

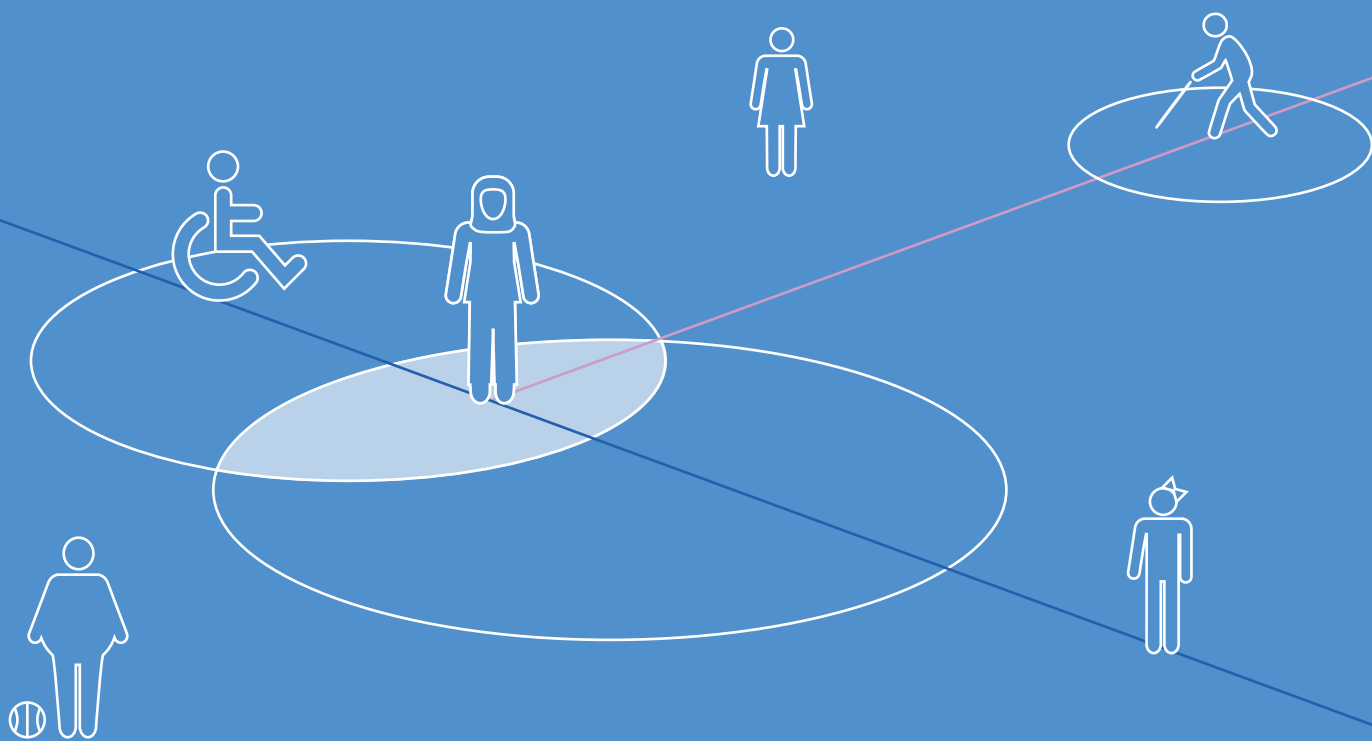


UNODC

United Nations Office on Drugs and Crime

SURVEILLANCE DE LA PRÉVALENCE DE LA TRAITE DES ÊTRES HUMAINS PAR UNE ESTIMATION DE SYSTÈMES MULTIPLES

Un manuel des Nations Unies pour les
décideurs, les praticiens et les
chercheurs qui participent à l'objectif
16.2 Du développement durable



Cette publication peut être reproduite en totalité ou en partie et sous quelque forme que ce soit à des fins éducatives ou non lucratives sans autorisation spéciale du détenteur des droits d'auteur, à condition que la source soit mentionnée.

Citation suggérée :

UNODC, Suivi de la prévalence de la traite des êtres humains par l'estimation de systèmes multiples 2022

Les commentaires sur le manuel sont les bienvenus et peuvent être envoyés à l'adresse suivante :

Centre international de Vienne, PO Box 500, 1400 Vienne, Autriche

Tél : (+43-1) 26060-0, Fax : (+43-1) 26060-5866.

Courriel : askresearch@unodc.org

Avis de non-responsabilité

Le contenu de cette publication ne reflète pas nécessairement les opinions ou les politiques de l'ONUDC, des États membres ou des organisations contributrices, et n'implique aucune approbation.

Édition non commerciale

Nations Unies, juin 2022. Tous droits réservés, dans le monde entier.

Les appellations employées dans cette publication et la présentation des données qui y figurent n'impliquent de la part du Secrétariat des Nations Unies aucune prise de position quant au statut juridique des pays, territoires, villes ou zones, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.

Cette publication n'a pas été formellement éditée.

PUBLICATION DES NATIONS UNIES

Remerciements

La préparation de ce manuel a été coordonnée par l'équipe de recherche sur la traite des personnes et le trafic illicite de migrants de la Section de la recherche criminelle de l'ONUDC. Les contributions des membres du personnel de l'ONUDC tels que Ragnie Johansen, Salomé Flores Sierra Franzoni et Fabrizio Sarrica sont vivement remerciées.

Des remerciements particuliers sont dus à Peter G.M. van der Heijden, Maarten Cruyff et Jan van Dijk et Joris Beijers pour la préparation de la version finale du présent manuel.

Ce Manuel a bénéficié de l'expertise et des conseils d'Annette Brunovskis, Davina Durgana, Ieke de Vries, Ruth King, Bernard Silverman et Kyle Vincent.

La Section de la recherche criminelle de l'ONUDC tient également à remercier le Service des statistiques démographiques et sociales de la Division des statistiques du Département des affaires économiques et sociales des Nations Unies, qui a rendu possible l'élaboration de ce manuel.

SURVEILLANCE DE LA PRÉVALENCE DE LA TRAITE DES ÊTRES HUMAINS PAR UNE ESTIMATION DE SYSTÈMES MULTIPLES

Un manuel des Nations Unies pour les décideurs, les praticiens et les chercheurs qui participent à l'objectif 16.2 Du développement durable

RÉSUMÉ EXÉCUTIF

Les politiques efficaces pour lutter contre la traite des êtres humains ne peuvent que se baser sur des connaissances solides et à jour sur la prévalence et les tendances. Étant donné que les statistiques sur les cas enregistrés reflètent en grande partie les efforts des institutions impliquées, elles ne peuvent pas être utilisées pour mesurer la prévalence réelle et sont donc inadaptées au suivi des progrès dans la réalisation de l'ODD 16.2. L'option consistant à mener des enquêtes régulières auprès de la population sur la victimisation autodéclarée par traite des êtres humains n'est pas facilement réalisable dans tous les pays. Une alternative économique est la triangulation des dossiers administratifs disponibles sur les victimes de la traite provenant des acteurs étatiques et non étatiques pour estimer les proportions de victimes non enregistrées, une méthode statistique connue sous le nom de capture-recapture ou estimation de systèmes multiples (MSE).

Dans ce manuel, les principes de base et les hypothèses de la MSE sont expliqués à l'intention du lecteur général. Ensuite, le manuel explique quels types de données administratives sont nécessaires pour la MSE et comment les ensembles préexistants de données de la police, des inspections du travail ou des ONG prestataires de services peuvent être utilisés pour la MSE ou adaptés à cette fin. Dans plusieurs pays européens, des systèmes intégrés de données sur les victimes de la traite existent dans le cadre des mécanismes nationaux d'orientation. Ceux-ci peuvent généralement être utilisés pour la MSE après des adaptations mineures. Dans les pays où il n'existe pas d'ensembles de données intégrés sur les victimes de la traite, des ensembles de données autonomes provenant d'institutions pertinentes peuvent être consolidés par des équipes de recherche spécialisées, comme cela a été fait avec succès dans les études de MSE aux États-Unis et en Australie. Cette partie du manuel

traite également de la façon dont la consolidation des données personnelles sensibles peut être effectuée conformément à la réglementation sur la protection des données, par exemple par le cryptage des données ou l'implication d'un bureau national de statistique. Il explique enfin comment la MSE produit non seulement des estimations de prévalence, mais peut également fournir un aperçu des catégories de victimes de la traite des êtres humains qui sont le plus cachées entre la population, ce qui semble malheureusement être le cas dans certains pays, les victimes enfants. De cette manière, les résultats de la MSE fournissent des indications sur la manière dont les tendances nationales de la traite des êtres humains se comparent au niveau international, sur les efforts nationaux de lutte contre la traite qui doivent être renforcés et sur les types de traite en particulier qui méritent une priorité plus élevée dans les projets de politiques publiques.

Contenido

Guide des lecteurs	9
Section 1: Présentation de l'estimation de systèmes multiples comme nouvelle façon d'estimer la prévalence de la traite et faire le suivi des progrès vers l'ODD 16.2	10
<i>Le Protocole de Palerme et la Déclaration du Millénaire</i>	11
<i>Pourquoi de meilleures statistiques nationales sur les victimes de la traite des êtres humains sont d'une urgente nécessité</i>	14
<i>L'option de l'estimation de systèmes multiples</i>	16
<i>Qu'est-ce que la MSE ?</i>	18
<i>Quelques hypothèses fondamentales derrière la MSE</i>	19
<i>Le «statut officiel» des estimations basées sur la MSE</i>	22
<i>Problèmes de protection des données</i>	23
<i>Points de synthèse</i>	25
Section 2: Identification, collecte et organisation des données utiles à la MSE	28
<i>Clarification des termes</i>	29
<i>Collecte de données</i>	30
<i>La sélection des listes appropriées</i>	31
<i>L'exigence de listes indépendantes</i>	39
<i>Les identifiants uniques</i>	40
<i>Les ensembles de données utilisés dans les études irlandaise et britannique</i>	41
<i>Utilisation de covariables pour améliorer la robustesse et la pertinence des résultats pour les politiques publiques</i>	44
<i>Points de synthèse</i>	46

Section 3: Modélisation statistique pour l'estimation de systèmes multiples	48
<i>Théorie</i>	49
<i>Covariables</i>	51
<i>Intervalle de confiance pour l'estimation de N</i>	52
<i>Trois listes</i>	53
<i>Sélection du modèle</i>	54
<i>Vérification des données et étapes préliminaires</i>	55
<i>Répartition des listes au fil des années</i>	57
<i>Figure 4 Histogramme des fréquences observées sur les listes R1, R2 et R3</i>	57
<i>La ' mse < de paquet R</i>	61
<i>Deux listes, pas de covariables, pas d'années</i>	62
<i>Deux listes, une covariable, pas d'année</i>	63
<i>Deux listes, une covariable et l'année</i>	65
<i>Toutes les listes, toutes les covariables et toutes les années</i>	68
Section 4: Aperçu des résultats des études MSE et discussion	74
<i>Quelques résultats clés de la MSE dans le monde</i>	76
<i>Comparabilité internationale</i>	78
<i>Nombres estimés par covariables</i>	79
<i>Proportion entre les cas enregistrés/observés et estimés</i>	80
<i>Observations finales et discussion</i>	81
Bibliographie	83

Guide des lecteurs

Ce manuel fournit des conseils pratiques sur la façon d'appliquer la technique statistique de l'estimation de systèmes multiples (MSE) pour générer de meilleures estimations de la dimension de la traite des êtres humains par extrapolation à partir des données administratives des cas enregistrés. Il est destiné à un public mixte de décideurs politiques et de praticiens dans le domaine de la lutte contre la traite, engagés à atteindre l'objectif de développement durable 16.2, ainsi que de chercheurs impliqués dans la collecte et/ou l'analyse de données sur les victimes de la traite des êtres humains à cette fin.

9

La première section d'introduction explique pourquoi la MSE est une approche prometteuse et rentable pour observer la prévalence de la traite et les progrès vers la réalisation de la cible 16.2 des objectifs de développement durable. Cette section est principalement destinée aux décideurs politiques et à d'autres personnes intéressées à placer la lutte contre la traite des êtres humains sur une base de preuves plus solide. Il vise à expliquer la théorie qui sous-tend la MSE pour le lecteur commun. Il explique ensuite les faibles coûts de la MSE et ses nombreuses utilisations pour la planification et l'évaluation des politiques publiques. Il répond également aux préoccupations concernant les problèmes de protection des données concernant les informations personnelles sensibles.

Les deuxième et troisième sections s'adressent aux responsables de la collecte, du stockage et de l'analyse des données sur les victimes de la traite des êtres humains, par exemple ceux qui travaillent pour des institutions publiques spécialisées, des ONG, des observatoires ou des instituts de recherche universitaires. La deuxième section définit les exigences en matière de données pour l'exécution de MSE, y compris les considérations de protection des données, et présente la théorie sous-jacente à MSE dans un langage plus technique. La troisième section montre étape par étape comment la MSE a été appliquée à un ensemble de données irlandais composé de trois listes et de quatre covariables, en utilisant un paquet MSE sur mesure dans le langage R librement disponible sur Internet. Cette section montre également comment l'adéquation d'un ensemble de données disponible pour MSE peut être vérifiée.

SECTION 1

Présentation de
l'estimation de systèmes
multiples comme nouvelle
façon d'estimer la
prévalence de la traite
et faire le suivi des
progrès vers l'ODD 16.2

Dans la dernière section, le manuel présente un aperçu des résultats des études MSE menées ces dernières années et discute des leçons apprises et des défis restants.

11

Le Protocole de Palerme et la Déclaration du Millénaire

Bien que la traite des êtres humains ne soit pas un phénomène nouveau, l'adoption du Protocole Additionnel à La Convention Des Nations Unies contre la Criminalité Transnationale Organisée Visant à Prévenir, Réprimer et Punir la Traite des Personnes, en Particulier des Femmes et des Enfants¹ (ci-après le *Protocole des Nations Unies sur la traite des personnes*) en 2000 à Palerme, en Italie, a marqué un bond en avant dans l'engagement de la communauté internationale à y faire face. Le Protocole des Nations Unies sur la traite des personnes a fourni la première définition internationalement acceptée de ce crime odieux très fréquent et une base juridique solide pour une action et une collaboration efficaces contre la traite à travers le monde. Son entrée en vigueur en 2003 a déchaîné une vague d'action, au fur et à mesure que les pays introduisaient une législation nouvelle ou adaptée, créaient des initiatives d'assistance aux victimes, élaboraient des plans d'action nationaux et amélioraient les performances de leurs fonctionnaires du système de justice pénale. Une communauté anti-traite plus élargie issue de la société civile a également commencé à se former, composée d'ONG, d'initiatives confessionnelles, de centres de recherche et de groupes de défense, pour n'en citer que quelques-uns.

1. Le Protocole additionnel à la Convention des Nations Unies contre la criminalité transnationale organisée visant à prévenir, réprimer et punir la traite des personnes, en particulier des femmes et des enfants.

La définition de la traite dans le Protocole des Nations Unies sur la traite des personnes fournit un terrain d'entente pour des réponses concertées face à ce crime. La définition complète se lit comme suit :

‘... le recrutement, le transport, le transfert, l'hébergement ou l'accueil de personnes, par la menace ou l'usage de la force ou d'autres formes de coercition, d'enlèvement, de fraude, de tromperie, d'abus de pouvoir ou d'une position de vulnérabilité ou de l'octroi ou de la réception de paiements ou d'avantages pour obtenir le consentement d'une personne ayant autorité sur une autre personne, à des fins d'exploitation. A/RES/55/25, 15 novembre 2000, art. 3 (a) ».

Comme il ressort de la définition du protocole du concept d'exploitation, celui-ci s'applique à un large éventail de différentes formes d'exploitation, y compris l'exploitation sexuelle et le travail forcé :

“L'exploitation comprendra, au minimum, l'exploitation de la prostitution d'autrui ou d'autres formes d'exploitation sexuelle, le travail ou les services forcés, l'esclavage ou des pratiques analogues à l'esclavage, la servitude ou le prélèvement d'organes.”

L'article 28 de la convention mère contre la criminalité transnationale organisée appelle les États parties à élaborer, en consultation avec les milieux universitaires, des « définitions et méthodologies communes » pour analyser les « tendances de la criminalité organisée ». En 2015, la quête de données solides et fiables sur la traite des êtres humains a été revigorée par l'adoption du Programme de développement durable à l'horizon 2030 par l'Assemblée générale des Nations Unies. Dans cet agenda, la traite des personnes est spécifiquement mentionnée dans trois cibles des ODD, à savoir 5.2, 8.7 et 16.2. La cible 16.2 appelle les pays à « mettre fin aux abus, à l'exploitation, à la traite et à toutes les formes de violence et de torture à l'égard des enfants ».² Le Groupe inter-agence et d'experts sur les indicateurs des ODD de la Commission de statistique des Nations Unies a adopté des ensembles d'indicateurs pour chaque cible des ODD. Pour la cible 16.2, l'un des indicateurs est : « nombre de victimes de la traite des êtres humains pour 100 000 habitants, par sexe, âge et forme d'exploitation » (indicateur ODD 16.2.2.). L'ONU DC a été désigné comme l'agence dépositaire de cet indicateur, responsable du développement méthodologique et de la facilitation des rapports au niveau des pays. Au vu des doutes persistants quant à l'utilité des statistiques officielles sur les cas enregistrés comme indicateur de prévalence, l'ONU DC a organisé en décembre 2013 un atelier d'experts intégré par des chercheurs spécialisés dans la mesure des populations insaisissables

2. Pour plus d'informations sur les objectifs de développement durable, veuillez consulter: <http://sustainabledevelopment.un.org>.

pour discuter d'options de mesure alternatives. L'une des recommandations issues de cette réunion était d'explorer la faisabilité de l'estimation de systèmes multiples comme méthode d'estimation de la prévalence de la traite des êtres humains grâce à une analyse des données sur les victimes enregistrées (Van der Heijden et al., 2015).

Ce manuel vise à aider les pays à estimer la prévalence nationale de la traite des personnes en appliquant l'estimation de systèmes multiples (MSE) à leurs données administratives sur les victimes de la traite. La MSE est une méthode statistique pour estimer la taille des populations insaisissables. Grâce à la mise en relation des registres disponibles provenant de diverses institutions concernant des personnes enregistrées faisant partie de cette population, on obtient la part de la population observée par n'importe quelle institution. Ensuite, le recouvrement entre les registres est utilisé pour estimer la partie non observée. Dans le cas des victimes de la traite des êtres humains, des exemples de registres possibles sont ceux des forces de police, des inspections du travail et des ONG fournissant des services aux victimes. L'objectif de ce manuel est d'expliquer aux décideurs politiques chargés de mettre fin à la traite des êtres humains pourquoi la MSE est un outil prometteur pour suivre les progrès de ces efforts et de fournir des conseils techniques aux praticiens et aux chercheurs sur la manière dont la MSE peut réellement être mise en œuvre en utilisant les données nationales dont ils disposent.

Le manuel est basé sur les expériences et les enseignements tirés d'une première vague de dix études MSE sur la traite des êtres humains, dont plusieurs ont été soutenues par l'ONUDC³. S'appuyant sur cette base de connaissances, le manuel décrit les principes de base sous-jacents à la MSE et le type de données requises. Il fournit également un guide pratique pour la procédure d'estimation proprement dite. Enfin, il résume les enseignements tirés des études MSE antérieures sur les données de la traite des êtres humains et présente des exemples d'études achevées avec succès qui peuvent servir de modèles à ceux qui cherchent à appliquer la MSE à des ensembles de données consolidées sur la traite des êtres humains dans leur propre pays.

3. La première étude MSE utilisant des données sur la traite a été réalisée au Royaume-Uni en 2014 (Silverman et al./Home Office, 2014). Cette étude a été suivie d'une série d'études initiées par l'ONUDC, utilisant des données des Pays-Bas (2017), d'Irlande, de Roumanie et de Serbie (2018). Les rapports sont disponibles sur : <https://www.unodc.org/unodc/en/data-and-analysis/tip.html>. Des études ultérieures incluent l'Australie (Lyneham, Dowling & Bricknell, 2019) et la Slovaquie (Walk Free Foundation, 2019), ainsi qu'une étude MSE non publiée sur des données de Biélorussie. Des études ont également été menées dans plusieurs endroits aux États-Unis, notamment à la Nouvelle-Orléans (Bales, Murphy et Silverman, 2020) et sur trois sites d'étude non spécifiés (Farrell et al., NIJ, 2019). À l'aide des données de la police sur les arrestations et les réarrestations pour exploitation sexuelle à des fins commerciales, des estimations du nombre total de victimes au Kansas ont été effectuées à l'aide d'une analyse de capture-recapture (Phillips, 2017).

Pourquoi de meilleures statistiques nationales sur les victimes de la traite des êtres humains sont d'une urgente nécessité

Les statistiques sur l'étendue et la nature de la traite des êtres humains sont utilisées pour sensibiliser à l'ampleur et à la gravité du phénomène, aider les gouvernements, les organisations internationales et les organisations non gouvernementales à élaborer des réponses fondées sur des données probantes et à suivre les progrès de leur mise en œuvre (De Vries & Dettmeijer -Vermeulen, 2015). Les statistiques internationales sur la traite des êtres humains permettent également de comparer les progrès achevés d'identifier les meilleures pratiques, comme l'envisage l'Agenda des Nations Unies pour le développement durable.

Les organismes d'État du monde entier recueillent des données sur le nombre de victimes de la traite des êtres humains détectées. Ces statistiques nationales sur les victimes enregistrées sont régulièrement rassemblées par l'Office des Nations Unies contre la drogue et le crime et publiées dans le *Rapport mondial biennal sur la traite des personnes*.⁴ La collecte de ces statistiques internationales sur les victimes détectées est facilitée par deux circonstances. Premièrement, la traite des personnes est l'un des rares types d'infractions pour lesquelles, comme on l'a dit, une définition juridique internationale a été adoptée, à savoir la définition donnée dans le Protocole des Nations Unies sur la traite des personnes. Cette définition est désormais largement reflétée dans les codes pénaux de 169 pays à travers le monde (ONUDC, 2020). Deuxièmement, un nombre croissant de pays ne comptent plus uniquement sur les forces de l'ordre pour atteindre les victimes. Conformément à l'esprit centré sur les victimes du Protocole de Palerme, ils ont mis en place des dispositions pour l'identification et l'orientation systémiques des victimes de la traite des personnes par, entre autres, la police, les inspections du travail et les ONG prestataires de services. Certains pays ont officialisé cette coopération multi-agences par le biais de mécanismes spéciaux tels qu'un mécanisme national d'orientation (NRM) qui stipulent les rôles et les responsabilités de chaque entité impliquée dans le processus d'identification (OSCE/ODHIR, 2004). Lorsque de tels dispositifs institutionnels sont en place, des statistiques complètes et multisources sur les victimes de la traite des êtres humains peuvent être recueillies auprès des diverses institutions publiques et non publiques concernées. C'est désormais le cas dans un nombre substantiel et en croissance rapide d'États membres de l'ONU.

Malgré ces améliorations, la collecte de statistiques internationalement comparables sur les tendances de la traite des êtres humains, comme le demande l'agenda des ODD,

4. Disponible sur : www.unodc.org/glotip.

laisse à désirer. Les rapports des organisations internationales montrent que les statistiques recueillies soulèvent de sérieuses questions quant à leur validité, leur fiabilité et leur comparabilité⁵. Par exemple, les taux de victimes de la traite pour 100 000 habitants dans le Rapport mondial des Nations Unies sont jusqu'à 100 fois plus élevés dans certains pays que dans d'autres. Cette énorme variation transnationale pourrait en théorie refléter les différences existantes dans la prévalence de la traite. Cependant, il est plus probable qu'il reflète la variabilité dans l'efficacité des agences nationales pour l'arrestation des trafiquants, le sauvetage des victimes et l'enregistrement des cas. Fait révélateur, certains pays accordent peu ou pas de priorité à la détection des victimes de ce crime. D'autres limitent leurs efforts à certaines sous-catégories, par exemple la traite des femmes à des fins d'exploitation sexuelle et ignorent les autres formes. En conséquence, le nombre de victimes enregistrées dans ces pays reste relativement faible alors que leur « nombre réel » peut être très élevé.

Les taux de victimes enregistrées ne varient pas seulement énormément d'un pays à l'autre. Les chiffres des pays montrent souvent aussi une grande instabilité dans le temps, augmentant ou diminuant parfois de 50 % ou plus d'une année à l'autre. Cette volatilité pourrait, encore une fois, refléter des changements dans la prévalence réelle, par exemple parce qu'au cours d'une année donnée, certaines opérations de traite d'une ampleur inhabituelle impliquant de nombreuses victimes ont été démantelées. Cependant, ces fluctuations sont plus susceptibles de refléter des changements dans les pratiques d'identification et/ou d'enregistrement. Le *Rapport mondial 2018* de l'ONUDC fournit plusieurs exemples d'augmentations significatives du nombre de victimes enregistrées dans un pays par an après l'introduction d'une nouvelle législation, d'institutions dédiées à la lutte contre la traite ou de programmes anti-traite.⁶ L'inverse se produit également, comme une chute soudaine du nombre de victimes identifiées en raison de la réduction des efforts de détection par les forces de l'ordre en cours de restructuration, comme cela s'est produit aux Pays-Bas après 2014 (CoMensha, 2019). Pour ces raisons, les changements dans le nombre de victimes identifiées au fil du temps peuvent ne pas être indicatifs des changements dans la prévalence réelle de la traite.

