in France[33], some 70% from traffic accidents, hence an incidence of 2.4/100 000. This is very close to our estimate of 2.0/100 000. For severe traumatic brain injuries, a registry in the French Aquitaine region reported[34] 642 cases in 1996, with a head injury of AIS 4 or 5, all causes. Restricted to traffic accidents (50% of all injured) and excluding early deaths (at least 18%), it goes down to 265. Our projection restricted to Aquitaine yields 244 cases (annual average over 1996-2004). These two figures are very similar.

Interpretation of findings

The estimated incidences obtained by projection are much higher than the police-based incidences. This changes the scale of the road injuries issue. These much higher figures are not so surprising since the French police under-reporting rate is 37.7%[26]. This rate is well within the average of other countries[1], implying that it is not only a French issue but one that concerns many countries.

The estimated incidences decrease less than the police-based ones, and the difference in decreases is larger for serious casualties. These decreases are coherent with the decrease in the number of road

CONFIDENTIEL

fatalities observed in France, which has been especially large since 2002, when numerous speed control cameras were announced and installed.
Estimated incidences according to age and gender give the same ranking of age and gender groups as those based on police data (not shown) since these two characteristics are not major bias factors in police reporting practices. Results confirm the higher incidence of road injuries among young people and among men[35].
The number of motorised two-wheel casualties hardly decreases; this confirms what is seen in the police-based statistics[36]. Some likely explanations are the ineffectiveness of speed cameras for motorised two-wheel users (because of difficulties in identifying them), and the increase in the fleet of motorised two-wheel vehicles[36].
The number of seriously injured (NISS 9+) motorised two-wheel users has reached the same level as seriously injured car users (whereas the number of killed motorised two-wheel users is three times lower than that of car users, and their share in the traffic is highly different). Cyclists roughly undergo the same number of injured (NISS 1+ and NISS 9+) as pedestrians. These two patterns are new. They can not be seen in the police-based figures because of the biases associated with police reporting[26].
The risk of suffering a major impairment is equal to the risk of being killed. For cyclists and motorised two-wheel users, this risk is in fact higher than the risk of being killed. This criteria should be used for prevention campaigns; the number of fatalities must not be the only criteria.
Motorised two-wheel users experience a large burden of traffic casualties, and much larger than what was known. These casualties are mostly teenagers and young adults[37]. It is crucial that prevention campaigns target these users.

Acknowledgements
Sources of funding: Ministère des Transports, Direction à la Sécurité et à la Circulation Routière, Programme National de Recherche et d'Innovation dans les Transports Terrestres (PREDIT-DSCR/INRETS n°02/70013 étude n° 6); Ministère de la Santé, Institut de Veille Sanitaire (J06-23)