En somme, les statistiques sur les victimes de la traite identifiées ne montrent que la part variable du phénomène qui est détectée par les autorités nationales et/ou

5. Voir, par exemple, ONUDC, , *Global Reports on Trafficking in Persons*; Conseil de l'Europe (2019), *9th General Report on GRETA's Activities* (GRETA, Groupe d'experts sur la lutte contre la traite des êtres humains) ; Commission européenne, *Data collection on trafficking in human beings in the EU*, 2020.

6. ONUDC, *Global Report on Trafficking in Persons 2018* (publication des Nations Unies, numéro de vente E.19.IV.2), pages 21-22.

les ONG à un certain moment. Ces statistiques ne peuvent pas être utilisées pour évaluer de manière fiable les variations de la prévalence réelle de la traite des êtres humains d'un pays à l'autre ou des changements au fil du temps par pays, et donc des progrès vers la réalisation de l'objectif de «mettre fin à l'exploitation et à la traite» tel que défini dans la cible 16.2 des ODD.

L'option de l'estimation de systèmes multiples

Le développement et la mise à l'essai de méthodes pour mieux estimer la prévalence de la traite des êtres humains sont depuis un certain temps une priorité pour plusieurs chercheurs et organisations internationales. Le Bureau international du travail (2009), en collaboration avec la Walk Free Foundation, a appliqué la méthodologie des enquêtes par sondage pour mesurer les taux autodéclarés de victimisation par la traite des êtres humains. Les résultats d'enquêtes menées dans quarante pays, principalement des pays en développement, ont été utilisés pour calculer les taux mondiaux et régionaux de victimisation de l'"esclavage moderne", ainsi que les taux d'une centaine d'autres pays par extrapolation (Walk Free Foundation, 2018).⁷ Lors d'une vague ultérieure en 2020, des enquêtes ont été menées dans trente autres pays, portant le total à soixante-dix. La méthodologie de l'enquête, bien que prometteuse à bien des égards, comporte des limites inhérentes.⁸ Il est difficile de parvenir à un consensus international sur les questions précises utilisées pour mesurer la traite des personnes, telle que définie dans le Protocole des Nations Unies, et sur les cadres d'échantillonnage et les modes d'entretien appropriés.⁹ Pour déterminer les changements statistiquement significatifs au fil du temps dans la prévalence de la traite des êtres humains, il faut répéter régulièrement des enquêtes à grande échelle avec des implications budgétaires importantes.

En 2014, Silverman et d'autres ont appliqué la MSE à la base de données sur les victimes de la traite compilée par la National Crime Agency du Royaume-Uni

7. L'Indice mondial de l'esclavage présente des estimations de la prévalence de «l'esclavage moderne», un concept qui n'a pas de définition juridique internationalement reconnue et est décrit comme suit: «L'esclavage moderne fait référence à des situations d'exploitation qu'une personne ne peut pas refuser ou quitter en raison de menaces, la violence, la coercition, la tromperie ou l'abus de pouvoir».

8. Pour un examen des différentes approches de mesure de la traite des êtres humains, voir Zhang & Joudo-Larsen, 2021, et Barrick & Pfeffer, 2021.

9. Les difficultés de la recherche par sondage dans ce domaine comprennent une distribution faussée de la traite des êtres humains parmi les populations de nombreux pays, avec une concentration dans les segments difficiles à atteindre tels que les migrants irréguliers, et la sensibilité des questions sur l'exploitation sexuelle et la criminalité forcée. Pour une étude sur la traite des êtres humains parmi les communautés de migrants, voir Zhang, 2012.

(Silverman, 2014 ; Bales, Hesketh et Silverman). Ensuite, une étude pilote, initiée par l'ONU DC, a appliqué la MSE à l'ensemble de données disponibles sur les victimes de la traite des êtres humains conservé par une ONG dédiée aux Pays-Bas (Co-Mensha, anciennement La Strada Pays-Bas) (Van Dijk & Van der Heijden, 2016). Une étude de suivi a différencié les différentes catégories de victimes, y compris les victimes de la traite à des fins d'exploitation sexuelle ou de travail forcé. Les résultats ont été copubliés par l'ONU DC et le Rapporteur national néerlandais sur la traite des êtres humains en 2017.¹⁰ Depuis lors, des études similaires ont été menées dans d'autres pays européens, ainsi qu'en Australie et aux États-Unis. Les résultats ont été publiés dans les Rapports mondiaux de l'ONU DC sur la traite des êtres humains et dans un numéro spécial du Journal sur la criminalité et la délinquance (2021). En 2021, l'étude néerlandaise a été répétée avec un financement du ministère de la Justice et de la Sécurité, en utilisant des données sur les années 2016-2019 (Van Dijk, Cruyff & Van der Heijden, 2021).

Les directives de l'OIT concernant la mesure du travail forcé, adoptées par la 20^e Conférence des statistiques du travail en 2018, mentionnent la MSE comme une méthode rentable pour estimer la prévalence en accompagnement des enquêtes sur la population.¹¹

En résumé, la MSE s'est développée au cours des dernières années en une méthode prometteuse et éprouvée pour estimer la prévalence de la traite des êtres humains grâce à une analyse des statistiques disponibles sur les cas enregistrés.

10. Disponible sur : <https://www.unodc.org/documents/data-and-analysis/tip/TiPMSE.pdf>. Voir aussi Cruyff, Van Dijk et Van der Heijden, 2017.

11. «Les dossiers administratifs, tels que les listes de personnes astreintes au travail forcé compilées par les autorités locales ou les dossiers de la police, ou établies par des organisations non gouvernementales et d'autres prestataires de services, peuvent être utiles pour produire des estimations de la prévalence du travail forcé à un coût relativement faible. Lorsqu'il existe différentes sources administratives qui se réfèrent à des périodes de référence communes et peuvent être confrontées les unes aux autres afin de mesurer leur chevauchement avec une précision raisonnable, des estimations de la prévalence du travail forcé peuvent être dérivées sous certaines hypothèses, appelées estimations de systèmes multiples ». (OIT, 2018).

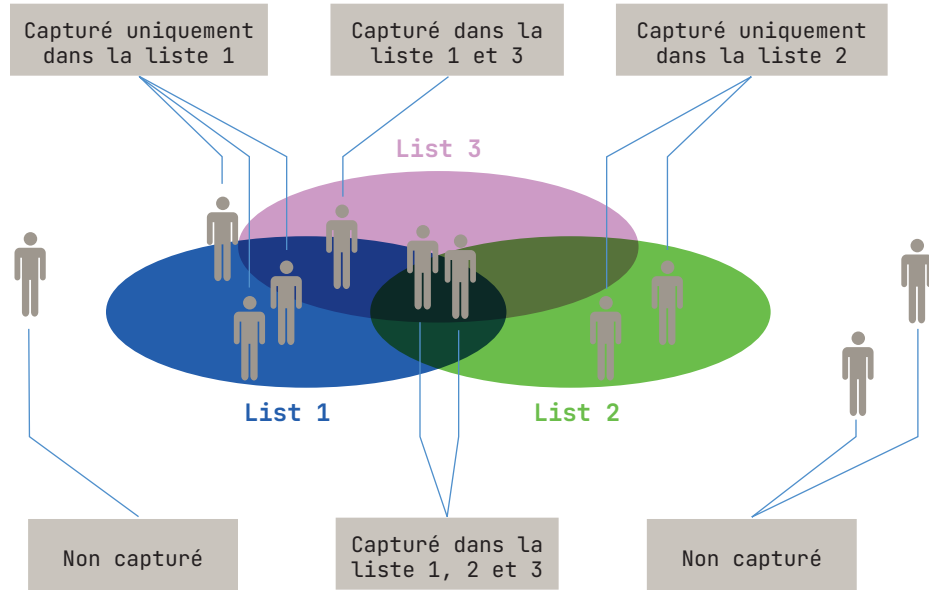
Qu'est-ce que la MSE ?

Dans la troisième section de cet article, une introduction plus formelle sera donnée à la MSE en tant que méthode statistique. Ici, nous allons essayer d'expliquer les principes de base de la MSE pour le lecteur non averti.

Selon certaines sources, la MSE remonte à la procédure de Capture-Recapture par laquelle le biologiste marin Petersen estimait la taille des populations de poissons dans les étangs ou les fjords. Selon d'autres, la méthode est déjà appliquée depuis bien plus longtemps (Goudie & Goudie, 2007). Quoi qu'il en soit, la procédure d'estimation des populations de poissons par les biologistes se pratiquait comme suit. Les chercheurs capturaient d'abord, par exemple, dix poissons, les marquaient et les remettaient à l'eau. Plus tard, ils revenaient pêcher dans le même étang pour voir quelle partie d'un nouveau lot de poissons capturés était marquée. Quand cela équivalait, par exemple, à deux poissons sur dix, ils savaient que sur les dix poissons marqués laissés dans l'étang, un sur cinq avait été pêché. Partant du principe que les autres poissons de l'étang avaient à peu près la même chance d'être capturés, ils en concluaient qu'il y avait une cinquantaine de poissons dans l'étang dont dix marqués et quarante non. Cette procédure d'estimation est connue sous le nom d'estimateur de Petersen.

La MSE peut être comprise comme une version avancée de la capture-recapture, dans laquelle la taille d'une population cachée d'humains est estimée en analysant le chevauchement entre trois ou plusieurs listes administratives sur lesquelles les personnes appartenant à cette population apparaissent. Les personnes appartenant à la population cachée des victimes de la traite peuvent, par exemple, être enregistrées par plusieurs agences gouvernementales telles que la police, l'immigration, les inspecteurs du travail ainsi que par des prestataires privés d'assistance juridique, médicale ou psychologique ou de garde d'enfants ou de jeunes. En modélisant la répartition des victimes enregistrées sur ces listes, une estimation peut être faite des victimes qui n'apparaissent sur aucune des listes de la police, des ONG ou d'une autre institution. Cet exercice fournit une estimation des nombres non enregistrés (ou non capturés), tel qu'illustré dans la figure ci-dessous.

Figure 1 Illustration du fonctionnement de la MSE en évaluant les chevauchements entre les listes de victimes observées pour estimer la population totale



Source: Élaboration de l'UNODC

Quelques hypothèses fondamentales derrière la MSE

Les hypothèses fondamentales de MSE sont mieux expliquées en exposant les hypothèses fondamentales de l'estimation à double système (DSE), c'est-à-dire la MSE avec seulement deux listes (ou capture-recapture). Selon une étude faisant autorité (IWGDME, 1995), les données utilisées pour la DSE doivent répondre à quatre hypothèses fondamentales :

- 1) les données doivent concerner des personnes appartenant à une population fermée ;
- 2) chaque personne doit être identifiable de manière unique afin d'être repérée dans plusieurs listes ;
- 3) chaque personne doit avoir la même chance d'être inscrite sur les listes ;
- 4) lorsque seulement deux listes sont utilisées, le placement d'une personne sur l'une des listes doit être statistiquement indépendant du placement sur l'autre.

La première hypothèse peut être facilement comprise dans l'exemple des pêcheurs revenant pour une seconde prise. Si l'endroit où les captures ont lieu est une rivière à courant rapide, les poissons marqués n'auraient pratiquement aucune chance d'être recapturés. Dans les DSE concernant les populations humaines, la condition peut être approchée en limitant l'analyse aux données concernant les personnes présentes dans une zone géographique pendant une période de temps limitée. Dans le cas des données sur les victimes de la traite enregistrées, l'hypothèse peut être approximée en utilisant un ensemble de données sur les personnes enregistrées comme victimes (présumées) de la traite sur le territoire d'un pays au cours d'une certaine année.

La deuxième hypothèse peut techniquement être facilement satisfaite lorsque les données sur les victimes proviennent des administrations, par exemple, des forces de l'ordre, des ONG fournissant un soutien, etc. Ces administrations contiennent généralement des données personnelles sur les personnes concernées. Cependant, les réglementations sur la protection de la vie privée peuvent faire obstacle à l'échange de ces données personnelles entre institutions. Le respect de ces réglementations peut donc nécessiter des mesures spéciales dans une étude de MSE, comme nous le verrons plus en détail ci-dessous.

Des recherches antérieures ont démontré que la troisième hypothèse d'égalité des probabilités d'inclusion doit être satisfaite par une seule des listes utilisées et non par toutes (Van der Heijden et al., 2012). Dans le cas des victimes de la traite des êtres humains, les administrations nationales de la police sont généralement qualifiées dans ce sens-là. Néanmoins, même cette hypothèse relâchée concernant une inclusion égale sur l'une des listes peut poser des problèmes pour la DSE avec les données sur la traite qui doivent être considérées. Diverses sous-catégories de victimes de la traite peuvent être hétérogènes dans leur probabilité d'être détectées par la police ou toute autre organisation pertinente. Par exemple, le travail forcé dans les zones reculées d'un grand pays peut être plus difficile à détecter que l'exploitation sexuelle dans les quartiers chauds bien connus des villes.

La quatrième hypothèse fondamentale sous-jacente à la DSE est que la probabilité pour une victime d'être inscrite sur la liste de l'organisation A, par exemple la police, devrait être en elle-même indépendante de la probabilité qu'une victime soit inscrite sur la liste B d'une autre organisation. La quatrième hypothèse découle logiquement de la troisième hypothèse car, lorsque la probabilité d'être inscrit sur la liste A est identique pour ceux qui sont sur la liste B ou qui ne sont pas sur la liste B, l'indépendance se tient. Dans l'exemple de la capture de poissons, le fait que des poissons soient capturés et marqués lors de la première capture ne devrait ni augmenter ni diminuer leur probabilité d'être capturés à nouveau lors des captures suivantes. A cet égard, il faut remarquer que les raisons pour lesquelles la probabilité

que des membres d'une population cachée soient enregistrés par une organisation particulière peuvent être co-dépendantes de leur placement sur la liste d'une autre organisation sont multiples.¹² Dans le domaine de la traite des êtres humains, l'inscription des victimes sur des listes de prestataires de services peut dépendre de leur inscription antérieure sur une liste d'institutions policières. En fait, dans les pays dotés de mécanismes nationaux d'orientation, les forces de l'ordre sont censées orienter les victimes de la traite identifiées vers des prestataires de services. Il faut donc s'attendre à une certaine mesure de dépendance statistique.

Nous concluons que les deux dernières hypothèses sont problématiques et ne tiendront pas en absolu. Il existe deux solutions à cela. Pour la troisième hypothèse, les personnes ayant la même probabilité d'être incluses dans chacune des listes, une solution consiste à ajouter des covariables pour les caractéristiques contextuelles telles que les informations sur le type d'exploitation de chaque victime à l'ensemble de données à utiliser dans la MSE. Lorsque cette « covariable » est ajoutée, l'hypothèse d'égalité des probabilités d'inscription des personnes sur une liste ne s'applique plus à toutes les victimes mais à des catégories de victimes telles que les victimes de travail forcé ou d'exploitation sexuelle.¹³ Ceci s'applique également à d'autres covariables pertinentes telles que l'âge et le sexe. L'inclusion de covariables clés dans l'analyse présente l'avantage supplémentaire que les résultats de l'estimation peuvent être désagrégés par type d'exploitation, âge et sexe, comme l'exige le suivi des progrès dans la mise en œuvre de l'ODD 16.2.

Malgré l'utilité d'inclure des covariables pour rendre les hypothèses du modèle DSE plus applicables, l'hypothèse problématique d'indépendance d'inclusion dans les deux listes persiste. Heureusement, ce problème est résolu dans la MSE, par l'utilisation de trois listes ou plus. Lorsque trois listes ou plus sont utilisées, il est possible de déterminer pour toutes les personnes de la liste 3, si et dans quelle mesure, la probabilité de leur placement sur la liste 2 (par exemple des foyers d'accueil) est co-dépendante de leur présence sur la liste 1 (par exemple la police). En incluant ces interactions entre listes dans le modèle d'estimation, il est possible de corriger

12. La dépendance entre les listes peut être basée sur le fait que des sous-groupes présentant certaines caractéristiques sont plus susceptibles d'être placés sur deux listes que d'autres. Cette « dépendance apparente » disparaît lorsque ladite caractéristique est incluse dans l'analyse en tant que covariable.

13. Dans l'exemple de la capture et la recapture de poissons, l'hypothèse sous-jacente de l'estimation était, comme nous l'avons dit, que les poissons non marqués dans l'étang ont à peu près la même probabilité d'être capturés que les poissons marqués. Dans la plupart des cas, cela ne peut pas être pris pour acquis car il peut y avoir diverses espèces de poissons nageant dans l'étang avec différentes probabilités d'être capturées par les pêcheurs. En incluant la covariable « type d'exploitation », la procédure d'estimation peut être appliquée séparément à différents types de poissons afin d'obtenir une meilleure estimation des chiffres réels dans l'ensemble.

les dépendances connues entre certaines listes dans l'estimation. Dans la troisième section, nous montrerons comment cela a fonctionné dans la MSE appliquée sur les données irlandaises sur la traite réparties sur trois listes.¹⁴

L'hypothèse d'indépendance dans la DSE est considérée comme un obstacle majeur, car la violation de cette hypothèse peut conduire à des estimations sérieusement biaisées. Pour cette raison, dans ce manuel, nous adoptons la position de conseiller de n'appliquer que la MSE, c'est-à-dire d'utiliser plus que deux listes. Nous conseillons également d'inclure des covariables, lorsqu'elles sont disponibles.

Le «statut officiel» des estimations basées sur la MSE

Comme expliqué ci-dessus, la MSE utilise idéalement les données de trois listes administratives officielles de victimes, ou plus, pour estimer la taille totale d'une population cachée. En utilisant les données administratives existantes pour estimer le «chiffre caché» des victimes non observées, le chercheur adopte les définitions juridiques et opérationnelles utilisées par les principales institutions impliquées. C'est particulièrement le cas si les listes administratives font partie d'un Mécanisme national d'orientation ou d'un autre ensemble de dispositifs de coopération interinstitutions pour l'aide aux victimes. Généralement, ces mécanismes de coopération fournissent des conseils aux praticiens de toutes les institutions impliquées concernant les définitions opérationnelles de la traite des êtres humains, sous la forme de listes de «signaux» d'un cas possible de traite d'êtres humains. Un excellent exemple de telles listes est la liste internationale d'indicateurs conçue par un groupe de travail conjoint de la Commission européenne et de l'OIT (OIT, 2009).

Du point de vue des politiques publiques, l'utilisation de définitions et d'indicateurs opérationnels officiellement approuvés des victimes de la traite des êtres humains est l'un des points forts de la MSE. Contrairement aux estimations basées sur des enquêtes, généralement basées sur des définitions opérationnelles conçues par des chercheurs et interprétées par les répondants individuels eux-mêmes, les estimations basées sur la MSE reflètent les définitions opérationnelles approuvées par l'État de chaque pays telles qu'interprétées par des fonctionnaires spécialisés, le personnel travaillant pour des ONG ou des chercheurs adoptant les définitions généralement admises. Les estimations de la MSE montrent le nombre de victimes qui seraient enregistrées si toutes les institutions impliquées dans la lutte contre la traite des êtres humains dans le pays étaient en mesure de mettre en œuvre de

14. Pour une étude de l'impact des références entre les listes sur la MSE, voir également Jones et al. 2014.

manière optimale leurs mandats existants de lutte contre la traite des êtres humains tels qu'officiellement définis dans le pays.

L'expérience de ces dernières années a montré que les estimations basées sur la MSE étaient souvent deux à quatre fois plus élevées que le nombre de victimes actuellement enregistré (pour un aperçu des résultats, voir le tableau 4 de la section 4). De telles estimations permettent aux décideurs de fixer des objectifs d'action améliorée qui peuvent être atteints de manière réaliste dans un avenir prévisible. En plus de fournir une évaluation réaliste de ce que l'amélioration des politiques nationales de lutte contre la traite pourrait accomplir dans l'ensemble, les résultats donnent un aperçu des catégories de victimes de la traite qui sont restées ces dernières années le plus cachées et devraient donc être prioritaires. Dans la première étude néerlandaise, utilisant des covariables, les victimes de travail forcé et les enfants victimes se sont avérés relativement moins susceptibles d'être détectés/enregistrés que d'autres catégories de victimes (ONU DC, 2017). Ces résultats fournissent des orientations factuelles aux décideurs concernant les priorités en matière de la détection ou la sensibilisation des prestataires de services.¹⁵ Pris ensemble, les résultats des études de MSE permettent aux décideurs politiques de rendre leurs politiques de lutte contre la traite axées sur les résultats, notamment en suggérant des objectifs de performance quantitatifs réalisables par type d'exploitation ou catégorie de victimes pour les années à venir.

Problèmes de protection des données

La conduite de la MSE repose sur la disponibilité d'au moins trois listes administratives appropriées concernant les victimes de la traite et sur la possibilité de déterminer celles qui figurent sur plusieurs listes (appariement). Étant donné que le statut d'une victime de la traite constitue une donnée personnelle sensible, par exemple dans le cas de victimes en situation irrégulière et/ou impliquées dans la prostitution illégale ou d'autres activités criminelles, l'échange de données personnelles sur les victimes de diverses institutions nécessite une attention particulière aux soucis de protection des données. Bien que les contextes et la législation varient, une base juridique claire pour l'échange et la collecte de données doit être établie. Une base

15. Aux Pays-Bas, un nouveau plan de politiques publié en 2019 a accordé une plus grande priorité à la détection des victimes enfants et des victimes du travail forcé (voir ministères de la justice et de la sécurité, des affaires sociales et de l'emploi, de la santé, du bien-être et des sports, et des affaires étrangères, gouvernement des Pays-Bas, *Together against human trafficking: An integrated program to tackling sexual exploitation, labor exploitation and criminal exploitation* (*Ensemble contre la traite des êtres humains: Un programme intégré de lutte contre l'exploitation sexuelle, l'exploitation par le travail et l'exploitation criminelle*), février 2019 (en néerlandais))

juridique courante et souvent préférée pour la collecte et l'échange de données personnelles des victimes de la traite dans les structures d'assistance est le consentement éclairé de la personne enregistrée ou, dans le cas de mineurs, de leur tuteur. En pratique, ce consentement peut parfois être difficile à obtenir. Les victimes ou les tuteurs peuvent avoir du mal à comprendre ses implications. En outre, il peut être difficile, voire déconseillé, de contacter les victimes pour obtenir leur consentement au partage de leurs données à des fins de recherche après qu'elles ont cessé d'être en contact avec des ONG ou autres, soit en raison d'information d'adresse manquante, soit en raison du risque que la demande tombe entre les mains d'un tiers. Lorsque des données sont collectées et partagées sur d'autres bases que le consentement, il convient d'examiner s'il y a des implications éthiques, en mettant particulièrement l'accent sur le fait de ne pas saper la confiance entre les victimes de la traite et les prestataires de services/autorités.

Dans les pays où on produit régulièrement des statistiques complètes sur les victimes de la traite des êtres humains, les données sont rassemblées auprès de différentes institutions publiques et privées. Pour éviter le double comptage des personnes enregistrées par plus d'une institution, l'ensemble de données comprend généralement soit des noms complets, soit des identifiants uniques tels que les données de naissance et les premières initiales. Pour répondre aux préoccupations en matière de protection des données, plusieurs pays ont mis en place des institutions publiques ou des ONG dédiées en tant que centres d'échange de données sur les victimes de la traite des êtres humains provenant de différentes sources, qui opèrent sous des régimes stricts de sécurité et de protection des données. Dans certains pays, par exemple la Roumanie, la Serbie, la Slovaquie et le Portugal, ces agences ont été dotées d'un statut légal. Dans les pays où de telles agences sont opérationnelles, les ensembles existants de données multi-sources peuvent être anonymisés après appariement et partagés avec les chercheurs à des fins de MSE. En l'absence d'agences de ce type, des équipes de recherche de confiance doivent être chargées de collecter et de traiter des données provenant de diverses sources en utilisant des identifiants uniques tels que la date de naissance et les initiales pour permettre la correspondance. Cette procédure a été suivie en Australie (Lyneham et al., 2019) et dans certaines études locales aux États-Unis (Farrell et al., 2019 ; Anderson et al., 2019). Pour se conformer aux exigences de protection des données, les identifiants peuvent également être cryptés, la « clé » n'étant connue que de l'agence contributrice et de l'équipe de recherche. Une autre option consiste à impliquer un tiers de confiance, tel qu'une autorité statistique ou un notaire public, dans la collecte et l'appariement des données sur les victimes enregistrées sur différentes listes.¹⁶ Dans ce scénario, les identifiants uniques peuvent être supprimés une

16. Le cryptage a été appliqué dans des études locales aux États-Unis (Farrell et al., 2019). Aux Pays-Bas, les institutions locales de deux villes, Utrecht et Ede, ont été invitées par une équipe de recherche

fois la correspondance effectuée et les chercheurs recevront un ensemble de données entièrement anonymisé pour analyse. Dans la section suivante, les mesures visant à assurer le respect des réglementations en matière de protection des données seront discutées plus en détail.