We thank Aurélie Moskal, Nathalie Nallet for critical reading and the following people for having participated in the data collection and data typing, in the framework of the Association for the registry of road traffic casualties in the Rhône (ARVAC, president YN Martin) and in the framework of INRETS UMRESTTE (B. Laumon scientific consultant and A. Ndiaye, director) : Ait Idir T, Ait Si Selmi T, Alloatti D, 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, Bec JF, Bejui J, Bel JC, Bérard E, Bérard J, Bernard JC, Berthet N, Bertrand JC, Besson L, Biot B, Biot V, Blanc C, Blanchard J, Bœuf C, Boisson D, Bonjean M, Bost J, Bouchedor C, Bouletreau P, Boyer V, Breda Y, Brilland R, Bussery S, Cabet N, Caillot L, Caillot JL, Cannamela A, Caregnato B, Carre M, Catala Y, Chagnon PY, Chambost M, Chantran C, Chardon P, Charnay P, Chatelain P, Chattard S, Chavane H, Chazot G, Chettouane I, Chevreton N, Chevrillon E, Chevrillon S, Chiron M, Chotel P, Cochard P, Combe C, Contamin B, Coppard E, Cot T, Crettenet Z, Cristini A, Dal Gobbo B, De Angelis MP, Decourt L, Delfosse A, Demazière J, Deruty R, Desjardins G, Devaux J, Dohin B, Emonet A, Escarment J, Eyssette M, Fallavier L, Fanton L, Felten D, 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, Goubsky A, Granger M, Grattard P, Gueniaud PY, Guenot C, Guérin AC, Guignand M, Haddak M, Hamel D, Heckel T, Herzberg G, Jacquemard C, Javouhey E, Joffre T, Kohler R, Lablanche C, Lafont S, Lagier C, Lapierre B, Laplace MC, La Rosa C, Laurent R, Lebel M, Leblay G, Le-Xuan I, Lieutaud T, Lille R, Linné M, Lucas R, Machin B, Maello E, Malicier D, Mangola B, Marduel YN, Marie-Catherine M, Martinand G, Marty F, Mazouzi S, Messikh C, Meyer F, Meyrand S, Molard S, Morel-Chevillet E, Mioulet E, Minjaud F, Mollet C, Monnet J, Moyen B, Neidhart JP, Ngandu E, Ny S, Ould T, Paget P, Paillot JC, Paris D, Patay B, Pauget P, Peillon D, Perrin-Blondeau D, Petit P, Piton JL, Plantier M, Pornon P, Pramayon C, Quelard B, Raquin L, Rezig M, 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, Sbraire N, Scappaticci N, Schiele P, Schneider M, Simonet C, Sindou M, Soldner R, Soudain M, Stagnara J, Stamm D, Suc B, Taesch MC, Tasseau F, Tell L, Thomas M, Tilhet-Coartet S, Tissot E, Toukou JC, Trifot M, Vallee B, Vallet G, Vancuyck A, Vergnes I, Verney MP,

CONFIDENTIEL

Voiglio EJ, Vourey G, Vuillard J, Westphal M, Willemen L.

Conflicts of interest: none declared

Key points

- We provide the first French nationwide estimations of non-fatal road casualties that are corrected for under-reporting and its biases (and based on medical classifications: NISS and IIS)
- The incidence is 871/100 000 for all injured (NISS 1+) and 103/100 000 for seriously injured (NISS 9+), respectively 3.4 and 2.2 times higher than police-based incidences.
- The number of seriously injured (NISS 9+) motorised two-wheel users is now equal to the number of seriously injured car users (while their share in traffic are highly different)
- The risk of suffering a long-term major impairment (IIS 3+) is equal to the risk of being killed, on average. For cyclists and motorised two-wheel users, the risk of major impairment is higher.
- In terms of morbidity, injury prevention in France should focus on motorised two-wheel users.