POINTS DE SYNTHÈSE

- Étant donné que les statistiques sur les cas enregistrés de victimes de traite reflètent en grande partie les efforts de détection et d'enregistrement des institutions compétentes, elles sont erronées en tant qu'indicateur de l'ampleur réelle du problème dans un pays et ne peuvent pas être utilisées de manière fiable à des fins de planification nationale, de définition d'objectifs ou de rapports sur les progrès de mise en œuvre de l'ODD 16.2 (éradication de toutes les formes de traite des êtres humains) auprès des Nations Unies.
- Une option testée pour estimer la prévalence réelle consiste à mener des enquêtes auprès de la population générale sur les expériences personnelles d'exploitation/traité des êtres humains. De telles études sont coûteuses, en particulier lorsqu'elles sont répétées périodiquement pour effectuer le suivi de tendances et tendent à produire des estimations de prévalence qui vont bien au-delà des chiffres enregistrés.
- Comme alternative, il semble utile d'envisager l'application de l'estimation de systèmes multiples (MSE) aux ensembles de données (potentiellement) disponibles sur les victimes enregistrées recueillies par les forces de l'ordre, les procureurs et les organisations du secteur privé fournissant des services aux victimes. De telles estimations peuvent être faites moyennant des coûts supplémentaires modestes. Elles fournissent une estimation du nombre de victimes de la traite des êtres humains officiellement définies, qui peuvent être secourues et soutenues de manière réaliste lorsque les capacités institutionnelles seront améliorées et/ou les priorités redéfinies.

à télécharger des données sur les cas de traite des êtres humains directement sur un site Web sécurisé géré par le Bureau central des statistiques des Pays-Bas. Après la mise en relation des ensembles de données fournis par ces institutions, le CBS a fourni un ensemble de données intégré anonymisé à l'institut de recherche Regioplan pour une analyse plus approfondie (Buimer et al., 2021).

- Bien que l'estimation soit possible en utilisant seulement deux listes distinctes de victimes, ce qui conduit à une estimation de double système, nous conseillons à MSE d'utiliser un minimum de trois listes distinctes de victimes, par exemple celles des forces de police, des inspections du travail, des organisations d'aide aux victimes et autres (par exemple, garde d'enfants services, cabinets d'avocats, organisations internationales). La raison de cet avis est que les hypothèses pour la MSE sont plus facilement remplies que pour l'estimation de double système. Il n'est pas nécessaire d'harmoniser complètement les critères d'inclusion de ces listes tant que chacune reflète les définitions opérationnelles de la traite en vigueur dans le pays.
- La plupart des bases de données administratives font des distinctions entre le type d'exploitation, le sexe, l'âge et la nationalité. Dans la mesure du possible, ces covariables doivent être incluses dans l'analyse, car cela rendra les estimations des victimes techniquement plus robustes. Étant donné que les estimations peuvent ensuite être différenciées selon les catégories de victimes, elles sont également plus utiles à des fins de politiques publiques, par exemple pour fixer de nouveaux objectifs pour les catégories des victimes le plus cachées (par exemple les mineurs).
- Si des accords multi-institutionnels sont en place pour détecter et soutenir les victimes de la traite et collecter des statistiques multi-sources, la MSE est généralement réalisable avec des implications financières minimales si ce n'est l'embauche d'un analyste qualifié (voir UNODC, 2020). À terme, la MSE pourrait devenir une tâche de rapport régulière pour l'agence nationale responsable des rapports statistiques sur la traite des êtres humains.
- S'il n'y a en place aucun mécanisme national d'orientation ou dispositif similaire pour détecter et soutenir les victimes de la traite, la MSE nécessite l'implication d'une équipe de recherche pour approcher les institutions impliquées et extraire des données sur les victimes présumées des registres administratifs existants en appliquant les définitions opérationnelles de la traite des êtres humains conformément à législation nationale et la jurisprudence (pour un exemple de telles études aux États-Unis, voir Farrell, 2019).
- Aux fins de la MSE, les personnes placées sur différentes listes doivent être comparées les unes aux autres à l'aide d'identifiants uniques tels que la date de naissance et les initiales. Dans certains pays, cette mise en correspondance peut être effectuée par des centres d'information existants mandatés pour rassembler des données sur la traite des êtres humains à des fins statistiques. Dans d'autres pays, les problèmes de protection

des données liés à la mise en rapport des listes doivent être résolus, par exemple en utilisant des mesures de cryptage et/ou en faisant appel à des tiers de confiance tels que les autorités statistiques. Des exemples de cela peuvent être trouvés dans des études menées en Australie (Lyneham et al., 2019), aux États-Unis (Farrell et al., 2019; Anderson et al., 2019) et aux Pays-Bas (Buimer et al., 2021).

SECTION 2

Identification, collecte
et organisation des
données utiles à la MSE

De toute évidence, il n'est pas possible d'appliquer la méthodologie statistique MSE sans des ensembles de données appropriés. Une fois qu'un pays a décidé de mener une étude nationale pour estimer le nombre de victimes de la traite avec la méthodologie MSE, la première tâche consiste à déterminer s'il existe suffisamment de données administratives appropriées disponibles. Comme indiqué, la plupart des pays ayant une réponse décisive à la traite collectent, conservent et publient déjà des données sur les victimes de la traite détectées à partir de différentes sources au niveau national ou local, bien que l'ensemble de données lui-même puisse ne pas être facilement accessible aux chercheurs. Cette partie du manuel abordera les exigences techniques en matière de données et les étapes à suivre pour collecter et préparer les données pour les estimations statistiques.

Clarification des termes

Dans le contexte de la MSE et aux fins de ce manuel, une « victime » de la traite est une personne enregistrée comme telle par les entités/institutions nationales compétentes, en tant que victime présumée ou définitivement identifiée de la traite des êtres humains. Cela implique qu'une « victime » est une personne qui a été victime de la traite et, de ce fait, est entrée en contact d'une manière ou d'une autre avec une entité qui fait partie de la réponse d'un pays à la traite des êtres humains. De plus, la personne a été enregistrée comme telle par l'entité comme répondant à certains critères communément acceptés.¹⁷

Une « liste » est un registre des victimes individuelles de la traite tenu par les entités qui entrent en contact direct avec les victimes. Les listes peuvent varier

17. Les victimes peuvent ne pas avoir le statut officiel de victimes identifiées de la traite mais peuvent être classées, entre autres, comme victimes « présumées », « suspectées », « possibles » ou « probables ».

considérablement en termes de qualité et de quantité d'informations sur chaque victime. Cependant, la plupart des listes comprendront, au minimum, certaines informations individuelles de base (par exemple, nom, numéro de dossier, date de naissance, date d'inscription au registre de l'organisation) ainsi que certaines informations de profil (sexe, nationalité et forme d'exploitation) sur chaque victime rencontrée par l'organisation.

Les « entités contributrices » sont les organisations qui peuvent fournir des données sur les victimes à utiliser pour l'étude, en raison de leur rôle dans la réponse du pays à la traite. Dans la plupart des pays, les entités contributrices les plus pertinentes seront les forces de l'ordre, les entités non gouvernementales (ONG) venant en aide aux victimes de la traite, les inspections du travail et les autorités chargées de l'immigration. Dans de nombreux pays, il existe souvent de nombreux autres contributeurs possibles, notamment des organismes de coordination de la lutte contre la traite, des prestataires d'assistance locaux/régionaux, des prestataires de services aux travailleurs migrants, des procureurs, des organisations internationales telles que l'OIM et des lignes d'assistance téléphonique publiques, pour n'en citer que quelques-uns. Si les nombres sont faibles, ces prestataires secondaires peuvent être regroupés sous une liste des Autres (comme cela a été fait dans l'étude sur l'Irlande présentée dans la section trois).

Collecte de données

Les pays dotés d'un Mécanisme national d'orientation formel encouragent généralement un large éventail d'institutions publiques et privées à attribuer le statut de victime présumée aux personnes atteignant un seuil d'indicateurs opérationnels officiellement établis. Ce statut offre à ces personnes certains droits tels qu'un logement et un permis de séjour temporaire au cas où elles risqueraient d'être expulsées. Dans la plupart de ces pays, le pouvoir d'attribuer le statut définitif de victime *identifiée* bénéficiant d'un ensemble plus large et plus permanent de droits est dévolu à la police nationale. Le Royaume-Uni fournit un exemple de mécanisme national d'orientation entièrement institutionnalisé qui attribue les deux statuts. Dans ce pays, environ un tiers de toutes les victimes présumées sont finalement reconnues comme des victimes définitivement identifiées.

Une question préliminaire à considérer lors de la planification d'une étude nationale est de savoir s'il faut utiliser des données relatives aux victimes *présumées* ou uniquement aux victimes *formellement identifiées*. Dans la plupart des pays, les victimes qui ont été enregistrées uniquement par des ONG sont considérées comme des victimes présumées. Si la méthodologie MSE est appliquée uniquement aux listes de victimes identifiées, les victimes qui n'entrent en contact qu'avec des ONG - par

exemple parce qu'elles ne veulent pas coopérer avec la police - sont exclues de l'analyse, ce qui restreint considérablement la portée de l'estimation. De plus, n'utiliser que les données des victimes formellement identifiées réduirait le nombre de pays où la MSE peut être appliquée, car les données des seules victimes identifiées seront souvent insuffisantes. Pour cette raison, les études MSE dans le domaine de la traite des êtres humains ont jusqu'à présent toujours utilisé des ensembles de données intégrés et multi-sources sur les victimes *présumées* (dont seules certaines ont finalement été formellement identifiées).¹⁸

La sélection des listes appropriées

Le principal défi pour la mise en œuvre d'une étude nationale avec la méthodologie MSE est d'obtenir des listes de victimes à peu près compatibles d'au moins trois entités différentes.

Pour assurer une fiabilité suffisante des estimations, des informations sur une masse critique de victimes sont nécessaires. Bien qu'il n'y ait, d'un point de vue méthodologique, aucune limite inférieure stricte, les analyses de petits ensembles de données produiront des estimations avec de grandes marges d'erreur, d'une utilité limitée aux fins des politiques publiques. Les études MSE menées jusqu'à présent ont été réalisées avec des ensembles de données contenant des informations sur au moins 150 victimes au total pour une étude d'une seule année, ou, avec des données sur plusieurs années, un minimum de 50 victimes par an. Bien qu'il soit impossible de donner des nombres minimaux, il devrait y avoir au moins un certain chevauchement entre les listes utilisées pour l'analyse. Une option pour augmenter le chevauchement est la collation de listes plus petites en plus grandes.¹⁹ Les chercheurs étudient actuellement l'application de la méthodologie MSE à un petit nombre d'ensembles de données, de sorte que cette recommandation pourrait faire l'objet de modifications à l'avenir.²⁰

Il faut au minimum des microdonnées concernant des personnes identifiables de manière unique contenant des informations sur les entités contributrices qui les ont déclarées au cours d'une année et, de préférence, un ensemble de covariables. Étant donné que les organismes centraux qui collectent ces données cherchent à

18. Théoriquement, il semble possible de réaliser des MSE sur un ensemble de données intégré distinguant les victimes *présumées* des victimes identifiées, mais cela n'a pas encore été fait à notre connaissance (Overstall et al., 2014).

19. Voir Sharifi Far et al., 2020.

20. Voir, par exemple, Chan, Silverman et Vincent, 2021.

corriger leurs totaux agrégés de victimes pour un «double comptage» par les institutions déclarantes, les informations sur l'enregistrement par différentes entités de la même victime sont généralement disponibles dans leurs ensembles de données réguliers. Dans certains cas, cependant, seule la première entité d'enregistrement est consignée dans l'ensemble de données et les informations sur un enregistrement ultérieur doivent être dérivées des dossiers de cas et ajoutées rétroactivement à l'ensemble de données régulier.

Dans les pays où il existe une entité centrale dotée d'un mandat national pour collecter ces données - ce que l'on appelle un «centre d'échange d'informations» sur la traite - cette entité est évidemment la mieux placée pour diriger le travail de collecte de données, en liaison avec les chercheurs chargés de l'étude MSE. De préférence, les analystes MSE assurent la liaison avec un homologue de l'agence faisant office de centre d'informations qui peut expliquer comment l'agence centrale et chaque entité contributrice assemblent et traitent les données.

Comme expliqué dans la première section, la MSE n'exige pas nécessairement des chances d'inclusion égales sur toutes les listes. Si toutes les victimes, ou catégories de victimes, ont une chance égale d'être placées sur au moins une des listes, par exemple celle de la police nationale, d'autres listes peuvent appliquer des critères d'inclusion différents, par exemple en se concentrant uniquement sur les victimes du travail forcé.

Bien qu'il n'y ait pas non plus d'exigence que toutes les listes appliquent exactement les mêmes définitions de la traite des êtres humains, il est conseillé que toutes les listes adhèrent largement aux définitions officielles de la traite des êtres humains du pays. Dans la mesure où il en est ainsi, l'estimation du nombre réel de victimes sera plus convaincante pour les décideurs puisqu'elle est conforme à leur propre définition officielle de la traite.

Comme on le sait, dans presque tous les pays, la définition de Palerme de la traite des êtres humains a été incorporée dans la législation nationale. Cette définition a été opérationnalisée dans des listes de « signaux » qui peuvent être reconnus par les travailleurs de première ligne et utilisés pour identifier les cas de victimisation présumée par la traite des êtres humains. Un excellent exemple d'une telle liste a été élaboré par l'OIT, en collaboration avec un groupe de travail du coordinateur de la lutte contre la traite des êtres humains de l'UE (OIT, 2009). Des listes similaires d'indicateurs ou de « signaux » détaillés ont été publiées par un consortium européen financé par l'UE et dirigé par le ministère français des Affaires étrangères (2013) et, plus récemment, par la Croix-Rouge croate (2019). Dans toute l'Europe, des versions de cette liste, adaptées aux situations locales, sont utilisées par les entités fournissant des données aux bases de données nationales centralisées.

Un exemple de la façon dont une entité fournissant des données sur les victimes présumées à un centre d'échange central identifie les victimes présumées avec une liste d'indicateurs opérationnels est fourni par l'ONG néerlandaise Fair Work, qui aide les travailleurs migrants exposés à de mauvaises conditions de travail. Le personnel et les volontaires sont renseignés des cas possibles de traite des êtres humains par un large éventail d'institutions dans le pays et par les victimes elles-mêmes. Il leur est demandé de saisir les cas en tant que victimes présumées de travail forcé/traité d'êtres humains dans la base de données de Fair Work si le cas est coché dans trois cases ou plus de la liste des « mauvaises conditions de travail » ou au moins une case de « mauvaises conditions de travail » et au moins une case de l'une des quatre autres listes (dépendance multiple, limitation de liberté, atteinte à l'intégrité, structure organisée d'exploitation). La figure 2 présente une vue d'ensemble des cinq listes d'indicateurs/cases.

Figure 2 Indicateurs opérationnels du travail forcé utilisés par FairWork Pays-Bas

0 <input type="checkbox"/>	1. Dépendance multiple (=signal)
<input type="checkbox"/> Oui	<i>L'employeur/l'exploiteur organise également le logement, l'habillement, le transport, etc.</i>
	<input type="checkbox"/> N'ont pas eux-mêmes organisé le voyage, le visa, etc.
	<input type="checkbox"/> Recrutés dans le pays d'origine
	<input type="checkbox"/> Travaille avec un passeport faux ou falsifié
	<input type="checkbox"/> N'a pas d'espace de vie propre aux Pays-Bas
	<input type="checkbox"/> Logement lié à l'emploi (utilisation obligée d'un logement proposé par une agence pour l'emploi, contraint de déménager après avoir été licencié)
	<input type="checkbox"/> Dort sur le lieu de travail
	<input type="checkbox"/> Méconnaissance de sa propre adresse professionnelle
<input type="checkbox"/> Oui	<i>Le client est dans une position vulnérable</i>
	<input type="checkbox"/> Capacité mentale limitée
	<input type="checkbox"/> Mauvaise situation économique dans le pays d'origine
	<input type="checkbox"/> Séjourne/travaille illégalement aux Pays-Bas
	<input type="checkbox"/> Le client ne parle pas néerlandais (ou anglais)
<input type="checkbox"/> Oui	<i>Le client est endetté</i>
	<input type="checkbox"/> Dettes envers quelqu'un d'autre que l'employeur
	<input type="checkbox"/> L'employeur a payé un montant pour le transfert de l'employé
<input type="checkbox"/> Oui	<i>N'est pas en mesure de contrôler (une partie de) ses propres revenus</i>
	<input type="checkbox"/> Ne possède pas son propre compte bancaire/carte de débit
	<input type="checkbox"/> L'employeur oblige l'employé à donner accès à son compte bancaire

Obligation d'acheter le transport, la nourriture et/ou d'autres services auprès d'un seul fournisseur

Oui Autre, à savoir :

TOTAL

2. Une forte limitation de la liberté fondamentale de la personne concernée (=signal)

Oui La victime peut ou non être en contact avec le monde extérieur.

Oui La victime n'a pas de liberté de mouvement indépendante

Oui Refus de l'aide médicale

Oui La victime n'est pas en possession de ses propres papiers d'identité

Oui L'employé est lié à l'employeur par une dette

Oui Activités criminelles forcées

Oui Autre, à savoir :

Total

3. Travail dans de mauvaises conditions (=signal)

Oui Le contrat n'est pas de qualité adéquate

Pas de contrat de travail

Pas de copie du contrat de travail

Le contrat de travail n'est pas dans la langue du migrant

Seulement un contrat de travail verbal

Oui Le travail diffère des promesses des employeurs

Pas ou moins de travail : travail sur la base d'un contrat zéro heure, fonctionnement en stand-by

Salaire moins élevé que promis

Oui Les droits du travail ne sont pas respectés

Beaucoup d'heures supplémentaires au travail (par jour ou par semaine)

Les heures supplémentaires ne sont pas payées

Tricherie dans l'enregistrement des heures

Pas de jours de congé (réguliers/vacances)

Licenciement sans motif

Salaire inférieur au salaire minimum officiel

Salaire payé en espèces

A seulement reçu de l'argent à l'avance

Pas de chèques de paie

Chèques de paie manuscrits

Pas de relevé annuel

Pas de versement de primes de vacances

Pas de paiement en cas de maladie

<input type="checkbox"/>	Travail forcé en cas de maladie ou travail non adapté pendant la grossesse
<input type="checkbox"/>	Pas de paiement de la dernière période avant le licenciement
<input type="checkbox"/>	Paiement du salaire juvénile lorsque l'employé a plus de 23 ans.
<input type="checkbox"/>	Paiement basé sur des salaires incitatifs
<input type="checkbox"/>	Pas de primes pour les heures supplémentaires, les vacances ou les heures de travail irrégulières
<input type="checkbox"/> Oui	<i>Des salaires restreints</i>
<input type="checkbox"/>	Amendes (en cas de renvoi, stationnement etc.)
<input type="checkbox"/>	Pour la scolarité
<input type="checkbox"/>	Pour le BSN (numéro d'identification du citoyen)
<input type="checkbox"/>	Pour les tenues de travail/l'équipement/les formations etc.
<input type="checkbox"/>	Pour l'ouverture de comptes bancaires
<input type="checkbox"/>	Sans raisons claires
<input type="checkbox"/>	Pour le transport depuis/vers le pays d'origine
<input type="checkbox"/> Oui	<i>Travail dans des conditions dangereuses ou insalubres</i>
<input type="checkbox"/>	Dommmages physiques ou mentaux causés par le travail
<input type="checkbox"/>	Travaille avec des substances dangereuses ou des matériaux vieillis/en mauvais état
<input type="checkbox"/>	Doit exécuter des activités (dangereuses) qui ne sont pas convenues
<input type="checkbox"/>	Aucune instruction
<input type="checkbox"/>	Aucune mesure de protection en cas de circonstances dangereuses ou insalubres
<input type="checkbox"/> Oui	<i>Mauvais logement</i>
<input type="checkbox"/>	La qualité du logement ne remplit pas les normes officielles
<input type="checkbox"/>	Prix de location très élevés
<input type="checkbox"/>	La maison est trop petite pour le nombre de personnes qui y vivent
<input type="checkbox"/>	Refus de la vie privée
<input type="checkbox"/>	Pas de système de chauffage
<input type="checkbox"/>	Installations sanitaires insuffisantes
<input type="checkbox"/> Oui	<i>Traitement inégal</i>
<input type="checkbox"/>	Salaire bas par rapport aux collègues
<input type="checkbox"/>	La CAO (convention collective de travail) n'est pas appliquée
<input type="checkbox"/>	Travail dans des conditions différentes (pires) de celles des employés néerlandais
<input type="checkbox"/> Oui	<i>Des primes d'assurance nationale restreintes</i>
<input type="checkbox"/>	En cas d'invalidité non enregistrée par l'UWV ou de licenciement immédiat après déclaration d'invalidité/accidents du travail
<input type="checkbox"/>	Pas de distribution des primes d'assurance nationale
<input type="checkbox"/>	Pas d'assurance maladie, ou pas d'information sur l'obligation de s'assurer
<input type="checkbox"/>	Il paye des primes à l'employeur mais n'est pas assuré

- Oui *Formes extrêmes de protection/contrôle*
- Bâtiments avec caméras (également à l'intérieur), gardes du corps, etc.
- Oui *Autre, à savoir :*

Total

4. Atteinte à l'intégrité physique ou mentale de la personne concernée (=signal)

- Oui *Intimidation sexuelle*
- Oui *Exécute des actes sexuels comme condition pour obtenir un emploi*
- Oui *Menacé ou confronté à la violence*
- Oui *Peur des collègues et/ou du manager, harcèlement etc.*
- Oui *Discrimination, à savoir...*
- Oui *Autre, à savoir :*

Total

5. L'exploitation n'est pas fortuite mais il existe un modèle ou une sorte de structure organisée (=signal)

- Oui *En structure organisée avec des agences d'emploi, de travail ou de voyage*
- Oui *L'exploitation est systémique, de nouvelles personnes sont recrutées encore et encore*
- Oui *La durée de l'exploitation a été longue (plus d'un an)*
- Oui *Autre, à savoir :*

En Europe, la création de bases de données multisources centralisées sur les victimes de la traite est activement encouragée par le Conseil de l'Europe (avec 40 États membres) et l'Union européenne (27 États membres). Cela signifie que les opportunités de mise en œuvre de la MSE augmentent progressivement dans cette région. Dans la plupart des pays en dehors de l'Europe, il n'existe souvent pas encore de listes consolidées de victimes gérées par un « centre d'informations » central, et encore moins au niveau national ou fédéral. Dans de nombreux pays, certaines données sur les victimes sont collectées par diverses organisations telles que la police et les ONG, mais ne sont pas rassemblées en un seul ensemble de données, géré par un « centre d'informations » dédié. L'intégration des données sur les victimes de la traite nécessite de niveaux élevés de coopération, d'harmonisation et de confiance qui peuvent être particulièrement difficiles à atteindre dans les pays dotés de structures étatiques fédérales en Amérique du Nord et du Sud. Dans ces pays, les études sur les MSE sont encore réalisables mais nécessitent une étape préparatoire au cours de laquelle les données existantes provenant de diverses entités sont collectées, standardisées et consolidées par des équipes de recherche. Des démarches vers un partage régulier des données sur les victimes de la traite ont notamment été entreprises au Brésil (Human Trafficking Data Lab) et au Chili (Quinteros Rojas, Dufrax Tapia & Ramos Rodríguez, 2019).

Lorsque les données sur les victimes enregistrées doivent être collectées aux fins de la MSE, il est essentiel que toutes les institutions contributrices utilisent les mêmes identifiants uniques. Ces études doivent donc être fondées sur des accords entre les entités contributrices pour partager des données personnelles avec une équipe de recherche, dans le respect des normes nationales et internationales de protection des données. De telles études MSE en plusieurs étapes ont jusqu'à présent été menées dans trois endroits non nommés aux États-Unis (Farrell et al., 2019) et, concernant les enfants victimes d'exploitation sexuelle, dans l'État de l'Ohio (Anderson et al., 2019)²¹. En Australie, une étude MSE a été menée au niveau fédéral dans laquelle les données devaient d'abord être collectées auprès de diverses agences par une équipe de recherche dédiée (Lyneham, Dowling et Bricknell, 2019).

Pour persuader différentes entités de partager leurs données sensibles sur les victimes, il peut être nécessaire que les hauts fonctionnaires ou les politiciens assument un rôle de leadership. En Australie, par exemple, le ministre de la Justice a chargé l'Institut australien de criminologie de mener une étude MSE sur la prévalence de la traite des êtres humains et de collecter les données nécessaires à cette fin.

Comme indiqué, les entités enregistrant des données sur les victimes présumées appliquent généralement des listes de contrôle d'indicateurs opérationnels de la traite des êtres humains selon la définition commune de Palerme. L'expérience a appris que les praticiens expérimentés travaillant dans l'application de la loi, l'aide aux victimes ou autres partagent une compréhension de base des indicateurs opérationnels de ce qu'est une victime présumée de la traite. Pourtant, les indicateurs opérationnels des diverses institutions peuvent différer à certains égards, reflétant leurs contextes et perspectives professionnels uniques. Les responsables de l'application des lois, par exemple, ont tendance à se concentrer un peu plus sur les indicateurs d'une éventuelle enquête criminelle réussie sur les trafiquants et les prestataires de soins visant les symptômes chez les victimes des effets du crime. En outre, les responsables de la police aux frontières examineront nécessairement les indicateurs visibles pendant le bref moment où ils peuvent interroger les non-ressortissants entrant dans le pays. Leur situation diffère de celle d'un policier qui interroge des victimes dans un commissariat ou d'un travailleur social qui accompagne une victime sur une plus longue période. Par conséquent, les entités peuvent parfois « rater » une victime présumée enregistrée par une autre entité en raison de critères différents ou d'informations manquantes dans des cas individuels. De telles incongruités entre certaines des listes n'empêchent pas, comme on l'a dit, la conduite d'une MSE valide.

21. Dans l'étude de l'Ohio, un décompte total des victimes enregistrées contrôlant le double comptage a été effectué, mais aucune MSE n'a été effectuée.