References

1. Elvik R, Mysen AB. Incomplete accident reporting; meta-analysis of studies made in 13 countries. Transp Res Rec. 1999;1665:133-40.
2. Hauer E, Hakkert A. Extent and some implications of incomplete accident reporting. Transp Res Rec. 1988;1185:1-10.
3. Hvoslef H. Under-reporting of road traffic accidents recorded by the police at the international level. IRTAD special report. Norway: Public Roads Administration; 1994.
4. James HF. Under-reporting of road traffic accidents. Traffic Eng Control. 1991;32:573-83.
5. Krug EG, Sharma GK, Lozano R. The global burden of injuries. Am J Public Health. 2000 April 1, 2000;90(4):523-6.
6. Laumon B, Martin JL, Collet P, et al. A French road accident trauma registry: first results. 41st annual conference of the Association for the Advancement of Automotive Medecine; 1997 November 10-11; Orlando, Florida; 1997. p. 127-37.
7. Association for the Advancement of Automotive Medicine, The Abbreviated Injury Scale: Des Plaines, Illinois; 1990.
8. Clark DE. Practical introduction to record linkage for injury research. Inj Prev. 2004 Jun;10(3):186-91.
9. Osler T, Baker S, Long W. A modification of the Injury Severity Score that both improves accuracy and simplifies scoring. J Trauma. 1997;43(6):922-6.
10. Association for the Advancement of Automotive Medicine. Injury
Impairment Scale. Des Plaines, IL: AAAM; 1994.
11. Amoros E, Martin JL, Laumon B, Chiron M. Road crash casualties: characteristics of police injury severity mis-classification. J Trauma. 2007;62(2):482-90.
12. Amoros E, Martin J-L, Laumon B. Estimating non-fatal road casualties in a large French county, using the capture-recapture method. Accid Anal Prev. 2007;39(3):483-90.
13. Brenner H. Application of capture-recapture methods for disease monitoring: potential effects of imperfect record linkage. Methods Inf Med. 1994 Dec;33(5):502-6.
14. Brenner H, Schmidtmann I. Determinants of homonym and synonym rates of record linkage in disease registration. Methods Inf Med. 1996 Mar;35(1):19-24.
15. Gallay A, Nardone A, Vaillant V, Desenclos JC. La méthode capture-recapture appliquée à l'épidémiologie: principes, limites et applications. [The capture-recapture applied to epidemiology: principles, limits and application]. Rev Epidemiol Sante Publique. 2002 Apr;50(2):219-32.
16. Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. Epidemiol Rev. 1995;17(2):243-64.
17. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation II: Applications in human diseases.. Am J Epidemiol. 1995 Nov 15;142(10):1059-68.
18. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation I: History and theoretical development. Am J Epidemiol. 1995 1047-1058;142(10).
19. Aptel I, Salmi LR, Masson F, Bourdé A, Henrion G, Erny P. Road accident statistics: discrepancies between police and hospital data in a French island. Accid Anal Prev. 1999;31(1-2):101-8.
20. Dhillon PK, Lightstone AS, Peek-Asa C, Kraus JF. Assessment of hospital and police ascertainment of automobile versus childhood pedestrian and bicyclist collisions. Accid Anal Prev. 2001;33(4):529-37.

CONFIDENTIEL

21. Meuleners LB, Lee AH, Cercarelli LR, Legge M. Estimating crashes involving heavy vehicles in Western Australia, 1999-2000: A capture-recapture method. Accid Anal Prev. 2006 2006/1;38(1):170-4.
22. Razzak JA, Luby SP. Estimating deaths and injuries due to road traffic accidents in Karachi, Pakistan, through the capture-recapture method. Int J Epidemiol. 1998;27:866-70.
23. Roberts I, Scragg R. Application of capture-recapture methodology to estimate the completeness of child injury surveillance. Journal of Paediatrics and Child Health. 1994 Dec;30(6):513-4.
24. Tercero F, Andersson R. Measuring transport injuries in a developing country: an application of the capture-recapture method. Accid Anal Prev. 2004 2004/1;36(1):13-20.
25. Tilling K, Sterne JA. Capture-recapture models including covariate effects. Am J Epidemiol. 1999 Feb 15;149(4):392-400.
26. Amoros E, Martin JL, Laumon B. Under-reporting of road crash casualties in France. Accid Anal Prev. 2006;38(4):627-35.
27. Esteve J, Benhamou E, Raymond L. Descriptive epidemiology. Lyon: International agency for research on cancer; 1994.
28. Diciccio TJ, Efron B. Bootstrap confidence intervals. Statistical Science. 1996;11(3):189-228.
29. Laumon B. (edited by) Recherches coordonnées sur les traumatismes consécutifs à un accident de la circulation routière, et sur leurs causes et conséquences. [Coordinated research studies on injuries from road crashes and on their causes and consequences]. UMRETTE report. Bron: INRETS; 2002. Report No.: 0205.
30. Jarvis SN, Lowe PJ, Avery A, Levene S, Cormack RM. Children are not goldfish - mark/recapture techniques and their application to injury data. Inj Prev. 2000;6:46-50.
31. Morrison A, Stone DH. Capture-recapture: a useful methodological tool for counting traffic related injuries? Inj Prev. 2000 Dec;6(4):299-304.
32. Broughton J, Yannis G, Evgenikos P, et al. Deliverable 1.6 Estimation of the real number of road accident casualties. SafetyNet project, intermediate progress report on Task 1.5; 2005.
33. Saillant G, Pascal-Moussellard H, Langeron O, Lazennec J. Les lésions traumatiques de la moelle épinière : épidémiologie et prise en charge hospitalière. [Spinal cord trauma: epidemiology and pre-hospital management]. Bull Acad Natl Med. 2005;189(6):1095-106.
34. Masson F, Thicoipe M, Aye P, et al. Epidemiology of Severe Brain Injuries: A Prospective Population-Based Study. J Trauma. 2001;51(3):481-9.
35. Robertson LS. Injury Epidemiology. 2nd ed. ed: Oxford University Press, USA; 1992.
36. ONISR. Exposition au risque [Exposure to crash risk]. In: ONISR, editor. La sécurité routière en France, bilan de l'année 2004 [Traffic safety in France, 2004 results]. Paris: La documentation Française; 2005. p. 211-6.
37. Martin JL, van Kampen B, Perez K. Deliverable D9 WP3 (Hospital data) data analysis: Pan-European Co-ordinated Accident and Injury Databases (PENDANT); 2006.