En principe, les listes peuvent provenir de n'importe quelle entité détectant et enregistrant les victimes de la traite, et de n'importe quelle combinaison de sources. Pour des raisons méthodologiques, peu importe qu'un pays utilise, par exemple, une liste d'application de la loi, une liste d'inspection du travail et une liste d'assistance téléphonique publique, ou, alternativement, trois listes de différentes ONG d'aide aux victimes, ou toute autre combinaison de listes.

Bien qu'un minimum de trois listes soit vital pour diverses raisons techniques, qui seront discutées ci-dessous, il n'y a pas de nombre maximum de listes. Lorsque des jeux de données contenant plus de quelques centaines d'entrées sont disponibles, l'utilisation de plus de trois listes est en fait recommandable car cela rendra les estimations plus robustes. À titre d'exemple, l'étude britannique a utilisé cinq listes et les études MSE approfondies aux Pays-Bas de 2017 et 2021 six.²² Lorsque plus de six listes sont disponibles, la sélection du modèle peut devenir trop compliquée. Pour éviter cela, certaines des plus petites listes disponibles peuvent être combinées en plus grandes. En général, lorsqu'il existe plusieurs listes avec peu de victimes, celles-ci peuvent être regroupées en une liste plus large, à condition que le nombre total de listes à utiliser pour l'étude reste au minimum de trois.²³

Dans certains pays, une entité peut tenir à jour une liste complète des victimes, et ces listes contiennent généralement également des indications sur quelle institution est initialement entrée en contact avec les victimes individuelles répertoriées. Sur la base de ces informations, la liste peut être subdivisée en différentes entités contributrices.²⁴

22. Il s'agissait de listes de la police nationale, de la police des frontières, des inspecteurs du travail, des coordonnateurs régionaux de la traite, des centres de traitement résidentiels et des refuges, et d'une liste «autres» qui comprenait, par exemple, les victimes qui étaient entrées en contact avec des centres de soins non résidentiels, des organisations fournissant des conseils juridiques et le service néerlandais de l'immigration et de la naturalisation.

23. Pour une enquête sur la robustesse de la combinaison et l'omission de listes, voir Sharifi Far et al., 2021.

24. Par exemple, en Serbie, le Centre pour les victimes de la traite des êtres humains tient une liste complète de toutes les victimes enregistrées par une ou plusieurs institutions. En remontant aux dossiers individuels des victimes disponibles, le Centre a réussi à désagréger la liste en trois listes distinctes d'institutions faisant office de points de contact primaires ou secondaires.

L'exigence de listes indépendantes

Une hypothèse sous-jacente à la méthodologie d'estimation de systèmes doubles, c'est-à-dire l'estimation à systèmes multiples de seulement deux listes est, comme expliqué, que l'inclusion dans la liste 1 est indépendante de l'inclusion dans la liste 2. Cependant, comme indiqué ci-dessus, dans le cas des victimes de la traite des êtres humains, les listes des victimes ne peuvent généralement pas être considérées comme indépendantes, car le fait d'apparaître sur une liste augmente souvent la probabilité de la victime d'apparaître sur une autre. Cela s'explique par le fait que les victimes sont, par exemple, d'abord détectées par les forces de l'ordre et ensuite orientées vers des ONG pour obtenir de l'aide, ou qu'elles s'adressent à des ONG pour obtenir de l'aide et sont ensuite orientées vers des agences gouvernementales. Il est également possible que les victimes placées sur une liste soient moins susceptibles d'être inscrites sur une autre liste (dépendance négative). L'exigence d'inclure une troisième liste (ou plus) dans le modèle statistique est une stratégie pour gérer de telles dépendances entre les listes.²⁵ Lorsqu'une troisième liste est incluse, les modèles utilisés dans la MSE permettent d'inclure des dépendances par paires entre les trois listes. Ainsi, une éventuelle dépendance peut être corrigée dans la procédure d'estimation.

Bien que l'hypothèse d'indépendance entre les listes puisse être assouplie en utilisant au moins trois listes, cela n'implique pas que la dépendance entre les listes soit sans conséquence pour la MSE. Une étude pilote avec des données provenant de listes de victimes entièrement basées sur des références a montré que ces données ne conviennent pas aux MSE.²⁶ La leçon clé est que les institutions qui fournissent des listes de victimes peuvent accepter des références les unes des autres, mais chacune doit également agir comme un point de contact principal distinct pour les victimes de la traite.

25. Ceci est expliqué plus en détail dans la section 3, Modélisation statistique.

26. Le test consistait à utiliser les données d'une base de données sur les victimes assistées gérée par l'Organisation internationale pour les migrations (OIM) au Bélarus, qui contenait des informations sur les différentes institutions qui avaient orienté les victimes vers le programme d'assistance de l'OIM. Grâce à un processus de séparation des listes, les victimes ont été identifiées sur la base des institutions qui les avaient référées à l'OIM, y compris celles qui avaient été référées par plus d'une institution. Bien que le modèle final ait produit une estimation raisonnablement stable, l'ensemble de données utilisé a été jugé insuffisamment conforme à l'hypothèse d'indépendance de la liste, et les résultats de l'estimation n'ont pas été publiés.

Les identifiants uniques

L'essentiel de la méthodologie MSE est d'établir les chevauchements entre les différentes listes de victimes et d'utiliser ces informations pour estimer le nombre de victimes qui n'apparaissent sur aucune liste. Pour ce faire, les analystes doivent déterminer sur combien de listes différentes apparaît la même victime. Cela se fait par un processus de 'matching', ou couplage d'enregistrements individuels, qui précise - victime par victime - le nombre de listes sur lesquelles chaque victime apparaît.

Pour que l'appariement soit effectué avec précision, un identifiant unique est requis pour chaque victime individuelle figurant sur l'une des listes. Des identifiants uniques comme la date de naissance ou le nom sont souvent présents dans les ensembles de données des institutions concernées. Dans plusieurs pays européens, des dispositifs légaux ou consensuels ont été mis en place pour la mise en commun des données sur les victimes identifiées. Dans de tels accords, les agences étatiques et non étatiques participantes partagent des données sur les victimes identifiées avec un dépositaire. Par exemple, aux Pays-Bas, les agences étatiques et non étatiques partagent ce type de données avec une organisation non gouvernementale dédiée et financée par l'État, CoMensha. Au Portugal, en Roumanie, en Serbie et en Slovaquie, entre autres, ce partage de données avec une institution étatique dédiée, généralement une unité de lutte contre la traite, est, comme indiqué, fondé sur la loi.²⁷

Dans de nombreux pays, les agences étatiques et les agences non étatiques ne sont pas autorisées ou ne consentent pas à partager des données personnelles de nature sensible telles que le statut d'une victime de la traite. Dans ces circonstances, des clés d'identité sécurisées peuvent être construites en concaténant des éléments d'identification. Une formule possible est l'utilisation des premières lettres du prénom et du nom de famille et la date de naissance. Dans une étude aux États-Unis, des chercheurs rassemblant des données provenant de différentes institutions ont utilisé les trois premières lettres du prénom d'un individu, les deux premières lettres de son nom de famille, son mois et son année de naissance et son état de résidence (Farrell, et al., 2019). Une option plus avancée consiste à brouiller les détails des identifiants des victimes. Par exemple, Jeanne Biche; date de naissance : 01/01/1990; numéro de sécurité sociale: 123-45-6789 pourrait être brouillé et recodé sous la forme 19J01D01678990 pour garantir l'anonymat. Un tel cryptage est recommandé par les chercheurs américains lorsqu'ils utilisent des données d'enregistrement sur la traite des êtres humains pour l'estimation de la prévalence (Anderson, Kulig &

27. Par exemple, au Portugal, l'Observatoire sur la traite des êtres humains a été mandaté par le décret/loi n° 229/2008 pour collecter des données sur la traite des êtres humains à partir de diverses sources.

Sullivan, 2019b ; O'Connell et al., 2020). Dans ce scénario, l'accès aux clés de déchiffrement doit bien entendu être strictement réservé aux agences contributrices et au dépositaire.

Si aucun identifiant unique ne peut être utilisé, d'autres informations – lieu, durée et forme d'exploitation, sexe, âge, pays d'origine, statut de résidence, méthode de recrutement, informations sur le trafiquant, lieu d'identification, etc. – peuvent se combiner dans un profil de victime complet et utilisé pour l'appariement,²⁸ même si cela est laborieux dans le cas de grands ensembles de données. Des discordances et des correspondances manquées peuvent se produire, ce qui pourrait affecter les estimations. Cependant, dans les études MSE entreprises à ce jour, l'impact des erreurs d'appariement a été marginal car des identifiants uniques pouvaient être utilisés.²⁹

L'appariement peut être effectué soit par l'entité « centre de partage d'informations », soit par les chercheurs chargés de l'étude MSE. L'appariement peut aussi, comme dit, être fait par un tiers de confiance, tel qu'un Office national de la statistique. Il est important de noter que, bien qu'il soit inévitable que l'entité qui procédera à l'appariement connaisse l'identité des victimes individuelles, ces informations peuvent être supprimées une fois l'appariement effectué. Les analystes chargés de la MSE travailleront normalement avec un ensemble de données entièrement anonymisé (pour un exemple, voir le tableau 1 ci-dessous présentant les données partagées par l'unité irlandaise de lutte contre la traite des êtres humains avec l'université d'Utrecht).

Les ensembles de données utilisés dans les études irlandaise et britannique

Pour illustrer le type d'ensemble de données requis pour les calculs statistiques de la MSE, le tableau 1 ci-dessous présente une partie de la liste consolidée des victimes individuelles collectée par l'Unité anti-traite des êtres humains (AHTU) du ministère irlandais de la justice et de l'égalité. Cet ensemble de données concerne les victimes identifiées en 2014 par trois groupes d'organisations différents : la Police nationale

28. Les champs de données exacts à utiliser dépendent des données capturées de manière cohérente dans toutes les listes de victimes. Il est important d'utiliser suffisamment de champs pour distinguer les profils similaires ; par exemple, une combinaison du sexe, de l'âge, de la nationalité et de la forme d'exploitation ne suffira probablement pas.

29. Pour une enquête sur l'utilisation d'ensembles de données avec un appariement incomplet pour la MSE, voir Sutherland et Schwarz, 2005.

(An Garda Síochána), les organisations de migrants/OIM et les ONG fournissant des services aux victimes. Les deuxième et troisième listes sont toutes deux des listes combinées, fusionnant des listes plus petites disponibles. Dans ce cas, AHTU a obtenu des listes avec des identifiants uniques de chacune de ces organisations, ce qui a permis une correspondance entre les listes. L'ensemble de données partagé avec les analystes était anonymisé et ne comprenait que le numéro de référence de la victime individuelle. Il comprenait également les covariables sexe, citoyenneté, âge (adulte/mineur) et type d'exploitation. La dernière covariable a été rendue dichotomique en distinguant l'exploitation sexuelle des autres uniquement. Les trois colonnes sur le côté droit présentent les résultats de l'appariement, car elles indiquent quelle(s) organisation(s) a/ont identifié les victimes individuelles.

Tableau 1 Échantillon de la liste consolidée des victimes présumées identifiées par la police nationale irlandaise (An Garda Síochána), les organisations de migrants/OIM et les ONG dédiées

Référence	Sexe	Irlandais / Non irlandais	Âge	Exploitation	An Garda Síochána	MRCI, ICI, OIM	Ruhama, SVCC, DL
2014-040	Homme	Non irlandais	Adulte	Travaux	1	1	0
2014-041	Homme	Non irlandais	Adulte	Travaux	1	1	0
2014-042	Homme	Non irlandais	Adulte	Travaux	1	0	0
2014-043	Homme	Non irlandais	Adulte	Travaux	1	0	0
2014-044	Femme	Non irlandais	Adulte	Sexuelle et travaux	1	0	0
2014-045	Femme	Non irlandais	Adulte	Travaux	1	0	1
2014-046	Femme	Non irlandais	Adulte	Sexuelle	1	0	0
2014-047	Femme	Non irlandais	Adulte	Criminalité forcée	0	0	1
2014-048	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-049	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-050	Homme	Non irlandais	Adulte	Criminalité forcée	0	1	0
2014-051	Femme	Non irlandais	Adulte	Mariage forcé	0	0	1
2014-052	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-053	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-054	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-055	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-056	Femme	Non irlandais	Adulte	Sexuelle	0	0	1

Référence	Sexe	Irlandais / Non irlandais	Âge	Exploitation	An Garda Síochána	MRCI, ICI, OIM	Ruhama, SVCC, DL
2014-057	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-058	Femme	Non irlandais	Adulte	Sexuelle	0	1	1
2014-059	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-060	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-061	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-062	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-063	Femme	Non irlandais	Adulte	Sexuelle	0	1	1
2014-064	Femme	Non irlandais	Mineur	Sexuel	0	1	0
2014-065	Femme	Non irlandais	Adulte	Mendicité forcée	0	0	1
2014-066	Femme	Non irlandais	Adulte	Sexuelle	0	1	1
2014-067	Femme	Non irlandais	Adulte	Travaux	0	1	0
2014-068	Femme	Non irlandais	Adulte	Sexuelle	0	0	1
2014-069	Homme	Non irlandais	Adulte	Criminalité forcée	0	1	0

Source : Unité de lutte contre la traite des êtres humains, ministère de la Justice et de l'Égalité, Irlande.

L'ensemble de données irlandais dont un échantillon est présenté dans le tableau 1 montre que plusieurs victimes ont été identifiées par une seule organisation, mais certaines ont été identifiées par deux (et dans d'autres parties de l'ensemble de données, certaines victimes ont été identifiées par les trois). Aux fins des calculs de la MSE, les données sont exprimées dans un tableau de contingence où les chevauchements entre les différentes listes sont quantifiés. Pour une MSE avec trois listes, nommées A, B et C à des fins d'illustration, un tel tableau de contingence comprendrait le nombre d'individus dans la liste A uniquement, la liste B uniquement, la liste C uniquement, les listes A et B, les listes A et C, les listes B et C et les listes A, B et C. Pourtant, il n'y a pas d'individus observés qui ne figurent pas dans les trois listes, et ce nombre est estimé dans la MSE, classé de manière croisée par les combinaisons de covariables si de telles covariables sont disponibles.

Dans l'ensemble de données utilisé pour une étude MSE au Royaume-Uni, les données de plusieurs petites listes ont été regroupées sous cinq listes principales, détaillées ci-dessous :

- LA : autorité locale
- NG : Organisation non gouvernementale
- PF : police/Agence nationale du crime

- GO : organisation gouvernementale (principalement des agences du ministère de l'Intérieur telles que la UK Border Force et la Gangmasters Licensing Authority)
- GP : Le grand public, à travers différents parcours.

Sur les 2 744 victimes incluses dans la base de données de 2013, certaines figuraient sur deux, et quelques-unes sur trois ou quatre des cinq listes. Le tableau de contingence de cette étude est présenté ci-dessous, montrant la répartition résumée des victimes présumées sur les cinq listes (tableau 2).

Tableau 2 Tableau de contingence des victimes présumées de la traite au Royaume-Uni, 2014

LA	X					X	X	X								X	X	X	
NG		X				X			X	X	X					X	X	X	X
PF			X				X		X			X	X			X	X		X
GO				X				X		X		X		X	X			X	X
GP					X						X		X	X					
number	54	453	995	695	316	15	19	3	62	19	1	76	11	8	4	1	1	1	

Source : Agence nationale du crime/Bales, Hesketh et Silverman, 2015)

Utilisation de covariables pour améliorer la robustesse et la pertinence des résultats pour les politiques publiques

Comme expliqué ci-dessus, l'inclusion de covariables - telles que le sexe, l'âge et la forme d'exploitation - dans l'étude est fortement recommandée, pour des raisons à la fois techniques et de politiques publiques. En l'absence de covariables, la MSE permet une dépendance entre les listes, où ces dépendances de liste sont égales pour différents segments de la population. Méthodologiquement, une fois que l'étude comprend des informations sur le sexe, l'âge et la forme d'exploitation des victimes, l'hypothèse d'égalité des dépendances de liste pour toutes les victimes peut être remplacée par l'hypothèse moins exigeante de dépendances de liste égales dans la classification croisée de toutes les covariables. Par exemple, une étude incluant la covariable sexe permettrait aux hommes et aux femmes d'avoir des dépendances de liste différentes. Ces différences peuvent être reconnues dans la procédure d'estimation et être corrigées dans l'estimation totale.

L'expérience néerlandaise a confirmé que l'inclusion de covariables est effectivement conseillée. En plus de rendre l'estimation totale plus robuste - une analyse utilisant des covariables a produit des estimations significativement différentes, dans ce cas des estimations moins importantes qu'une étude plus ancienne n'utilisant pas de covariables - l'inclusion d'informations sur les covariables a permis de calculer des estimations séparées de « nombres cachés » de, par exemple, classifications croisées d'âge (victimes mineures, victimes adultes), de sexe (victimes féminines, victimes masculines) et de type d'exploitation (victimes d'exploitation sexuelle et non sexuelle (toutes les autres catégories)).³⁰

L'indicateur ODD 16.2.2 demande spécifiquement des taux de victimisation par sexe, âge et forme d'exploitation. L'inclusion du sexe et de l'âge comme covariables reflète l'accent particulier mis par le protocole de Palerme sur la protection des femmes et des enfants. Les formes d'exploitation les plus importantes à distinguer peuvent différer d'un pays à l'autre. Dans de nombreux pays, les formes les plus courantes sont l'exploitation sexuelle et le travail forcé. Dans la plupart des études, le type covariable d'exploitation faisait la distinction entre l'exploitation sexuelle et d'autres formes (telles que le travail forcé, la mendicité forcée, la criminalité forcée et le prélèvement d'organes). Dans les études utilisant des données provenant de Roumanie, de Serbie et de Slovaquie, la mendicité/activité criminelle a été incluse en tant que troisième forme distincte d'exploitation en raison de sa prévalence dans l'ensemble de données.

Selon la nature de la traite dans un pays, d'autres covariables peuvent également être incluses dans une étude MSE. L'inclusion de covariables supplémentaires renforce, comme expliqué, la robustesse des résultats et améliore également la pertinence des résultats pour les politiques publiques. Par exemple, dans de nombreux pays, les victimes de la traite détectées comprennent à la fois des citoyens locaux et des étrangers. C'est particulièrement le cas des pays de destination typiques de la traite où les étrangers sont transférés à des fins d'exploitation (en plus de la traite intérieure des citoyens locaux). Dans ces pays, il est recommandé d'inclure la nationalité comme covariable, car les citoyens locaux et les étrangers peuvent être confrontés à différentes formes et niveaux de traite et également à différentes probabilités d'être détectés.

30. Voir Ministères de la justice et de la sécurité, des affaires sociales et de l'emploi, de la santé, du bien-être et des sports, et des affaires étrangères, gouvernement des Pays-Bas, *Together against human trafficking: An integrated programme to tackling sexual exploitation, labour exploitation and criminal exploitation*, février 2019 (en Néerlandais).

De plus, dans les pays qui sont surtout des pays d'origine, les victimes détectées peuvent avoir été exploitées en dehors de leur pays d'origine, et éventuellement rapatriées par la suite. Lorsqu'on travaille avec des ensembles de données de ces pays, il est essentiel d'inclure le territoire d'exploitation dans l'ensemble de données afin de pouvoir estimer séparément le nombre d'exploités sur le propre territoire du pays et le nombre de ressortissants exploités ailleurs.³¹ Si les taux de prévalence dans les pays qui servent principalement de pays d'origine pour les victimes de la traite doivent être comparés à ceux des pays de destination typiques, ils doivent représenter les victimes qui ont été exploitées dans leur pays d'origine. Nous reviendrons sur cette question dans la dernière section.

POINTS DE SYNTHÈSE

- Dans les pays dotés de mécanismes nationaux d'orientation, les données sur les victimes présumées de la traite des êtres humains sont souvent collectées par un bureau central de lutte contre la traite des êtres humains ou une ONG dédiée. Si un tel centre d'échange existe, ses ensembles de données disponibles répondront généralement aux exigences clés pour les MSE énumérées ci-dessous ou peuvent être reconstruits pour le faire.
- S'il n'existe pas d'ensemble de données centralisé et multi-sources, les données doivent être collectées auprès des institutions collectant ces données dans le cadre d'un projet de recherche ad hoc pour développer une base de données intégrée dans le but de mener des MSE.
- Des données doivent être disponibles sur les personnes qui répondent aux indicateurs opérationnels de victimes présumées conformément à la législation nationale, provenant d'au moins trois entités différentes qui entrent en contact avec les victimes de la traite, par exemple, la police, les ONG, les autorités de l'immigration, les inspections du travail et les organisations internationales compétentes.
- Si les données de différentes entités doivent être consolidées, un ensemble de définitions et d'indicateurs opérationnels d'une victime présumée peut devoir être appliqué rétroactivement par l'équipe de recherche pour extraire les cas pertinents à partir des ensembles de données administratives existants.

31. Dans les publications de l'OIT et de la Fondation Minderoo/Walk Free sur la traite des êtres humains, les taux régionaux et nationaux de victimes se rapportent aux victimes exploitées sur les territoires nationaux des régions/pays.

- Les institutions qui fournissent des listes de victimes doivent toutes fonctionner comme des points de contact indépendants pour les victimes de la traite, en plus d'accepter les dépositaires les unes des autres. Par exemple, lorsque la police oriente les victimes identifiées vers une ONG pour l'obtention d'un soutien, cette ONG devrait également entrer directement en contact avec les victimes par d'autres moyens.
- Chaque victime qui apparaît sur une liste quelconque a besoin d'un identifiant unique tel que la date de naissance et les initiales, ou une clé d'identité codée ou cryptée afin de permettre la correspondance entre les listes pour déterminer les chevauchements (soit qu'une victime apparaisse sur une, deux ou plusieurs listes). Pour des raisons de sécurité des données, l'appariement peut également être réalisé par un tiers de confiance qui supprimera ensuite les identifiants.
- Pour arriver à des estimations suffisamment fiables, le nombre total de victimes de la traite dans toutes les listes doit être d'au moins 150, ou lorsque des ensembles de données pluriannuels sont utilisés, d'au moins 50 victimes par an ; il devrait y avoir un chevauchement raisonnable entre les diverses listes.
- Idéalement, l'ensemble de données devrait inclure des données sur les caractéristiques de base des victimes telles que l'âge (adulte ou enfant), le sexe et le type d'exploitation (sexuelle, travail forcé, mendicité forcée/exploitation criminelle). D'autres covariables pertinentes peuvent inclure la nationalité et le lieu d'exploitation (dans le pays d'origine ou à l'étranger dans un pays de destination).
- Étant donné que l'utilisation de plusieurs listes et covariables nécessite, comme cela sera démontré dans la section suivante, un exercice de modélisation statistique avancé, il est conseillé d'impliquer un analyste ayant une expérience préalable de la MSE, de préférence avec des données sur la traite des êtres humains. La Section de recherche sur la criminalité de l'ONUDC est prête à offrir des conseils aux gouvernements intéressés par la conduite d'une MSE pour suivre les progrès accomplis dans la réalisation de l'ODD 16.2.

SECTION 3

Modélisation statistique pour l'estimation de systèmes multiples

Dans cette section, nous présentons d'abord la théorie sous-jacente à l'estimation de systèmes multiples. La MSE est également connue sous une variété d'autres noms, principalement Capture-recapture. Dans notre exposé, nous suivons une étude du Groupe de travail international sur la surveillance et la prévision des maladies (1995).³² Nous complétons la théorie avec le *code langage R*. Une MSE utilisant des données collectées en Irlande servira d'exemple illustratif.³³

Théorie

Bien que l'application MSE pour estimer le nombre de victimes non observées de la traite des personnes devrait toujours, comme indiqué, être basée sur les données d'au moins trois listes, la discussion commencera par un exemple de deux listes pour illustrer le fonctionnement de la méthode.

Dans la MSE, plusieurs listes d'individus sont liées. Une liste fait référence à une liste d'individus qui font partie d'une population commune (dans le contexte de ce manuel : la population des victimes de la traite des êtres humains). Une situation simple est celle de deux listes R1 et R2, toutes deux de niveau 1 lorsqu'elles sont observées dans la liste, et 0 lorsqu'elles ne sont pas observées dans la liste.

32. D'autres références clés sont Van der Heijden et al. (2012), Bird & King (2018) et Boehning, van der Heijden & Bunge (2018).

33. Dans cette élaboration, nous expliquons le choix par étapes du modèle le mieux ajusté tenant compte et spécifiant les interactions entre les listes et les covariables. Pour des approches alternatives de modélisation dans lesquelles les estimations de divers modèles sont moyennées, voir Burnham et Anderson (2002) et Silverman (2020).

Tableau 3 Tableau de *contingence avec deux listes*

		R2	
		1	0
R1	1	n_{11}	n_{10}
	0	n_{01}	n_{00}

Le tableau ci-dessus montre le croisement de deux listes, R1 et R2. Nous dénotons les valeurs attendues par ' m ' et les valeurs observées par ' n '. Après avoir lié les deux listes, il y a n_{11} individus qui sont à la fois dans R1 et R2, il y a n_{10} individus qui sont dans R1 mais pas dans R2, et il y a n_{01} individus qui sont dans R2 mais pas dans R1. Il y a n_{00} individus qui ne figurent dans aucune des listes. Ce nombre est inconnu et à estimer. La partie observée de la population (victimes documentées de la traite) est notée $n = n_{11} + n_{10} + n_{01}$. Le problème statistique est d'estimer la taille de la population totale, que l'on note N . Si nous connaissons n_{00} , cela nous donnerait N comme $N = n + n_{00}$. Par conséquent, nous visons à trouver une estimation pour n_{00} , de sorte que nous ayons une estimation de N .

Nous pouvons trouver une estimation de n_{00} en faisant un certain nombre d'hypothèses. Nous supposons qu'il est possible d'apparier les individus de R1 à ceux de R2 sans faire d'erreur. Deuxièmement, nous supposons que la probabilité d'inclusion dans R1 est statistiquement indépendante de la probabilité d'inclusion dans R2. Liée à cela, il y a l'hypothèse que soit les probabilités d'inclusion pour R1, soit les probabilités d'inclusion pour R2, sont homogènes pour les individus.

L'hypothèse d'indépendance nous permet de trouver une estimation pour n_{00} . Sous indépendance, nous avons que les cotes des valeurs dans les lignes sont identiques aux cotes des valeurs dans les colonnes. Par exemple, pour les lignes, $n_{11} / n_{10} = n_{01} / n_{00}$. En d'autres termes, le rapport de cotes est de 1. Si nous désignons l'estimation de la valeur attendue pour la cellule (0,0) par m_{00} , et une estimation en ajoutant un chapeau '^', alors il s'ensuit que :

$$\hat{m}_{00} = \frac{n_{10}n_{01}}{n_{11}}$$

Pour deux listes, c'est très simple, mais cela devient plus compliqué dans la situation où il y a plus de deux listes et covariables. Par conséquent, nous passons au modèle log-linéaire plus général, qui a le modèle d'indépendance comme cas particulier. Le modèle log-linéaire a l'avantage de pouvoir être facilement étendu à davantage de listes et de covariables. En général, si nous indexons les niveaux de R1 et R2 par i et j respectivement, alors la valeur attendue pour la cellule (i, j) est donnée par

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)}$$

où 'log' représente le logarithme naturel, et nous devons imposer des restrictions pour identifier le modèle. Comme il n'y a que trois comptages observés, seuls trois paramètres peuvent être identifiés. Pour un tableau général 2×2 où une cellule (0,0) est observée, un modèle log-linéaire aura un terme d'interaction $u_{12(ij)}$, mais dans notre contexte où la valeur de la cellule (0,0) n'est pas observée, ce terme doit être ajusté à zéro car il ne peut pas être identifié. De plus, pour identifier les paramètres, nous définissons tous les paramètres égaux à 0 si un indice est 0, nous définissons donc $u_{1(0)} = u_{2(0)} = 0$. Le résultat est que

$$\log m_{00} = u.$$

En d'autres termes, nous ajustons un modèle pour les trois cellules observées et le projetons sur la quatrième cellule (0,0). Une fois que nous avons une estimation pour u , alors nous pouvons insérer cette estimation dans l'équation et nous obtenons

$$\hat{m}_{00} = \exp \hat{u},$$

où $\exp(x)$ est la fonction exponentielle de x . Cette estimation est utilisée comme estimation de n_{00} pour trouver une estimation de N , la taille de la population.

Covariables

Lorsqu'il y a deux listes, R1 et R2, une dépendance statistique des probabilités d'inclusion peut survenir, ce qui violerait une hypothèse clé. Il se peut que certaines victimes soient plus susceptibles d'être incluses dans R1 ainsi que dans R2 que d'autres. Cela peut être dû à leur visibilité. Supposons, par exemple, que le sexe soit lié à cette visibilité, par exemple parce qu'il existe une politique active d'identification des femmes victimes de la traite à des fins d'exploitation sexuelle, et celles-ci sont donc plus facilement retrouvées. Dans ce cas, ne pas inclure le sexe conduirait à une violation de l'hypothèse du modèle. Cependant, si nous incluons le sexe comme covariable et procédions à des estimations distinctes pour les femmes et les hommes, la violation serait moins grave.

Supposons donc que nous ayons deux listes et une covariable, disons le sexe, notée x , avec des niveaux $x = 0, 1$ pour les femmes et les hommes respectivement. Alors le modèle log-linéaire qui peut être ajusté est

$$\log m_{ijx} = u + u_{1(i)} + u_{2(j)} + u_{4(x)} + u_{14(ix)} + u_{24(jx)}$$

Ce modèle suppose que pour les hommes comme pour les femmes, l'inclusion dans la liste R1 et celle dans la liste R2 sont indépendantes, mais que les probabilités d'inclusion des hommes peuvent différer de celles des femmes. Ce modèle serait indiqué par les marges ajustées les plus élevées entre crochets, c'est-à-dire [14][24].

Dans ce contexte, nous devons estimer le nombre de deux cellules, la cellule pour les femmes (0,0,0) et la cellule pour les hommes (0,0,1). Nous identifions à nouveau les paramètres en les mettant à 0 si un indice est 0, et estimons

$$\hat{m}_{000} = \log(\hat{u})$$

pour les femmes et

$$\hat{m}_{001} = \log(\hat{u} + \hat{u}_{4(1)})$$

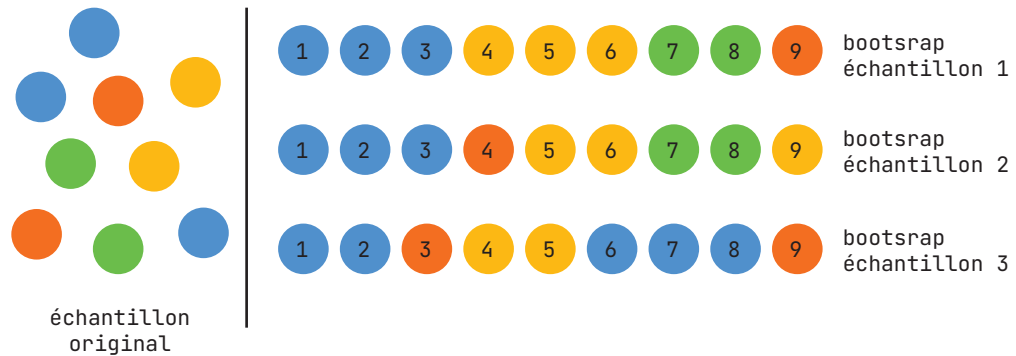
pour les hommes.

Intervalle de confiance pour l'estimation de N

Nous décrivons comment trouver un intervalle de confiance pour l'estimation de N dans le contexte de deux listes, car cela se généralise à plus de deux listes de manière directe. Nous utilisons le bootstrap paramétrique (comparer Groupe de travail international sur la surveillance et la prévision des maladies, 1995). Nous illustrons la façon dont cela fonctionne avec la figure avec les neuf boules colorées ci-dessous (Figure 3).

Nous avons trois comptes observés symbolisés par les boules orange, vertes et jaunes, et nous avons un compte estimé de la cellule (0,0) symbolisé par les boules bleues. La somme de ces quatre chiffres donne la taille estimée de la population. Ces quatre comptes sont utilisés pour calculer quatre probabilités d'une distribution multinomiale. À l'étape 1, nous tirons un premier échantillon bootstrap avec remise en utilisant les quatre probabilités de la distribution multinomiale et l'estimation arrondie de N . Dans la figure 3 ci-dessous, nous avons trois boules bleues, donc à chaque tirage, la probabilité de tirer une boule bleue est de 3/9, et pour chacune des autres couleurs, la probabilité est de 2/9. À l'étape 2, nous éliminons les observations qui sont tombées dans la cellule (0,0), c'est-à-dire les boules bleues dans les échantillons bootstrap. À l'étape 3, nous estimons N pour le premier échantillon bootstrap, c'est-à-dire que nous estimons le nombre de boules bleues compte tenu du nombre de boules vert orange et jaune. Ceci termine la procédure pour le premier échantillon bootstrap. Nous faisons cela, disons, 1 000 fois, et cela nous donne 1 000 estimations d'échantillon bootstrap de N . Nous les classons dans une distribution d'estimations. En prenant les valeurs de 2,5 et 97,5 centiles, cela donne un intervalle de confiance de 95 % pour le N estimé trouvé dans l'échantillon d'origine (voir la figure 3).

Figure 3 Recherche d'intervalles de confiance par bootstrap ; une illustration



Cet intervalle de confiance a l'avantage sur les approches analytiques qu'il n'est pas supposé a priori que l'intervalle de confiance est symétrique. De plus, en incluant cette procédure compliquée pour la cellule (0,0), en l'échantillonnant puis en éliminant le décompte, nous permettons une fluctuation du nombre d'observations n .

Trois listes

Dans la section pour deux listes, nous avons vu qu'il n'y a que trois comptages observés et donc nous ne pouvons ajuster qu'un modèle avec trois paramètres. Cependant, lorsqu'il y a trois listes, le nombre de comptages observés est $2 \times 2 \times 2 - 1 = 7$, où le '-1' représente le comptage qui doit être estimé. Ainsi, un modèle à 7 paramètres peut être adapté. On note la troisième liste R3 d'indice k ($k = 0, 1$). Le modèle log-linéaire le plus général que nous puissions ajuster aux données, c'est-à-dire les sept comptes, est

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$$

Cela montre que le grand avantage d'inclure une troisième liste est que nous ne supposons plus l'indépendance statistique entre les listes car le modèle a trois paramètres d'interaction à deux facteurs, à savoir $u_{12(ij)}$, $u_{13(ik)}$ et $u_{23(jk)}$. Dans des situations pratiques, cela permet d'avoir un modèle plus réaliste qui peut être ajusté. Il y a seulement la restriction qu'il n'y ait pas d'interaction à trois facteurs, ce qui signifie qu'une interaction entre deux listes est identique sur les niveaux de la troisième liste, ou en d'autres termes, que les chances d'inclusion dans deux listes ne dépendent pas de l'inclusion ou l'exclusion d'une troisième liste.

Le modèle est ajusté de la même manière que pour deux listes : nous estimons le nombre de cellules (0,0,0) à l'aide de

$$\hat{m}_{000} = \log(\hat{t})$$

Cette estimation est utilisée comme estimation de n_{000} pour trouver une estimation de N , la taille de la population.

Il est de pratique courante de n'ajuster que les modèles loglinéaires dits hiérarchiques (Agresti, 2013). Cela signifie que, si un terme d'ordre supérieur (ici : une interaction à deux facteurs) est dans le modèle, les termes d'ordre inférieur (ici : les termes marginaux) sont également dans le modèle. (Si de tels termes marginaux font défaut, le modèle deviendrait difficile à interpréter.) Par conséquent, il existe un moyen efficace de désigner les modèles log-linéaires, à savoir en plaçant les termes les plus élevés entre crochets. Par exemple, [12][23] est le modèle log-linéaire où $u_{13(ik)} = 0$.

Sélection du modèle

Bien que l'inclusion d'autant de listes/registres et de covariables que possible soit avantageuse pour respecter les hypothèses d'indépendance et de probabilités d'inclusion égales, il y a un inconvénient. Premièrement, le nombre de modèles potentiels augmente très rapidement avec le nombre de variables. Par exemple, avec deux listes, il y en a un, avec deux listes et une covariable, il y en a quatre, et avec trois listes, il y a huit modèles potentiels parmi lesquels choisir. Ainsi, lorsque le nombre de variables augmente, le problème de trouver le meilleur modèle deviendra de plus en plus difficile. Deuxièmement, avec un nombre croissant de variables, les fréquences dans les cellules du tableau de contingence deviennent de plus en plus petites, ce qui peut conduire à des estimations instables de la taille de la population. Par conséquent, la qualité des estimations dépend dans une large mesure du juste équilibre entre la taille de l'échantillon, le nombre de listes et de covariables et la complexité du modèle.

Il est de pratique courante dans la modélisation log-linéaire d'ajuster une série de modèles dans le but de trouver le modèle le plus restrictif qui s'adapte de manière adéquate aux données. Ce modèle le plus restrictif a la propriété qu'en général, l'intervalle de confiance des estimations sous ce modèle restrictif sera plus petit que l'intervalle de confiance du modèle saturé (Agresti, 2013).

Nous adaptons une série de modèles en utilisant la procédure dite STEP programmée en R . Nous partons du modèle n'ayant qu'une interception, puis commençons

à ajouter des termes log-linéaires marginaux et d'interaction dans des étapes consécutives, où à chaque étape, il est vérifié si, lorsqu'un terme est ajouté, d'autres termes deviennent redondants. Dans le contexte de l'estimation de la taille de la population, deux approches sont préférées pour décider quand s'arrêter et choisir un modèle final. L'une est l'AIC³⁴ et l'autre est le critère BIC³⁵. Les deux sont des fonctions de la déviance d'un modèle, où dans l'AIC il y a une pénalité pour l'ajout de paramètres, et dans le BIC pour l'ajout de paramètres qui impliquent n . Le BIC a tendance à trouver un modèle plus restrictif que celui trouvé dans l'AIC si la taille de l'échantillon est grande.

Un avertissement général est en place ici. Il est possible de se retrouver avec un modèle trop restrictif, et donc avec un intervalle de confiance pour N trop petit. L'intervalle de confiance est optimal sous le modèle auquel on aboutit mais ne tient pas compte de la procédure d'ajustement du modèle. Ce problème a déjà été remarqué dans le Groupe de travail international sur la surveillance et la prévision des maladies (1995) et fait toujours l'objet de recherches actives.

Vérification des données et étapes préliminaires

Dans cette section, l'analyse des données irlandaises servira d'illustration. Cet ensemble de données relativement petit comprend 302 personnes, et le fichier de données Excel d'origine contient un registre pour chaque personne (voir le tableau 4 pour plus de détails).

La première étape consiste à convertir le fichier Excel en un bloc de données appelé «d» pour une utilisation dans R. Les trois listes sont désignées par la lettre R et les chiffres de 1 à 3, et les covariables sont désignées par une seule lettre, y compris les années variables (voir tableau 4).

34. Critère d'information d'Akaike.

35. Critère d'information bayésien.

Tableau 4 Dénotation de l'ensemble de données irlandais

Variable	La description	Niveau de mesure	Niveaux/catégories
R1	An Garda Síochána (police)	Numérique	0 = non observé dans la liste, 1 = observé dans la liste
R2	MRCI, ICI, OIM (aide au migrant)	Numérique	0 = non observé dans la liste, 1 = observé dans la liste
R3	Ruhama, SVCC, DL (ONG)	Numérique	0 = non observé dans la liste, 1 = observé dans la liste
S	Sexe	Factoriel	F = femme, M = homme
A	Âge	Factoriel	A = adulte, M = mineur
N	Nationalité	Factoriel	I = irlandais, N = non irlandais
E	Exploitation (Type d')	Factoriel	S = sexuel, O = autre
Y	Année d'observation	Factoriel	2014, 2015, 2016

Nous fournissons maintenant une liste de contrôle pour la qualité des données, y compris des instructions sur le retour à ceux qui collectent les données pour des clarifications ou des demandes de réorganisation des données.

Sur ce fichier de données nous effectuons les vérifications suivantes :

1. Répartition des listes au fil des années
 - a. vérifier les fluctuations extrêmes au fil des ans
 - b. vérifier les fréquences nulles

2. Chevauchement entre paires de listes
 - a. vérifier un chevauchement suffisant (bon)
 - b. vérifier le chevauchement complet (pas bon)

3. Fréquences des covariables vérifier un étalement suffisant au fil des années

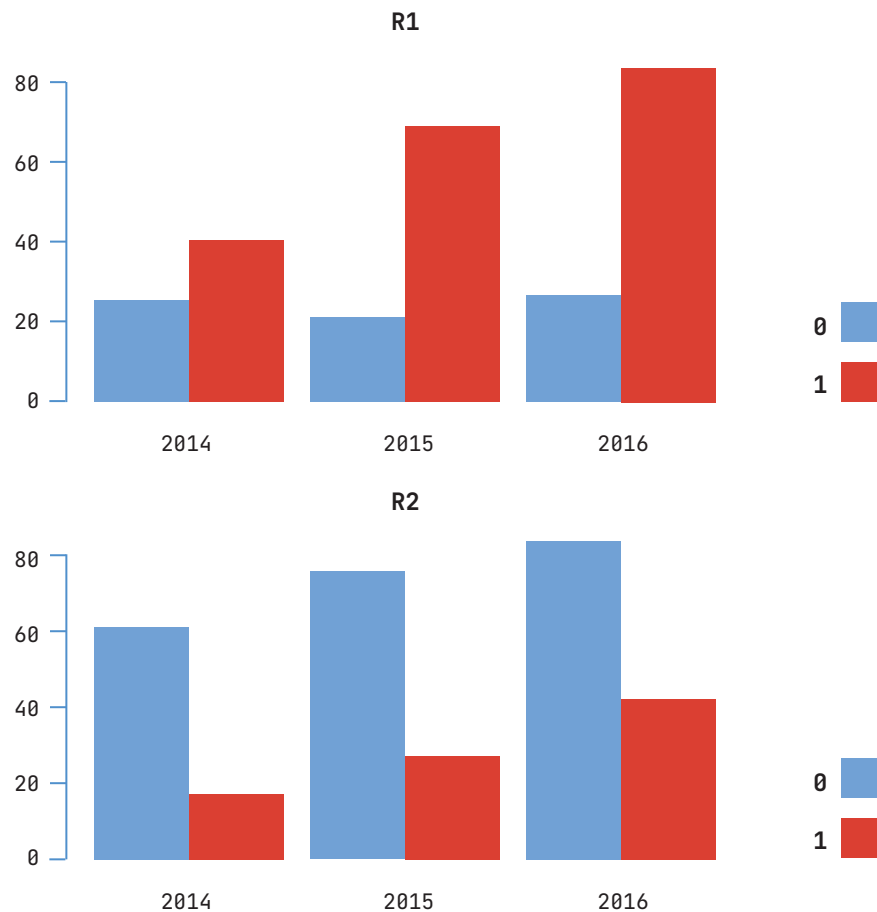
S'il y a un chevauchement insuffisant entre les paires de listes - par exemple, il n'y a pas ou très peu d'observations qui apparaissent dans les deux listes - et/ou la propagation des covariables - par exemple, certaines valeurs de covariables ont une fréquence de 0 ou proche de zéro dans une année particulière - des problèmes numériques peuvent survenir lors de l'estimation des effets d'interaction entre ces listes ou entre les covariables et les années. En règle générale, les estimations des paramètres correspondants prennent de grandes valeurs et/ou ont des erreurs types importantes. En soi, cela ne doit pas conduire à des estimations irréalistes de la taille de la population, mais il peut également arriver que l'estimation de la taille de la

population explose. Chan, Silverman et Vincent (2020) discutent plus en détail de la détection et du traitement des cellules avec des zéros d'échantillonnage.

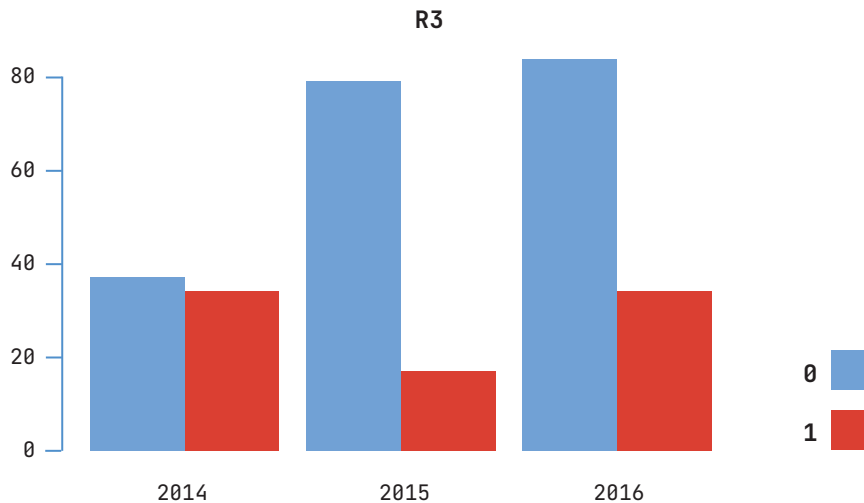
Répartition des listes au fil des années

Nous faisons les affichages graphiques suivants (diagrammes à barres) avec les fréquences observées dans l'ensemble de données irlandais sur chacune des trois listes (Figure 4)³⁶.

Figure 4 Histogramme des fréquences observées sur les listes R1, R2 et R3



36. Code R pour faire le 1^{er} graphique : `barplot(xtabs(~ R1 + Y, data = Ireland), beside = TRUE)`



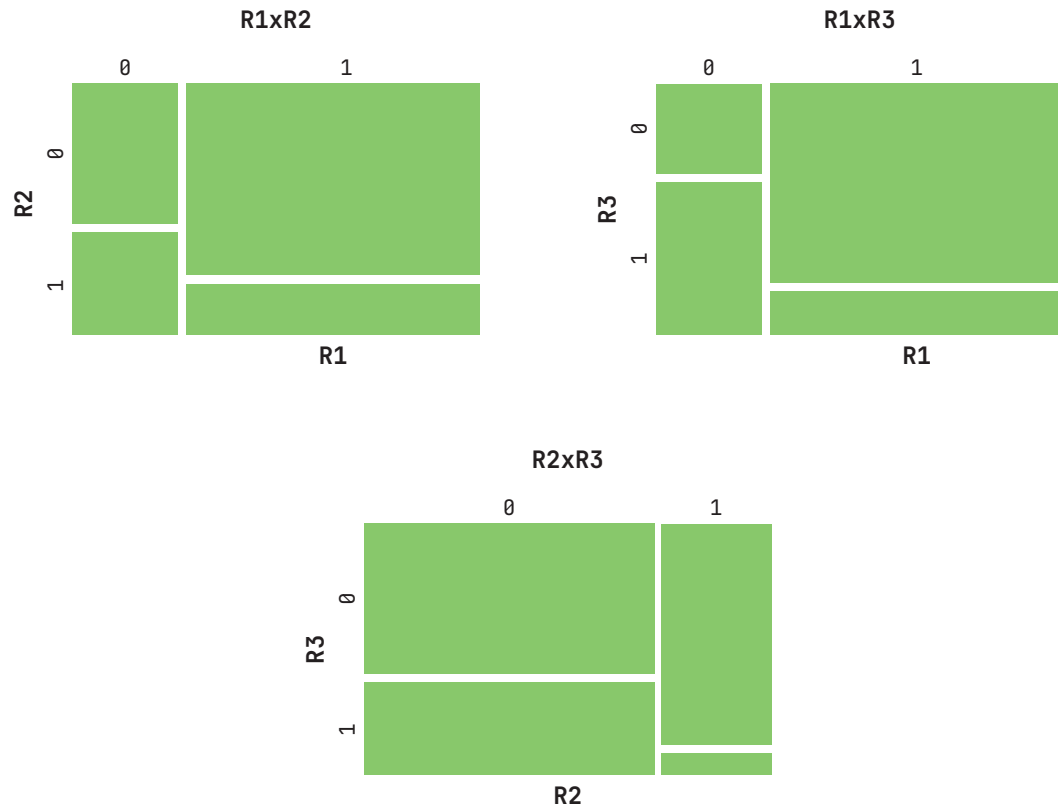
Il y a une augmentation des observations au fil des ans. Cette augmentation semble être en grande partie due à un nombre croissant d'observations par l'An Garda (R1).

Ensuite, nous faisons des mosaïques de toutes les paires de listes pour vérifier s'il y a suffisamment de chevauchement. Les surfaces des carrés dans les graphiques représentent les fréquences dans lesquelles on est observé dans l'une des deux listes, les deux listes, ou on n'est pas observé dans l'une ou l'autre des listes. On aimerait que les surfaces de toutes ces combinaisons soient présentes. L'absence d'une surface signifie qu'il manque une combinaison, ce qui, en fonction de la combinaison manquante, peut être alarmant :

- si la combinaison 00 est manquante : les deux listes couvrent toutes les observations dans les données, de sorte que les listes restantes n'apportent aucune nouvelle observation. Cela peut être une indication que l'une de ces deux listes est une combinaison des listes restantes. Cela doit être vérifié auprès de l'entité fournissant la liste.
- Si 01 ou 10 manque : c'est une indication qu'une des listes est une combinaison d'autres listes. Cela doit être vérifié auprès de l'entité fournissant la liste.
- Si 11 manque : il n'y a pas lieu de s'alarmer. Cela indique très probablement que les deux listes sont négativement associées ; l'inclusion dans une liste réduit la probabilité d'inclusion dans l'autre liste. L'effet d'une fréquence nulle de 00 est que les paramètres du modèle seront gonflés, mais pas les estimations de la taille de la population.

Dans chacun des trois tracés de la figure 5 ci-dessous, nous voyons 4 rectangles, dont la surface reflète la proportion d'observations dans les cellules 00, 10, 01 et 11 respectivement. S'il y avait eu une cellule sans observations, l'un des rectangles serait manquant, c'est-à-dire aurait une surface nulle.