CONFIDENTIEL

TITRE :
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)

RÉSUMÉ :
Les victimes d'accident de la route sont recensées par les forces de l'ordre, en France, et dans la plupart des pays. L'enregistrement des blessés non-décédés est incomplet et biaisé (biais de sélection et de classement de la gravité).
Un registre médical couvre le département du Rhône. La coexistence des recensements policier et médical permet par capture-recapture d'estimer un bilan exhaustif au niveau rhodanien, et ainsi d'estimer les facteurs de correction des données policières. En faisant l'hypothèse d'homogénéité, sur le territoire national, des pratiques policières d'enregistrement des blessés, nous appliquons les coefficients de correction aux données policières nationales, en redressant sur les facteurs de biais.
Les effectifs annuels moyens sur 1996-2004 sont alors estimés à 514 000 blessés dont 61 000 blessés graves, soit 3,7 et 2,2 fois les décomptes des forces de l'ordre. Les blessés avec séquelles lourdes sont estimés à 8000 annuellement, soit autant que les tués.

DISCIPLINE :
Épidémiologie

MOTS-CLÉS :
Épidémiologie, accidents de la route, blessés, séquelles, biais de sélection, erreurs de classement, registre, incidence, capture-recapture, projection, standardisation indirecte, redressement.

PHD TITLE :
Non-fatal road casualties: estimation of frequency and injury severity, France 1996-2004, modelled from a medical registry (Rhône area) and police data (France)

ABSTRACT :
In France, as in most countries, road casualty numbers are provided by the police. Non-fatal casualties are under-estimated. This under-reporting is rather large and biased (selection and classification bias).

A road trauma registry covers the Rhône area. By using capture-recapture method on those two registrations, we estimate the total number of casualties in the Rhône area, and hence the police data correction factors. Those are applied to the nationwide police data, with standardisation on bias factors. The underlying assumption for that projection is the homogeneity of police reporting practices across the national area.

Between 1996 and 2004, an annual average of 514 000 non-fatal casualties is estimated, including 61 000 seriously injured. Those are respectively 3.7 and 2.2 times police results. Casualties suffering major impairment are estimated at 8000 annually, which is as many as fatal casualties.

KEYWORDS :
Epidemiology, traffic accidents, injury, impairment, selection bias, mis-classification, registry, incidence, capture-recapture, projection, indirect standardisation, weighting adjustment.

INTITULÉ ET ADRESSE DU LABORATOIRE :
Unité Mixte de Recherche Épidémiologique et de Surveillance Transport Travail Environnement (UMRESTTE INRETS-UCBL-InVS) UMR T 9405
INRETS
25 avenue François Mitterrand
Case 24,
69675 BRON CEDEX