Figure 5 Représentation graphique de la distribution des cas sur trois listes

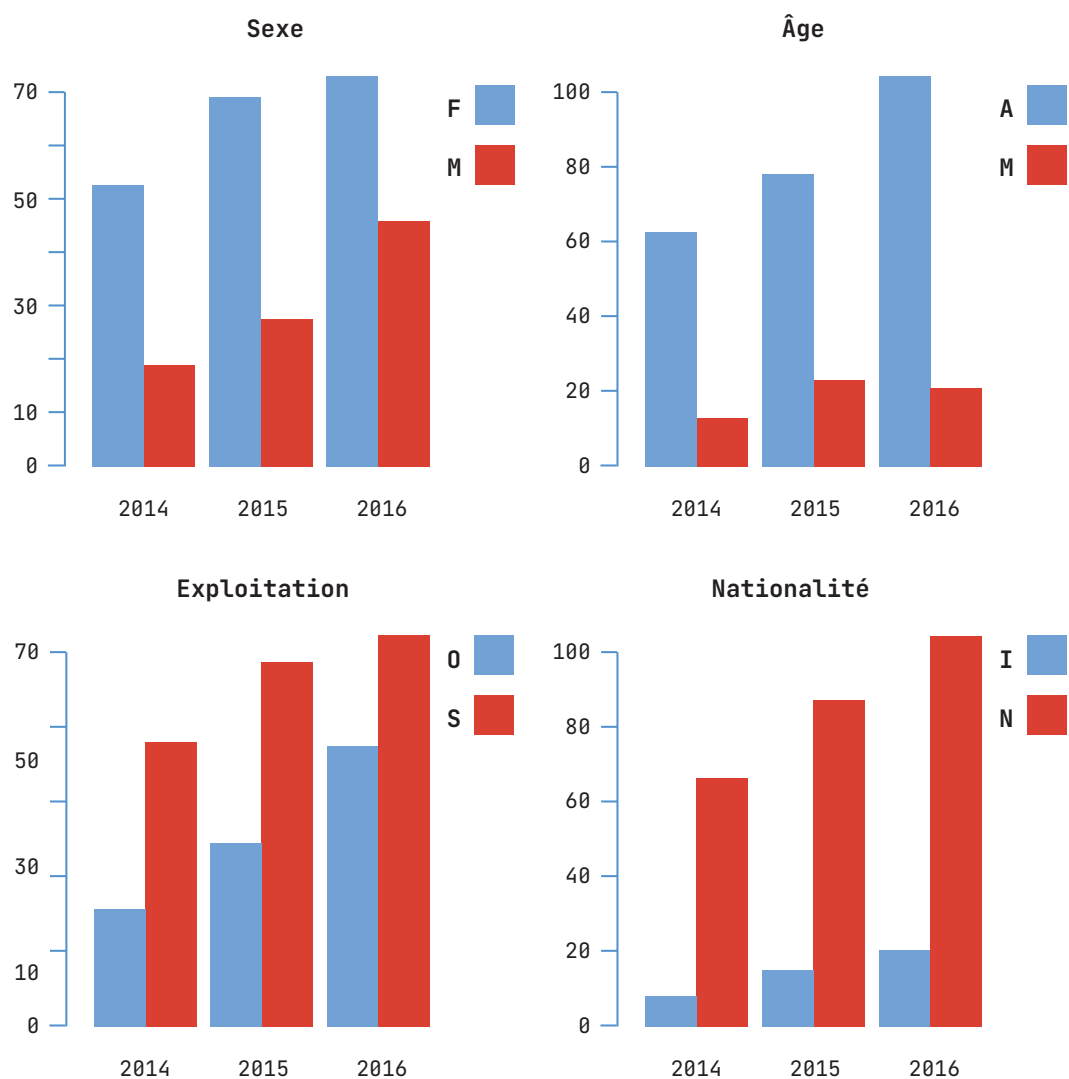


On voit que la surface dans le troisième tracé de $R2 = 1, R3 = 1$ est la plus petite. Cette cellule a tout de même 7 observations. Ainsi, aucune des combinaisons de listes n'a de fréquences proches de zéro, et il y a un chevauchement suffisant entre les paires de listes pour une estimation fiable des paramètres.³⁷

Nous vérifions également la distribution des covariables au fil des ans (voir figure 6).

37. Code R pour faire le 1er tracé : `mosaicplot(xtabs(~ R1 + R2, data = Ireland))`

Figure 6 Répartition des quatre covariables dans les diagrammes à barres par année



L'augmentation des observations au fil des ans ne semble dépendre d'aucune des covariables en particulier ; les niveaux des quatre covariables semblent montrer une augmentation relativement constante au fil des ans, à l'exception peut-être de la légère diminution de la fréquence des mineurs observés de 2015 à 2016. Cela signifie qu'il y a peu ou pas d'interaction entre l'année et chacune des covariables : sinon, l'augmentation des fréquences au fil des ans serait différente pour les niveaux à l'intérieur des covariables.

La 'mse' de paquet R

La « mse » de paquet R a été spécialement développée pour ce manuel par l'auteur Maarten Cruyff afin d'être utilisée pour plusieurs systèmes d'estimation des victimes de la traite des êtres humains. Le paquet sera disponible sur CRAN, mais pourra également être installé depuis GitHub avec la commande R :

```
devtools::install_github("MaartenCruyff/mse")
```

Le paquet contient deux fonctions, une pour la sélection du modèle et une pour la détermination des intervalles de confiance. La fonction pour effectuer une recherche du modèle le mieux adapté est :

```
model_search(x, lists = NULL, max_model = Freq ~ .^2,  
            year = NULL, degree_year = NULL)
```

où x désigne les données (un bloc de données avec un enregistrement par observation individuelle, ou un tableau de contingence), répertorie les numéros de colonne avec les listes, `max_model`, le modèle maximal pouvant entrer dans le chemin de recherche, `year`, le nom de la variable avec l'année de l'observation (si les observations sont collectées sur plusieurs années), et `degree_year`, le degré du polynôme pour l'année (si elle n'est pas spécifiée, des variables factices seront créées).

Le modèle minimal dans la recherche de modèle est spécifié par l'argument `min_model`, qui est par défaut le modèle `Freq ~` à effets principaux. L'argument `start_model` offre la possibilité de démarrer la recherche avec un modèle défini par l'utilisateur et, s'il n'est pas spécifié, la recherche de modèle démarre avec le modèle spécifié dans `min_model`.

La fonction qui effectue un bootstrap paramétrique sur le modèle le mieux ajusté est

```
boot_mse(object, modelnr = NULL, rep = 1000, seed = 1)
```

où `object` est l'objet créé avec `model_search()`, `modelnr` est le numéro du modèle dans le chemin de recherche (la valeur par défaut `NULL` sélectionne le modèle avec le BIC le plus bas), `rep` est le nombre d'échantillons bootstrap, et `seed` désigne la reproductibilité des résultats, et renvoie des tableaux et des graphiques des estimations de la taille de la population, stratifiés au long des années (si plus d'une) et sur les covariables et paires de covariables (si présentes).

La section suivante explique à nouveau les principes de l'estimation de systèmes multiples, maintenant dans le contexte des données irlandaises, et illustre le fonctionnement du paquet mse sur ces données. La section commence par le cas le plus simple de deux listes sans covariables, et part de là jusqu'à l'analyse simultanée de toutes les variables disponibles dans les données irlandaises.

Deux listes, pas de covariables, pas d'années

Nous appliquons maintenant les principes théoriques discutés ci-dessus. Pour les listes R1 et R2, le tableau de contingence pour l'année 2016 ressemble à celui de la figure 7.

Tableau 4 *Tableau de contingence de deux listes pour l'année 2016 dans l'ensemble de données irlandais*

	R1	R2	Freq
1	0	0	0
2	1	0	64
3	0	1	11
4	1	1	31

La première ligne de ce tableau de contingence fait référence aux personnes qui n'ont été observées sur aucune des deux listes, et a par conséquent des fréquences nulles.

Nous commençons par discuter des hypothèses et des violations des hypothèses pour ce tableau de contingence. L'hypothèse d'indépendance peut nécessiter quelques éclaircissements. Cela signifie que la probabilité d'inclusion dans R1 est indépendante de la probabilité d'inclusion dans R2. Parfois, cela est mal compris, lorsque les gens supposent que lorsque les listes sont construites indépendamment les unes des autres, cette hypothèse est remplie. Cependant, une violation clé se produit lorsque certaines victimes ont une probabilité beaucoup plus élevée d'être incluses dans les deux listes que d'autres victimes, et cela peut se produire même si les listes sont construites indépendamment les unes des autres, par exemple par des références croisées entre les entités qui maintiennent les listes.

Il a été indiqué précédemment que l'indépendance statistique des probabilités d'inclusion signifie que le rapport des cotes est égal à 1, et donc

$$\eta_{00} = \frac{n_{10}n_{01}}{n_{11}} \frac{64 \cdot 11}{31} = 22.7$$

Cela donne une estimation de la taille de la population totale $64 + 11 + 31 + 22.7 = 128.7$.

À l'aide du paquet `mse`, la recherche de modèle produit le modèle suivant, les coefficients log-linéaires et l'intervalle de confiance à 95% :

```
R1R2 <- model_search(x = d[d$Y == 2016, 1:2], lists = 1:2)
```

Step	Df	Deviance	Resid..Df	Resid..Dev	AIC	AICc	BIC	Nhat
1	NA	NA	0	0	0	0	0	128.7

```
R1R2$coefs
```

```
[[1]]
(Intercept)          R11          R21
 3.1227912  1.0360919 -0.7248959
```

```
boot_mse(R1R2)
```

```
$x
  nobs Nhat min95 max95
all  254  379   326   447
```

Notez que la spécification des données `d[d$Y = 2016, 1:2]` sélectionne les observations faites en 2016, et les variables R1 et R2, et qu'avec ces deux variables, le seul modèle qui peut être ajusté est $\text{Freq} \sim R1 + R2$. L'estimation de la taille de la population est de 379 avec un intervalle de confiance à 95% allant de 326 à 447.

Deux listes, une covariable, pas d'année

Nous incluons maintenant la covariable `sexe (S)`. Nous renvoyons aux sections précédentes pour une explication de l'importance d'inclure des covariables, lorsqu'elles sont disponibles. Le tableau de contingence de ces données est donné dans le tableau 8.

Tableau 5 *Tableau de contingence pour deux listes et la covariable sexe (F/M) pour l'année 2016 uniquement*

R1	R2	S	Freq
1	0	F	0
2	1	F	114
3	0	F	27
4	1	F	20
5	0	M	0
6	1	M	57
7	0	M	8
8	1	M	28

La recherche de modèles pas à pas a produit trois modèles : un modèle sans interaction, une interaction entre le sexe et la liste 1 et avec une interaction entre le sexe et la liste 2 respectivement. Le programme affiche également les Nhat (nombres estimés) des modèles et les valeurs des meilleures mesures d'ajustement (AIC, AICc et BIC). En bas, les résultats du bootstrap (estimations minimales et maximales) sont présentés.

```
R1R2S <- model_search(x = d[d$Y == 2016, c(1, 2, 4)], lists = 1:2)
```

Step	Df	Deviance	Resid..Df	Resid..Dev	AIC	AICc	BIC	Nhat
1	NA	NA	2	21.0	17.0	16.7	11.3	128.7
2 + R2:S	-1	8.8	1	12.3	10.3	10.1	7.4	128.7
3 + R1:S	-1	12.3	0	0.0	0.0	0.0	0.0	163.0

```
R1R2S$coefs[[3]]
```

(Intercept)	R11	R21	SM	R21:SM	R11:SM
4.0125145	-0.2513144	-1.8152900	-3.4528988	1.9488214	2.7362211

```
boot_mse(R1R2S)
```

```
$x
  nobs Nhat min95 max95
all  125  163   123   275

$S
  nobs Nhat min95 max95
F    77  114   74   225
M    48   49   38   62
```


Les mesures d'ajustement AIC, AICc et BIC sont toutes les plus faibles pour le 3^e modèle, c'est donc le modèle préféré (le modèle inclut une interaction entre le sexe et la deuxième liste). L'estimation de la taille globale de la population du modèle préféré est de 163, avec un intervalle de confiance à 95 % allant de 123 à 275 ; pour les femmes séparément, l'estimation est de 114 (74, 225) et pour les hommes de 49 (38, 62).

Deux listes, une covariable et l'année

Nous étendons maintenant le modèle en incluant la variable Année, indiquée par Y, en tant que covariable. Le tableau de contingence (avec les cellules R1 = 0, R2 = 0 supprimées) est présenté dans le tableau 6.

Tableau 6 Tableau de contingence de deux listes, la covariable sexe (Femmes et Hommes) et année (2014,

	R1	R2	S	Y	Freq
2	1	0	F	2014	30
3	0	1	F	2014	6
4	1	1	F	2014	1
6	1	0	M	2014	13
7	0	1	M	2014	5
8	1	1	M	2014	2
10	1	0	F	2015	41
11	0	1	F	2015	12
12	1	1	F	2015	12
14	1	0	M	2015	23
15	0	1	M	2015	1
16	1	1	M	2015	2
18	1	0	F	2016	43
19	0	1	F	2016	9
20	1	1	F	2016	7
22	1	0	M	2016	21
23	0	1	M	2016	2
24	1	1	M	2016	24

2015 et 2016)

La variable année a 3 niveaux/catégories, qui sont représentés par deux variables fictives avec l'argument année = «Y», ce qui permet à l'effet de l'année sur le log des fréquences ajustées d'être non linéaire. Les résultats de la recherche de modèle sont désormais les suivants :

```
R1R2SY2 <- model_search(x = d[, c(1, 2, 4, 8)], lists = 1:2, year = "Y")
```

Step	Df	Deviance	Resid..Df	Resid..Dev	AIC	AICc	BIC	Nhat
1	NA	NA	12	52.5	17.8	16.6	1.7	378.7
2 + R1:S	-1	3.5	11	49.0	16.2	15.2	3.9	378.7
3 + R2:S	-1	9.7	10	39.4	8.6	7.7	0.0	424.2
4 + S:Y	-2	5.3	8	34.0	7.2	6.6	6.0	424.2
5 + R2:Y	-2	3.8	6	30.2	7.4	7.1	13.7	424.2
6 + R1:Y	-2	11.5	4	18.7	0.0	0.0	13.6	573.6

```
round(R1R2SY2$coefs[[3]], 3)
```

(Intercept)	R11	R21	SM	Y2015	Y2016
3.542	-0.300	-1.740	-2.246	0.468	0.620
R11:SM	R21:SM				
1.553	1.030				

L'argument supplémentaire `degree_year = 1` traite l'année comme une variable numérique, avec un seul paramètre ayant un effet linéaire sur le logarithme des fréquences ajustées. Les résultats de la recherche de modèle sont légèrement différents, comme indiqué ci-dessous.

```
R1R2SY1 <- model_search(x = d[, c(1, 2, 4, 8)], lists = 1:2, year = "Y",
  degree_year = 1)
```

Step	Df	Deviance	Resid..Df	Resid..Dev	AIC	AICc	BIC	Nhat
1	NA	NA	13	53.9	21.1	20.7	6.3	378.7
2 + R2:Y	-1	4.1	12	49.8	19.1	18.7	7.9	378.7
3 + R1:Y	-1	11.9	11	38.0	9.2	9.0	1.8	465.5
4 + R1:S	-1	3.5	10	34.5	7.7	7.6	4.0	465.5
5 + R2:S	-1	9.7	9	24.8	0.0	0.0	0.0	542.7

```
round(R1R2SY1$coefs[[3]], 2)
```

(Intercept)	R11	R21	SM	Y	R21:Y
3.58	0.00	-1.58	-0.55	-3.34	3.34
R11:Y					
4.08					

Le BIC privilégie le 3^{ème} modèle avec deux paramètres pour l'année, et le 6^{ème} modèle avec un seul paramètre pour l'année. Ces deux modèles ne peuvent pas être comparés directement, car dans chaque tableau, les valeurs BIC sont mises à l'échelle de sorte que la valeur minimale soit 0. Les valeurs non mises à l'échelle sont obtenues avec

```
R1R2SY2$minima["BIC"]
```

```
BIC  
156.3
```

```
R1R2SY1$minima["BIC"]
```

```
BIC  
147.4
```

Ceci montre que le modèle à un seul paramètre pour l'année est plus parcimonieux que le modèle à deux paramètres pour l'année, et est donc à privilégier.

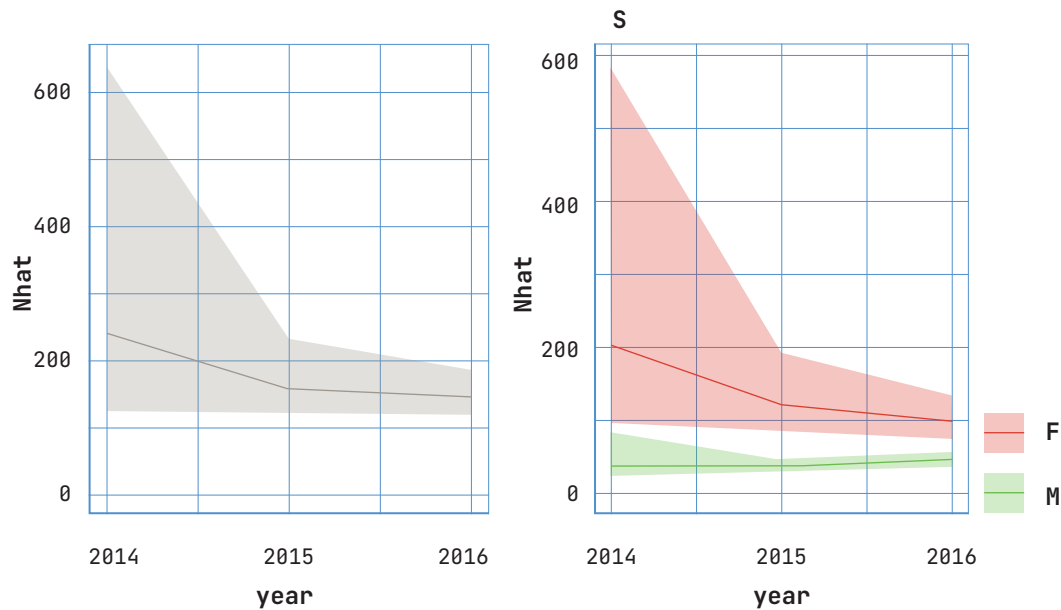
Lorsque l'argument `year` est spécifié, la fonction `boot_mse()` renvoie des tableaux ainsi que des tracés avec les estimations et les intervalles de confiance à 95%. Les tableaux obtenus sont les suivants :

```
out <- boot_mse(R1R2SY1)  
out$tables
```

```
$x  
  nobs Nhat min95 max95  
all  302  543   393  1012  
  
$Y  
  Year nobs Nhat min95 max95  
2014 2014   75  241   124   641  
2015 2015  102  157   120   235  
2016 2016  125  145   118   185  
  
$S  
  Year nobs Nhat min95 max95  
F 2014   55  203   97   587  
F 2015   73  121   86   195  
F 2016   77   98   74   134  
M 2014   20   38   23   85  
M 2015   29   36   28   47  
M 2016   48   47   37   58
```

Les tracés sont stockés dans l'élément de liste `out$plots`. La figure 7 présente les graphiques montrant les estimations minimales et maximales, globalement, et pour les hommes et les femmes séparément.

Figure 7 Estimation du nombre de victimes selon le modèle le mieux ajusté comprenant 2 listes, la covariable, le sexe et trois ans.



Les tableaux et les graphiques montrent de larges intervalles de confiance pour les estimations de 2014. L'inclusion d'une autre liste et/ou de plusieurs covariables peut aider à réduire la largeur de ces intervalles.

Toutes les listes, toutes les covariables et toutes les années

Comme dernier exemple, nous effectuons une recherche de modèle sur toutes les listes et toutes les covariables. Comme dans l'exemple précédent, nous commençons par effectuer une recherche de modèle avec un effet linéaire et non linéaire pour l'année, et une comparaison des BIC des deux modèles. Dans ce cas, le modèle à deux paramètres pour l'année fonctionne légèrement mieux (BIC de 473,2 contre 475,2). Les résultats de la recherche de modèle pour ce modèle sont :

```
ms2 <- model_search(d, lists = 1:3, year = "Y", degree_year = 2)
```

	Step	Df	Deviance	Resid..Df	Resid..Dev	AIC	AICc	BIC	Nhat
1		NA	NA	326	679.0	505.4	501.0	453.2	464.9
2	+ A:N	-1	165.9	325	513.1	341.5	337.3	293.0	464.9
3	+ S:E	-1	97.9	324	415.2	245.6	241.6	200.8	464.9
4	+ R3:S	-1	50.4	323	364.8	197.2	193.4	156.1	464.9
5	+ N:E	-1	42.0	322	322.9	157.3	153.6	119.9	464.9
6	+ R3:A	-1	41.6	321	281.2	117.6	114.2	84.0	464.9
7	+ R2:N	-1	35.3	320	245.9	84.3	81.1	54.4	474.9
8	+ S:N	-1	24.1	319	221.7	62.1	59.2	35.9	475.1
9	+ R2:E	-1	18.6	318	203.2	45.6	42.9	23.0	464.5
10	+ R3:Y	-2	18.9	316	184.2	30.6	28.5	15.5	464.5
11	+ R1:S	-1	12.1	315	172.2	20.6	18.8	9.2	459.1
12	+ R2:R3	-1	10.6	314	161.6	12.0	10.5	4.3	432.3
13	+ R2:A	-1	7.9	313	153.7	6.1	4.9	2.1	432.4
14	- R2:N	1	4.0	314	157.7	8.1	6.6	0.4	432.9
15	+ R3:E	-1	6.1	313	151.6	4.0	2.8	0.0	429.6
16	+ R2:N	-1	4.1	312	147.4	1.8	1.0	1.6	429.0
17	+ S:Y	-2	5.8	310	141.6	0.0	0.0	7.2	429.0
18	+ R2:Y	-2	3.6	308	138.0	0.4	1.2	15.0	429.7
19	+ R1:Y	-2	3.6	306	134.3	0.7	2.5	22.8	438.0
20	+ R3:N	-1	1.8	305	132.5	0.9	3.1	26.6	437.9
21	+ R1:N	-1	1.2	304	131.3	1.7	4.4	31.1	285687136.5
22	+ N:Y	-2	2.1	302	129.1	3.5	7.4	40.4	282461493.4
23	+ A:Y	-2	6.2	300	122.9	1.3	6.3	45.6	279138606.6

Le BIC favorise le modèle 15 et les AIC et AICc le modèle 17, mais les deux modèles aboutissent à la même estimation de la taille de la population (430 et 429 respectivement). Les trois derniers modèles montrent ce qui peut arriver si les données sont surajustées ; *les estimations de la taille de la population explosent à des chiffres irréalistes.*

La recherche du modèle comprend une interaction entre les listes, à savoir entre R2 et R3 (voir modèle 12). L'interprétation de cette interaction peut être déduite des estimations des paramètres. Les estimations paramétriques du modèle 15 (le meilleur modèle selon le BIC, présenté en annexe A (dernière page du document) montrent une estimation paramétrique négative (-1,3) pour l'interaction entre R2 et R3. Cela signifie que les personnes observées dans R2 sont moins susceptibles d'être observées dans R3, et vice versa, que ce à quoi on pourrait s'attendre en cas d'indépendance (conditionnelle) des listes.

Les pages suivantes montrent les tableaux et les graphiques créés en appliquant la fonction `boot_mse()` au modèle `ms2` (modèle 15), globalement et séparément pour les niveaux des covariables (Sexe : femmes/hommes ; âge (adulte/mineur ; nationalité (irlandais / non irlandais) et exploitation (sexuelle et non sexuelle).

```

$x
  nobs Nhat min95 max95
all  302  430   386   485

$Y
  Year nobs Nhat min95 max95
2014 2014   75   98    76   121
2015 2015  102  153   123   186
2016 2016  125  179   151   212

$S
  Year nobs Nhat min95 max95
F 2014   55   75   58   94
F 2015   73  106   82  134
F 2016   77  128  104  156
M 2014   20   23   15   32
M 2015   29   47   35   61
M 2016   48   51   38   66

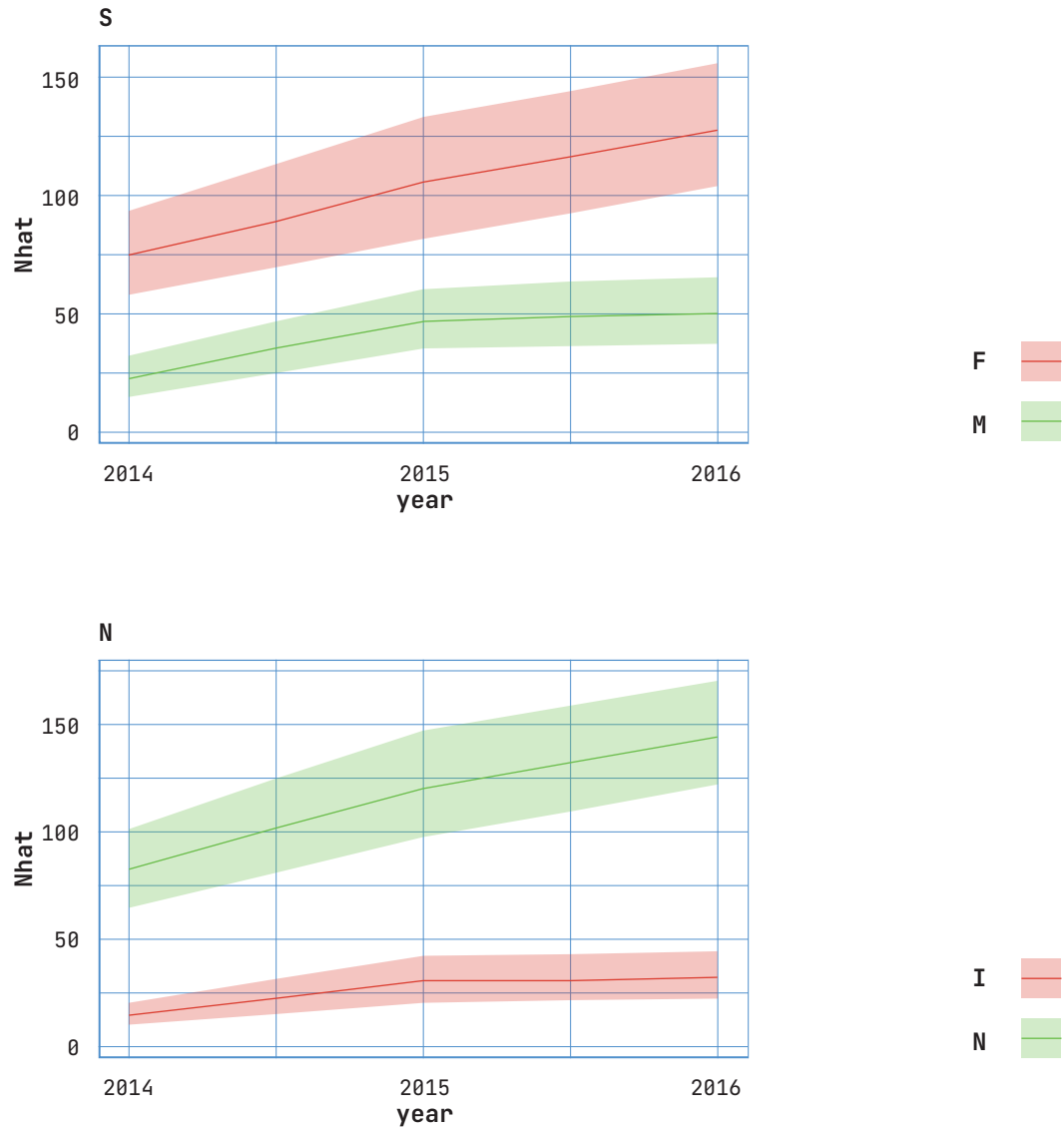
$A
  Year nobs Nhat min95 max95
A 2014   62   78   61   97
A 2015   78  110   88  135
A 2016  104  133  113  160
M 2014   13   20   13   28
M 2015   24   42   30   57
M 2016   21   45   33   60

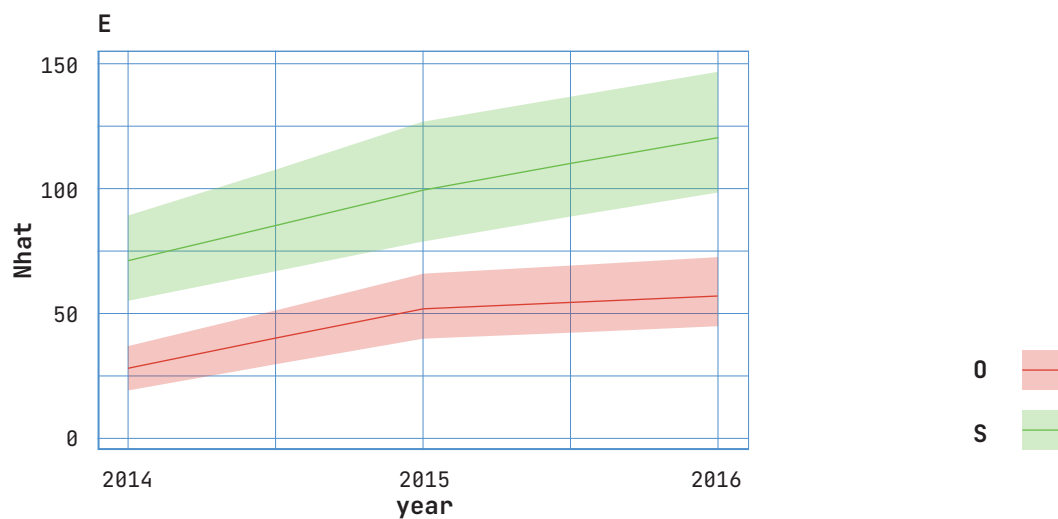
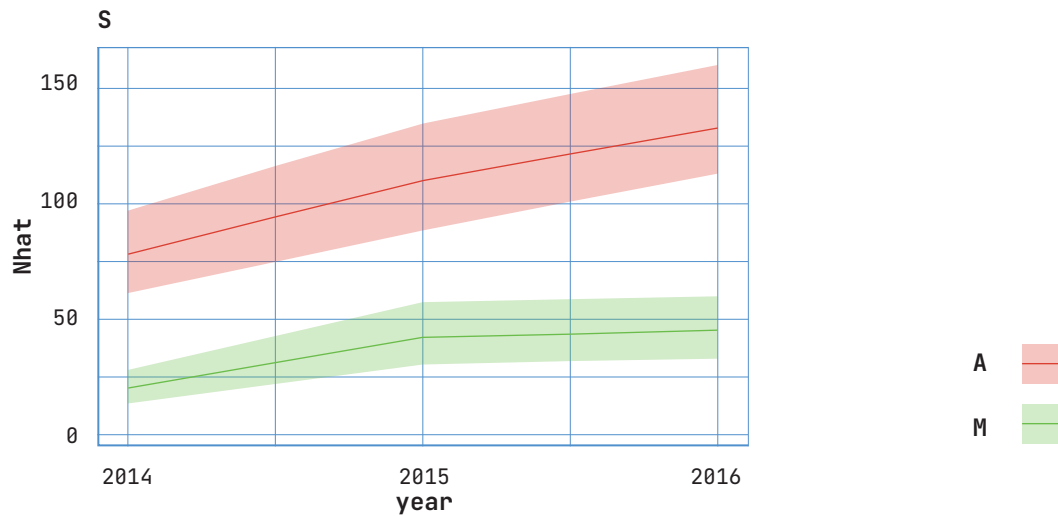
$N
  Year nobs Nhat min95 max95
I 2014    8   15   10   21
I 2015   15   31   21   43
I 2016   20   33   23   45
N 2014   67   83   65  102
N 2015   87  121   98  148
N 2016  105  145  123  172

$E
  Year nobs Nhat min95 max95
O 2014   22   28   19   37
O 2015   34   52   40   66
O 2016   52   57   45   73
S 2014   53   71   55   89
S 2015   68  100   79  127
S 2016   73  121   99  148

```

Figure 8 Estimation du nombre de victimes selon le modèle le mieux ajusté comprenant 3 listes, quatre covariables (S, A, N et E) et trois ans.





Les résultats présentés dans les tableaux et graphiques ci-dessus montrent que l'intervalle de confiance des estimations pour 2014 a été considérablement réduit par l'inclusion d'une troisième liste et de plusieurs covariables (comparer la figure 8 avec la figure 7 ci-dessus).

Les résultats pour les combinaisons de covariables sont présentés à l'annexe A.

SECTION 4

Aperçu des résultats des études MSE et discussion

Pouvoir mesurer un phénomène de manière fiable est le premier pas vers une réponse concertée et durable. Le rapport de l'Institut australien de criminologie sur son étude basée sur la MSE tire la conclusion suivante : «La quantification de l'étendue de la traite des êtres humains et de la victimisation par l'esclavage améliore notre compréhension du problème et fournit une base de preuves solides pour éclairer l'élaboration des politiques, l'aide aux victimes et les activités d'application de la loi » (Lyneham, Dowling & Bricknell, 2019, p.1). La valeur de la méthodologie MSE réside dans le fait qu'elle fournit aux pays une voie solide et rentable pour estimer l'ampleur et la nature réelles de leur problème de traite.

Une étude nationale utilisant la méthodologie MSE produira une estimation du nombre total de victimes de la traite des êtres humains dans le pays. Cette estimation des victimes est une statistique globale qui, à l'intérieur de modestes marges statistiques d'incertitude, indique le niveau véritable de victimisation de la traite dans le pays tel qu'actuellement définie par les autorités. Il montre également quelle part de cette estimation totale a été portée à l'attention des forces de l'ordre et/ou des ONG spécialisées et quelle part ne l'a pas été.

Si une étude nationale est réalisée avec des données sur le sexe, l'âge et la forme d'exploitation des victimes, les résultats peuvent également être utilisés pour mettre en évidence les types les plus courants de traite des êtres humains - par exemple, l'exploitation sexuelle ou l'exploitation non sexuelle - et quels groupes de la population sont les plus à risque. Cela peut fournir des informations inestimables pour la planification des programmes de prévention et l'établissement des priorités en matière d'enquête ou d'aide aux victimes. Plus précisément, l'utilisation de covariables par la MSE donne un aperçu des groupes de victimes qui sont plus ou moins susceptibles d'être détectés par les forces de l'ordre ou une organisation de soutien. Si les données de plusieurs années sont utilisées pour produire une étude MSE, il est possible de produire des analyses des changements dans la détection des victimes au fil du temps. Les résultats peuvent démontrer à la fois les tendances de la prévalence

dans la traite des êtres humains et si certaines victimes sont devenues plus ou moins « visibles » pour les autorités ou les ONG au fil du temps.

Quelques résultats clés de la MSE dans le monde

Les principaux résultats des études MSE sont des estimations du nombre total de personnes qui ont pendant un certain temps été victimes de traite/exploitées dans un pays au cours d'une année.³⁸ Aux fins de comparaison internationale, ces chiffres doivent être exprimés en taux pour 100 000 habitants, comme requis pour l'indicateur 16.2.2 des ODD. Une autre mesure comparative intéressante est le rapport entre le nombre enregistré et le nombre estimé par pays, car il donne un aperçu de l'efficacité des programmes existants pour détecter les victimes.

Bien que les résultats des études MSE dans différents pays ne soient pas strictement comparables, comme nous le verrons en détail ci-dessous, nous présenterons quelques résultats internationaux à titre d'illustration. La figure 9 présente les principaux résultats de dix études MSE menées ces dernières années pour un examen plus approfondi.

Figure 9 Nombre enregistré et estimé de victimes de la traite, intervalles de confiance des estimations, ratios entre les nombres enregistrés et estimés et nombres estimés pour 100 000 habitants, avec notes ajoutées, tirés de dix études MSE récentes (Durgana & Van Dijk, 2021)

Pays/Lieu	An	Enregis- tré	Estimation	Min max	Proportion de Rec/Est	Victimes	
						pour 100.000	Remarques
Australie	2016/ 2017	414	1567	1300-1900	1:4	3	adultes seulement
Irlande	2016	125	197	148-213	1:2	4	
Pays-Bas	2014/ 2015	1450	6500	6250-6750	1:4	38	
Roumanie	2015	880	1328	1062-2083	1:2	7	y compris à l'étranger

38. Ce taux fournit une estimation dite de flux, à distinguer du taux de personnes exploitées dans un pays à un moment donné (une estimation dite de stock). Les estimations de stock peuvent être calculées en multipliant l'estimation du débit par la durée moyenne de l'exploitation. Par exemple, si la durée moyenne d'exploitation est de trois mois, une estimation de flux de 1000 victimes par an donne une estimation de stock de 250 victimes à un moment donné.

Pays/Lieu	An	Enregis- tré	Estimation	Min max	Proportion de Rec/Est	Victimes	Remarques
						pour 100.000	
Slovaquie	2016	45	87	59-164	1:2	1.5	y compris à l'étranger
Serbie	2016	146	970	670-1560	1:7	14	y compris à l'étranger
Royaume- Uni	2014	2744	11000	10000- 13000	1:4	17	
États-Unis 1 (Est)	2016	290	657	549-806	1:2	109	mineurs seulement
États-Unis 2 (Ouest)	2016	345	2235	1 606-3 609	1:6	972	
États- Unis 3 (Nouvelle- Orléans)	2016	185	1000	650-1600	1:5	69	

D'un point de vue comparatif, les taux de chiffres estimés pour 100 000 habitants semblent la statistique la plus pertinente. La première étude MSE réalisée sur les données de la traite des êtres humains, réalisée au Royaume-Uni, a produit un taux de victimisation de 17 pour 100 000 habitants en 2014 (Bales, Hesketh & Silverman, 2015). L'étude de l'ONU/DC/Rapporteur national néerlandais sur les données néerlandaises, utilisant des covariables, a trouvé un taux de 38 pour 100 000 habitants aux Pays-Bas pour 2014/2015. L'étude MSE australienne, n'utilisant pas de covariables, a rapporté un taux estimé de 3,3, principalement composé d'adultes (Lyneham, Dowling & Bricknell, 2019). D'autres études MSE – toutes utilisant des covariables – ont produit des taux de 12 pour la Serbie, 6 pour la Roumanie et 3 pour l'Irlande et la Slovaquie. Dans les trois études locales aux États-Unis, des taux de prévalence beaucoup plus élevés ont été estimés que dans les études nationales en Europe et en Australie.

Des explications possibles de ces différences peuvent d'abord être trouvées dans les méthodologies des études. Certaines études incluaient des covariables et d'autres non. Bien que l'utilisation de covariables puisse conduire à des estimations à la fois supérieures et inférieures, elles affectent souvent les estimations et compromettent ainsi la comparabilité. Les résultats divergents des études américaines pourraient également être liés à la méthodologie appliquée. Premièrement, il s'agissait des résultats d'études locales. La prévalence locale de la criminalité peut diverger fortement d'un taux national, comme cela a été constaté dans de nombreuses études criminologiques. Les études peuvent avoir été menées dans des endroits à forte concentration de traite des êtres humains. Un autre facteur expliquant peut-être la prévalence élevée dans les études américaines est que la MSE a été menée sur des listes de victimes présumées telles que définies par les chercheurs, plutôt que sur

des listes administratives produites par des policiers et d'autres praticiens appliquant les définitions officielles de la traite des êtres humains.

Comparabilité internationale

Lors de l'interprétation de la variation des taux de victimisation, plusieurs autres facteurs compromettant la comparabilité doivent être pris en compte. Un premier facteur est la définition de la traite des êtres humains dans le droit national. En termes de législation, le Protocole des Nations Unies sur la traite des personnes a inspiré de nombreuses lois nationales, car la grande majorité des pays disposent désormais d'une législation largement conforme aux dispositions énoncées dans le Protocole des Nations Unies sur la traite des personnes.³⁹ Néanmoins, il existe une ambiguïté juridique et une marge d'interprétation dans la définition de la traite des personnes du Protocole des Nations Unies sur la traite des personnes. Cette ambiguïté se reflète dans les différentes manières dont les pays ont choisi de criminaliser la traite, dans la manière dont la définition est opérationnalisée par les agences impliquées et dans la priorisation de l'application des lois. Dans certains pays, la détection de la traite des êtres humains est par exemple largement axée sur la traite des femmes à des fins d'exploitation sexuelle, tandis que dans d'autres, les cas de traite domestique ne sont pas inclus. Ces facteurs ont une forte influence sur les types de victimes qui apparaissent sur les listes utilisées pour la MSE et peuvent compromettre la comparabilité internationale des résultats⁴⁰.

Un deuxième facteur entravant la comparabilité internationale est le concept de victime présumée. Ni dans le Protocole des Nations Unies ni dans aucune autre loi internationale, ce concept n'a été défini de manière étroite. Le concept de victime présumée fait souvent partie d'un mécanisme national d'orientation. Au Royaume-Uni, par exemple, les victimes présumées sont strictement définies comme des personnes qui réussissent un « test de motifs raisonnables » indiquant qu'elles ont été victimes de la traite (un expert de l'autorité compétente « soupçonne mais ne peut pas prouver » qu'une personne est plus susceptible que non d'être victime de la traite). Aux Pays-Bas, la notion de victime identifiée n'existe pas formellement. Pour des raisons pratiques, les forces de l'ordre et les ONG ont pour instruction de

39. 169 des 181 pays évalués pour le *Rapport mondial sur la traite des personnes (Global Report on Trafficking in Persons)* de 2020.

40. En Europe, la législation et les politiques relatives à la traite des êtres humains ont été davantage harmonisées par la directive de l'UE de 2012 et la convention du Conseil de l'Europe sur la lutte contre la traite des êtres humains de 2009. Toutes deux exigent l'incorporation de la définition de Palerme dans le droit national et encouragent la collecte de statistiques de multiples sources sur les victimes ainsi définies.

considérer comme une « victime présumée » toute personne qui, aux yeux des premiers intervenants, « *montre quelque signe qui soit qu'elle pourrait avoir été victime de la traite* ». La définition néerlandaise semble plus large et place donc la barre un peu plus bas pour inclure les personnes en tant que victimes. Les définitions utilisées dans la plupart des autres pays ressemblent à celle utilisée au Royaume-Uni.

Troisièmement, l'univers des victimes présumées de la traite diffère fondamentalement entre les typiques pays d'origine de traite des êtres humains et les pays de destination. Sur les dix pays où la MSE a été appliquée aux données sur les victimes, la Serbie, la Roumanie et la Slovaquie sont principalement des pays d'origine de la traite. Les victimes présumées ont pour la plupart été victimisées à l'étranger et rapatriées dans leur pays d'origine. Le nombre de victimes enregistrées et le nombre total estimé de ces victimes fournissent des informations importantes aux décideurs politiques sur le nombre de victimes qui devraient bénéficier d'une protection et d'un soutien dans le pays. Cependant, ces personnes n'ont pas toutes été exploitées sur le territoire national du pays, et les taux de ces pays ne sont donc pas strictement comparables à ceux des pays de destination tels que le Royaume-Uni, les Pays-Bas, l'Australie et l'Irlande. Pour calculer les taux de personnes exploitées sur le territoire d'un pays, les taux slovaque, serbe et roumain doivent être décomposés en taux pour les personnes exploitées à l'étranger et pour les personnes exploitées sur le territoire national. Ces derniers taux sont considérablement inférieurs aux taux totaux présentés à la figure 9.

Nombres estimés par covariables

Comme indiqué, si des covariables sont utilisées dans les calculs, les estimations globales sont plus robustes et des taux distincts de victimisation de la traite peuvent être calculés pour différents groupes de population, comme le demande l'indicateur 16.2.2 des ODD.

Dans l'étude néerlandaise, quatre covariables ont été incluses : le sexe (masculin/féminin), l'âge (mineur/adulte), le type d'exploitation (sexuelle/non sexuelle) et la nationalité (néerlandaise/non néerlandaise). Les taux de victimisation pour 100 000 habitants pour exploitation sexuelle et non sexuelle étaient respectivement de 25 et 13, ce qui indique qu'aux Pays-Bas, la traite à des fins d'exploitation sexuelle était la forme prédominante de traite au cours de cette période. Cependant, une répétition de l'étude utilisant des données de 2016-2019 montre que les totaux estimés de cas d'exploitation sexuelle et de travail forcé ont convergé et sont maintenant presque égaux (Van Dijk, Cruyff & Van der Heijden, 2021). Un tel changement dans la composition du nombre total estimé de victimes de la traite a des implications évidentes pour les politiques de lutte contre la traite.

Les études MSE utilisant des covariables peuvent produire des taux de victimisation par sous-groupe de population. Il a été constaté que ces taux varient considérablement. Aux Pays-Bas, les taux par an pour 2015 étaient de 32 victimes pour 100 000 femmes néerlandaises adultes et de 1 pour 100 000 hommes néerlandais adultes. La raison de cette différence est que les hommes ne sont que rarement exploités sexuellement et sont également un peu moins susceptibles d'être soumis au travail forcé. En supposant que les enfants victimes de la traite des êtres humains aux Pays-Bas ont généralement entre 12 et 17 ans, les taux de victimisation des mineurs sont de 235 pour 100 000 filles (entre 12 et 17 ans) et de 14 pour 100 000 garçons (entre 12 et 17 ans). Le taux de ressortissants néerlandais est estimé à 22. Le taux parmi les résidents non néerlandais (réguliers ou irréguliers) est estimé à 326 victimes pour 100 000. Certains de ces migrants ont sans aucun doute été introduits dans le pays dans le but même d'être exploités par la traite.

Sur les dix études, six ont utilisé des covariables. Si à l'avenir les résultats d'un plus grand nombre de pays deviennent disponibles, cela permettra des comparaisons internationales plus ciblées des nombres estimés par type d'exploitation, sexe et âge. Avec de tels raffinements, la comparabilité des résultats sera considérablement accrue.

Proportion entre les cas enregistrés/observés et estimés

Dans la figure 9, sont présentés des résultats internationaux concernant les rapports entre les nombres enregistrés et estimés. Les ratios par pays diffèrent dans une fourchette limitée de 1:2 et 1:7. Le ratio typique dans les grands pays de destination – Australie, Pays-Bas, Royaume-Uni et États-Unis (Nouvelle-Orléans) est de 1:4 ou 1:5. Dans tous ces pays, les entités d'enregistrement telles que la police et les ONG détectent ensemble un quart de toutes les victimes présumées ou moins. Avec des efforts supplémentaires, ils pourraient facilement doubler ou même tripler leur nombre enregistré et ainsi le nombre de victimes secourues et soutenues. Les possibilités d'amélioration de la détection se sont révélées considérablement plus faibles en Serbie (3), en Roumanie (2) et en Irlande (1,5).

Bien que le nombre limité de pays participants ne permette pas de tirer des conclusions définitives, les résultats suggèrent que les ratios entre les victimes enregistrées et estimées sont plus faibles dans les pays d'origine typiques que dans les pays de destination, bien que le ratio relativement faible en Irlande soit une exception notable. Une explication possible de cette tendance pourrait être que dans les pays d'origine, la majorité des victimes présumées ou identifiées sont des ressortissants victimisés rapatriés, et souvent principalement détectés et identifiés à l'étranger. Les personnes précédemment identifiées par les institutions et/ou par elles-mêmes comme

victimes dans un pays de destination pourraient être plus visibles pour les institutions concernées dans leur pays d'origine lors de leur rapatriement que les victimes dans les pays de destination. Dans ce contexte, il convient également de reconnaître que pour l'identification des victimes de la traite, les institutions des pays d'origine dépendent largement des efforts de détection de leurs homologues dans les pays de destination. Dans les pays d'origine typiques, l'identification des victimes dépend corrélativement des efforts d'application de la loi dans les pays où leurs ressortissants sont exploités et/ou des références internationales des ONG prestataires de services de ces pays. La nature transfrontalière de la traite des êtres humains complique les implications en matière de politiques pour les pays producteurs de résultats basés sur la MSE concernant les victimes exploitées à l'étranger. Ces résultats mettent principalement en lumière l'efficacité des dispositions nationales pour gérer les renvois de victimes étrangères.

Observations finales et discussion

Des études MSE qui ont été réalisées jusqu'à présent, plusieurs enseignements peuvent être tirés.

L'expérience a appris que pour rendre les ensembles de données existants appropriés pour la MSE, ils doivent souvent être quelque peu adaptés, par exemple en combinant des listes plus petites d'institutions similaires de plus grandes listes. Une étroite consultation entre les gestionnaires opérationnels du système de données d'origine et les chercheurs est la clé d'une étude MSE réussie.

Pour les gouvernements, un avantage de la MSE est qu'elle prend comme point de départ les propres données administratives des gouvernements, enregistrées par des institutions officiellement dédiées travaillant au sein de l'infrastructure juridique et opérationnelle du pays. Pour cette raison, les estimations basées sur la MSE fournissent des informations pertinentes aux décideurs politiques sur le nombre maximum de victimes qui pourraient être enregistrées si toutes les institutions impliquées dans la détection et le soutien des victimes étaient bien établies et dotées de ressources optimales. Le fait de savoir que, par exemple, trois ou quatre fois plus de victimes pourraient potentiellement être atteintes si les efforts de détection étaient intensifiés, permet aux décideurs de fixer des objectifs réalistes et fondés sur des preuves pour les politiques de lutte contre la traite dans leur pays dans un délai choisi.

Un avantage évident des études MSE par rapport à d'autres méthodes pour mesurer des chiffres cachés tels que les enquêtes de population à grande échelle est leur rapport coût-efficacité. Dans de plus en plus de pays, des systèmes nationaux

complets d'information sont introduits pour améliorer la détection et le soutien des victimes de la traite. Un effet secondaire positif de cette évolution est que, dans un nombre croissant de pays, des ensembles de données numérisés et intégrés sur les victimes seront disponibles dans des environnements sécurisés de manière appropriée. Si ces ensembles de données répondent aux spécifications appropriées, des estimations basées sur la MSE des nombres réels peuvent être réalisées à des coûts minimes.

L'expérience néerlandaise a montré que des études MSE répétées peuvent fournir un aperçu factuel des tendances à long terme des différents types de traite des êtres humains, révélant dans ce cas un déplacement progressif du volume de victimes de la traite des victimes de la traite à des fins d'exploitation sexuelle vers les victimes de traite à des fins d'exploitation par le travail entre 2010 et 2019.

L'utilisation de données administratives pour les MPE s'accompagne également de certaines limites. Certaines catégories de victimes peuvent être si profondément cachées dans la société qu'aucune d'entre elles n'a jamais été enregistrée. Dans la terminologie de la biologie marine - où la méthodologie de capture-recapture a été appliquée pour la première fois en capturant, marquant et recapturant des poissons dans un étang - ces segments sont appelés « mangeurs terrestres », c'est-à-dire des poissons qui n'apparaissent jamais dans les prises normales de poissons. Ces segments qui n'apparaissent sur aucune des listes disponibles resteront hors de portée de l'estimation de systèmes multiples. Des exemples dans le domaine de la traite des êtres humains sont des sous-groupes particulièrement profondément cachés tels que les immigrants irréguliers se livrant au commerce du sexe ou travaillant dans des sites isolés à la campagne ou sur des navires. Un autre exemple est celui des nouveaux groupes de victimes qui n'ont pas encore attiré l'attention des forces de l'ordre ou des prestataires de services. Un exemple pertinent au Royaume-Uni est celui des mineurs qui travaillent comme passeurs de drogue pour ce que l'on appelle les « filières de comtés » (county lines) dans différentes parties du pays. À l'heure actuelle, ces mineurs constituent une part importante des victimes identifiées de la traite, mais ils ont été largement oubliés dans l'étude MSE qui a utilisé des données de 2013, car aucune ou presque aucune victime de ce type n'avait été enregistrée à l'époque. La possibilité que certains sous-groupes ou groupes de victimes profondément cachés qui n'ont pas encore été reconnus comme tels ne soient enregistrés sur aucune liste et soient donc omis dans les estimations, est une autre raison pour laquelle les études MSE devraient être répétées régulièrement, en utilisant les dernières listes disponibles. C'est également une raison de compléter l'estimation basée sur la MSE par des études de terrain sur la victimisation autodéclarée parmi les groupes vulnérables afin de détecter les nouvelles formes émergentes de traite et d'exploitation des êtres humains.

Bibliographie

- Agresti, A., *Categorical Data Analysis*, 3^e éd., Wiley & Sons, New Jersey, 2013.
- Anderson, RV, TC Kulig & Chr. J. Sullivan, *Estimating the Prevalence of Human Trafficking in Ohio*, School of Criminal Justice at the University of Cincinnati under grant number 2016-JG-HTP-6096 for the Ohio Office of Criminal Justice, février 2019a.
- Anderson, RV, TC Kulig & Chr. J. Sullivan, Agency Records as a Method for Examining Human Trafficking, In M. Delen & DMD Siva (Eds), *Methods of Criminology and Criminal Justice*, Emerald Group of Publishing, pages 165-178, 2019b.
- Bales, KO, O. Hesketh & BW Silverman, “Modern Slavery in the UK; how many victims”, *Significance*, pages 16-21, Vol 12 juin 2015 .
- Bales, KO, LT Murphy & BW Silverman, How many trafficked people are there in Greater New Orleans; Lessons in Measurement, *Journal of Human Trafficking*, Vol 6, Numéro 4, pages 375-385, 2020.
- Barrick, K. & Pfeffer, R. *Human Trafficking Policy and Research Analyses Project: Comprehensive Review of Prior Prevalence Studies and Recommendations for Field Testing in the United States*. Rapport OPRE n° 2021-XXX. Washington, DC : Bureau de la planification, de la recherche et de l'évaluation, Administration pour les enfants et les familles, Département américain de la santé et des services sociaux, 2021.
- Buimer, L., A. Mack, M. van Meeteren et Y. Bleeker, *Mensenhandel Herkennen en Begrijpen; Lessen van Utrecht en Ede*, Regioplan, (Reconnaître et comprendre la traite des êtres humains ; Leçons à partir d'Utrecht et Ede), décembre 2021.
- Bird, SM & R. King, “Multiple Systems Estimation (or Capture-Recapture Estimation) to Inform Public Policy”, *Annual Review of Statistics and Its Application*, vol. 5, p. 95-118, mars 2018.
- Böhning , D., PGM van der Heijden & J . Bunge, *Capture-Recapture Methods for the Social and Medical Sciences*, CRC Press, Boca Raton, 2018.
- Burnham, K. et Anderson, D., *Model selection and multimodel inference: a practical information-theoretic approach*. New York : Springer, 2002

- Chan, L., Silverman, B.W. & Vincent, K., Multiple Systems Estimation for Sparse Capture Data: Inferential Challenges When There Are Nonoverlapping Lists, *Journal of the American Statistical Association*, 116:535, 1297-1306, 2021.
- CoMensha, *Slachtoffermonitor Mensenhandel 2014-2018 uitgebracht: reactie CoMensha op de vijf aanbevelingen*, 18 octobre 2019 (disponible en ligne : <https://www.comensha.nl/actualiteiten/item/slachtoffermonitor-mensenhandel-2014-2018-vijf-aanbevelingen>).
- Conseil de l'Europe, *9e rapport général d'activités du GRETA*, 2019 (disponible en anglais sur : <https://rm.coe.int>).
- Croix-Rouge croate, *The identification of victims of human trafficking in transit and destination countries in Europe; A practical guideline for frontline workers*, projet STEP, 2019.
- Cruyff, M., van Dijk J. & van der Heijden P., The Challenge of Counting Victims of Human Trafficking: Not on the record: A multiple systems estimation of the numbers of human trafficking victims in the Netherlands in 2010–2015 by year, age, gender, and type of exploitation. *CHANCE* 30(3), p. 41–49, 2017.
- De Vries, I. & C. Dettmeijer -Vermeulen, Extremely wanted: human trafficking statistics- what to do with the hodgepodge of numbers? In Kangaspunta, K. (ed.) *Forum on Crime and Society, Special Issue: Researching Hidden Populations: Approaches to and Methodologies for Generating Data on Trafficking in Persons*, 8, pages 15-37, 2015.
- Durgana, D. & JJM van Dijk, Measuring the Hidden Numbers of Human Trafficking Through Multiple Systems Estimation: Lessons Learned and Challenges Outstanding, *J. of Crime and Delinquency*, première publication en janvier 2021.
- Farrell, A., M. Dank, M. Kafafian, S. Lockwood, R. Pfeffer, A. Hughes et K. Vincent, *Capturing Human Trafficking Victimization Through Crime Reporting*, National Institute of Justice, janvier 2019.
- Ministère français des Affaires étrangères, *Guidelines for the first level identification of victims of trafficking in Europe (Lignes directrices pour l'identification de premier niveau des victimes de la traite en Europe)*, Euro TrafGuID, juin 2013.
- Goudie IBJ & M. Goudie, Who captures the marks for the Petersen estimator, *JR Statist. Soc. A* 170, partie 3, p. 825–839, 2007.

- Bureau international du travail, *Operational indicators of trafficking in human beings, results from a Delphi Study*, Genève : OIT, 2009.
- Bureau international du travail, Special Action Programme to Combat Forced Labour, *ILO Global Estimate of Forced Labour: Results and Methodology*, 2012.
- Bureau international du travail, Département des statistiques, ICLS/20/2018/Guidelines, *Guidelines concerning the measurement of forced labour*, 2018.
- Groupe de travail international pour la surveillance et la prévision des maladies (IWGDMF), Capture-recapture and multiple-record systems estimation I: History and theoretical development. *American Journal of Epidemiology*, 15 novembre;142(10), pp1047-58, 1995 .
- Jones, HE, Hickman, M., Welton, NJ, De Angelis, D., Harris, RJ, & Ades, AE, Recapture or precapture? Fallibility of standard capture-recapture methods in the presence of referrals between sources. *American Journal of Epidemiology*, 179 (11), 1383–1393, 2014.
- Lyneham, S, C. Dowling & S. Bricknell, Estimating the dark figure of human trafficking and slavery victimisation in Australia, *Australian Institute of Criminology: Statistical Bulletin* 16, février 2019.
- Rapporteur national sur la traite des êtres humains et la violence sexuelle contre les enfants (Pays-Bas), *Victims of Human Trafficking. Periodical Report 2012-2016. Summary*, 2017.
- O'Connell, D, R. Brittingham, D. Borton & L. Mutis, *Human Trafficking in Delaware*, Centre d'études sur les drogues et la santé, Université de Delaware, septembre 2020.
- OSCE/BIDDH, *National Referral Mechanisms; Joining efforts to Protect the Rights of Trafficked Persons, A Practical Handbook*, 2004.
- Overstall, AM, King R., Bird SM, Hutchinson SJ, Hay G. Incomplete contingency tables with censored cells with application to estimating the number of people who inject drugs in Scotland. *Stat Med.* 30 avril; 33(9):1564-79, 2014.
- Phillips, AR *Commercial sexual exploitation; an analysis of prostitution in Kansas City*, Thèse, Université du Missouri-Kansas City, 2017.

- Quinteros Rojas., DR Dufrax Tapia, R. Ramos Rodríguez, Human Trafficking Cases in Chile: Challenges for Reducing the “Dark Figure” *The Palgrave International Handbook of Human Trafficking*, pages 1151-1164, 2019.
- Sharifi Far, S., R.King, S. Oiseau, A. Dépassement, H. Worthington & N. Jewell, Multiple Systems Estimation for Modern Slavery: Robustness of List Omission and Combination, *Crime & Delinquency*, vol. 67, 13-14 : pages 2213-2236, première publication le 22 août 2020.
- Silverman, BW *Modern slavery: an application of multiple systems estimation*. Gouv. Royaume-Uni. 2014.
- Silverman, BW, ‘Multiple Systems Estimation for Sparse Capture Data: Inferential Challenges when there are Non-Overlapping Lists’, *Statistics in Society, Série A*, Volume 183, Numéro 3, pages 689-1326, 2020.
- Silverman BW et K. Vincent, ‘Multiple Systems Estimation for Sparse Capture Data: Inferential Challenges when there are Non-Overlapping Lists’, *Journal de l’American Statistical Association*, vol. 00, non. 0, p. 1-10, février 2020.
- Sutherland , J. & CJ Schwarz , Multi-List Methods Using Incomplete Lists in Closed Populations, *Biometrics. J of the Biometric Society*, Vol 61, 1, pages 134-140, 2005.
- ONUDC/Rapporteur national sur la traite des êtres humains et la violence sexuelle contre les enfants (Pays-Bas), *Monitoring Target 16.2 of the United Nations' Sustainable Development Goals; a multiple systems estimation of the numbers of presumed human trafficking victims in the Netherlands in 2010-2015 by year, age, gender, form of exploitation and nationality*, 2017.
- ONUDC en collaboration avec Walk Free Foundation, *Research Briefs: Monitoring Target 16.2 of the United Nations Sustainable Development Goals: multiple systems estimation of the numbers of presumed victims of trafficking in persons, Ireland, Romania and Serbia*, 2018 (disponible sur : <https://www.unodc.org/unodc/en/data-and-analysis/tip.html>).
- ONUDC, *Research Brief: Multiple Systems Estimation for estimating the number of victims of human trafficking across the world*, 2015.
- ONUDC, *UNODC helps countries uncover real magnitude of trafficking in persons and report on SDG 16*, 2018 (voir www.unodc.org).

ONUUDC, *Global Report on Trafficking in Persons*, 2020 (Publication des Nations Unies, numéro de vente F.20.IV.3), p. 61, 2020

Van der Heijden, PGM, J. Whittaker, M. Cruyff, B. Bakker & R. Van der Vliet, 'People born in the Middle East but residing in the Netherlands: invariant population size estimates and the role of active and passive covariates,' *The Annals of Applied Statistics*, Vol. 6, n° 3, pages 831 à 852, 2012.

Van der Heijden, PGM, de Vries, I., Böhning, D., & Cruyff, M. Estimating the size of hard-to-reach populations using capture-recapture methodology, with a discussion of the International Labour Office's global estimate of forced labour, In Kangaspunta, K. (ed.) *Forum on Crime and Society, Special Issue: Researching Hidden Populations: Approaches to and Methodologies for Generating Data on Trafficking in Persons*, 8. pages 109-136, 2015

Van Dijk, JJM & PGM van der Heijden *Research Brief. Multiple Systems*

Estimation for estimating the number of victims of human trafficking across the world, Vienne : UNODC, 2016

Van Dijk, J. & PGM van der Heijden (2016). *On the potential of Multiple Systems Estimation for estimating the number of victims of human trafficking across the world*. Écrit pour l'ONUUDC, où il est paru comme un mémoire de recherche. (disponible sur : <https://www.unodc.org/documents/data-and-analysis/tip/Ti-PMSE.pdf>).

Van Dijk, JJM, Cruyff, M., Van der Heijden, PGM & Kragten-Heerdink, SLJ, *Monitoring Target 16.2 of the United Nations sustainable development goals: a multiple systems estimation of the numbers of presumed human trafficking victims in the Netherlands in 2010-2015 by year, age, gender, form of exploitation and nationality*, Vienne : UNODC, pages 51, 2018.

Van Dijk, JJM, Cruyff, M. & Van der Heijden, PGM ., *A Multiple Systems Estimation of presumed victims of human trafficking in the Netherlands during 2016-2019*, La Haye : ministère de la Justice et de la Sécurité, novembre 2021.

Fondation Walk Free, *The Global Slavery Index*, 2018.

Walk Free Foundation, Van Dijk, JJM, Cruyff, M. & Van der Heijden, PGM ., *Monitoring Target 16.2 of the United Nations' Sustainable Development Goals; a multiple systems estimation of the numbers of presumed human trafficking victims in Slovakia in 2016-2018 by year, age, gender and form of exploitation*, 2019.

Zhang, SX, *Trafficking of Migrant Laborers in San Diego County: Looking for a*

Hidden Population, San Diego, Californie : Université d'État de San Diego, 2012.

Zhang, SX & J. Loudo -Larsen, Estimating the Size of the Human Trafficking Problem: MSE and Other Strategies, *Crime and Delinquency*, Vol 67, Numéro 13-14, pages 1-9, 2021

Annexe A

Tableaux et graphiques des estimations des victimes de la traite des êtres humains en Irlande avec intervalles de confiance ; résultats du modèle le mieux ajusté comprenant trois listes, par combinaisons de covariables par paires (S et A ; S et N ; S et E ; A et N ; A et E ; N et E).

\$SxA

	Year	nobs	Nhat	min95	max95
F:A	2014	49	61	47	77
F:A	2015	55	77	59	96
F:A	2016	65	97	79	118
F:M	2014	6	14	9	21
F:M	2015	18	29	19	42
F:M	2016	12	31	21	44
M:A	2014	13	17	11	24
M:A	2015	23	33	25	43
M:A	2016	39	36	27	48
M:M	2014	7	6	4	10
M:M	2015	6	13	8	20
M:M	2016	9	14	9	21

\$SxN

	Year	nobs	Nhat	min95	max95
F:I	2014	3	10	6	16
F:I	2015	11	21	13	32
F:I	2016	11	23	14	34
F:N	2014	52	65	49	81
F:N	2015	62	85	66	106
F:N	2016	66	105	86	129
M:I	2014	5	5	2	8
M:I	2015	4	10	5	16
M:I	2016	9	11	6	17
M:N	2014	15	19	12	26
M:N	2015	25	37	27	48
M:N	2016	39	40	29	52

\$SxE

	Year	nobs	Nhat	min95	max95
F:O	2014	8	11	7	15
F:O	2015	13	18	12	25
F:O	2016	14	20	14	28
F:S	2014	47	65	49	81
F:S	2015	60	88	67	114
F:S	2016	63	108	86	132
M:O	2014	14	17	11	24
M:O	2015	21	34	25	44
M:O	2016	38	37	27	49
M:S	2014	6	6	4	10
M:S	2015	8	13	8	20
M:S	2016	10	14	8	21

```

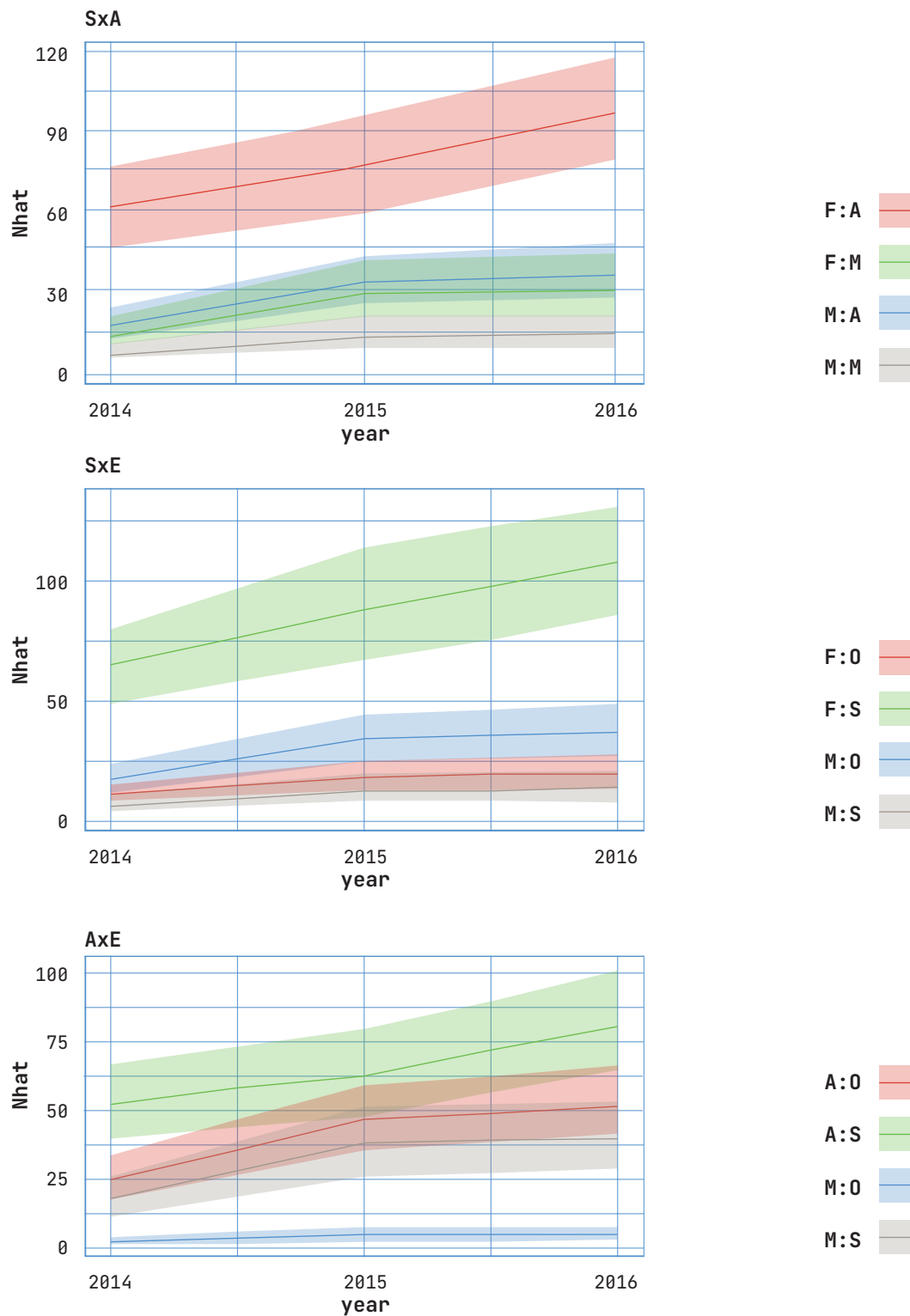
$AxN
  Year nobs Nhat min95 max95
A:I 2014   0   0   0   1
A:I 2015   0   0   0   1
A:I 2016   1   1   0   2
A:N 2014  62  78  61  96
A:N 2015  78 110  88 134
A:N 2016 103 133 112 159
M:I 2014   8  15   9  21
M:I 2015  15  31  20  43
M:I 2016  19  33  22  45
M:N 2014   5   6   3   9
M:N 2015   9  12   6  18
M:N 2016   2  13   7  20

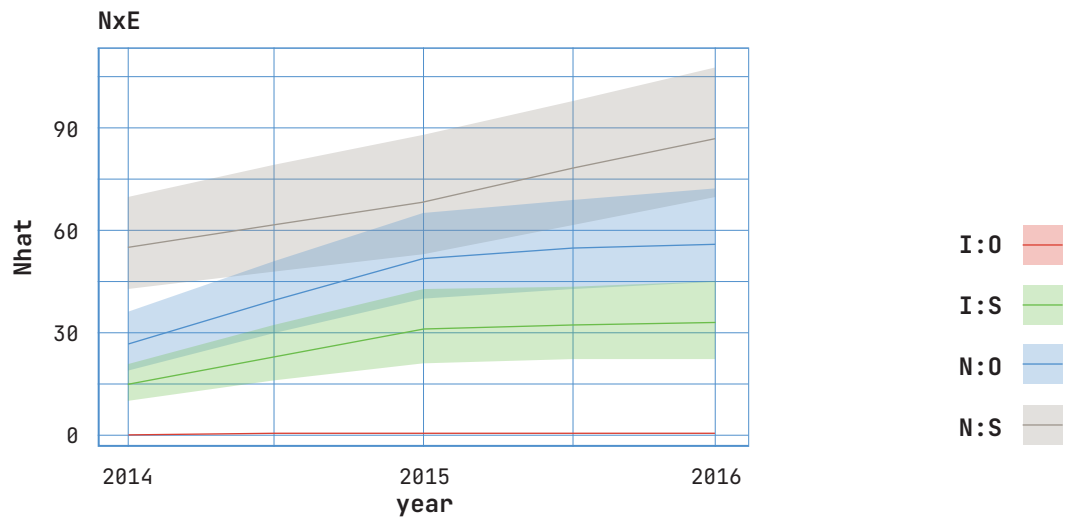
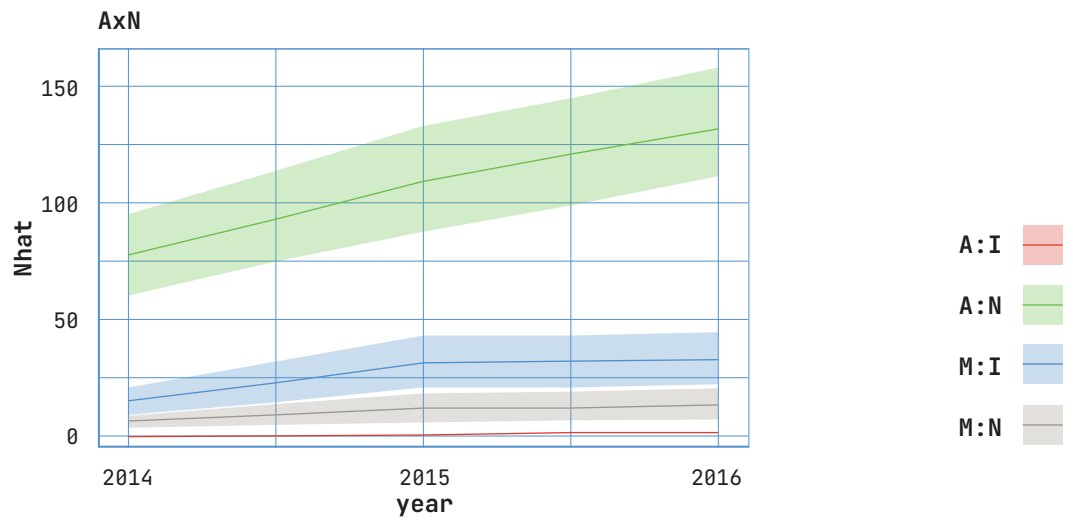
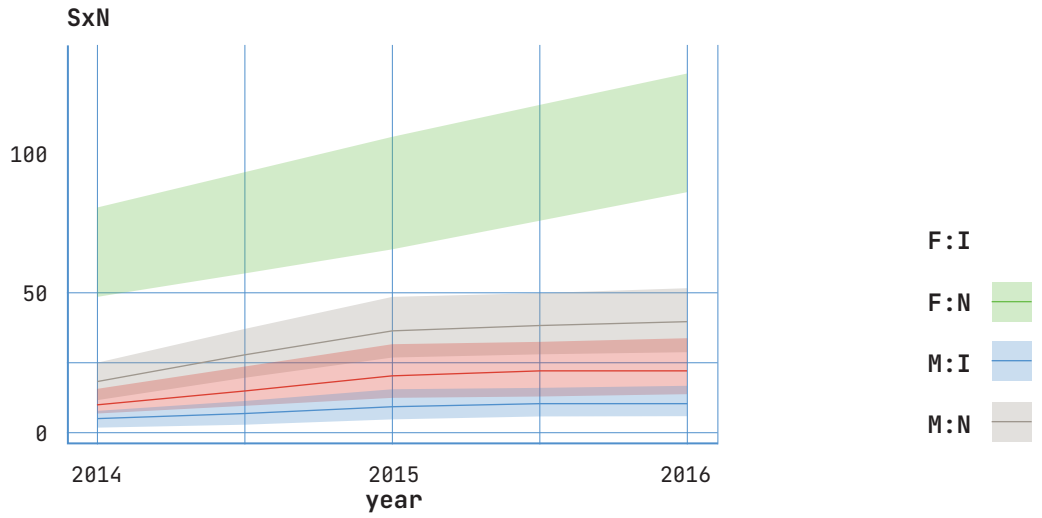
$AxE
  Year nobs Nhat min95 max95
A:O 2014  19  25  18  34
A:O 2015  31  47  36  60
A:O 2016  52  52  42  66
A:S 2014  43  53  40  67
A:S 2015  47  63  48  80
A:S 2016  52  81  65 101
M:O 2014   3   2   1   4
M:O 2015   3   5   2   8
M:O 2016   0   5   3   8
M:S 2014  10  18  12  25
M:S 2015  21  38  26  51
M:S 2016  21  40  29  54

$NxE
  Year nobs Nhat min95 max95
I:O 2014   0   0   0   0
I:O 2015   0   0   0   0
I:O 2016   0   0   0   0
I:S 2014   8  15  10  21
I:S 2015  15  31  21  43
I:S 2016  20  33  23  45
N:O 2014  22  28  19  37
N:O 2015  34  52  40  66
N:O 2016  52  57  45  73
N:S 2014  45  56  43  71
N:S 2015  53  69  53  89
N:S 2016  53  88  71 109

```

Figure 9 Estimation du nombre de victimes en Irlande selon le modèle le mieux ajusté comprenant 3 listes, quatre covariables (S, A, N et E) et trois ans par combinaisons par paires de covariables.





Modèle 15 d'estimation des paramètres

	est
(Intercept)	-24.3
R11	-0.3
R21	-0.2
R31	-1.3
SM	2.3
AM	4.7
NN	25.6
ES	21.9
Y1	6.6
Y2	-3.3
AM:NN	-6.4
SM:ES	-3.7
R31:SM	-1.3
NN:ES	-20.4
R31:AM	-3.4
SM:NN	-2.2
R21:ES	-1.2
R31:Y1	-6.6
R31:Y2	9.7
R11:SM	1.1
R21:R31	-1.3
R21:AM	-2.8
R31:ES	1.0



UNODC

United Nations Office on Drugs and Crime