

N°26
Juin 2024

Méthode

Les bases statistiques du SSMSI sur la délinquance enregistrée

Auteur : Laurianne Salembier
Rédacteur en chef : Gwennaël Solard

Table des matières

1. Sources de données	3
2. La statistique institutionnelle historique – « Etat 4001 »	5
3. Contenu et périmètre des bases statistiques.....	6
3.1. Contenu des bases statistiques	6
3.2. Périmètre et champ géographique	8
3.3. Dates de référence des bases statistiques	8
4. Calendrier de constitution.....	10
5. Limites et précautions d'usage.....	12
5.1. Données administratives.....	12
5.2. Données enregistrées.....	13
5.3. Suivi des victimes et des mis en cause	14
6. Contrôles de cohérence, redressements et apurements.....	15
6.1. Communes.....	15
6.2. Dates incohérentes.....	15
6.3. Homicides et tentatives d'homicides	15
6.4. Mauvaises indexations	17
6.5. Imputation de valeurs manquantes	18
6.6. Délits routiers et contraventions enregistrés par la gendarmerie nationale.....	19
7. Avantages des bases statistiques par rapport aux données administratives d'origine	20
8. Perspectives.....	20
8.1. Expertises complémentaires	20
8.2. Appariements et enrichissements.....	21
8.3. Modalités d'accès pour les chercheurs	21
Bibliographie.....	22
Textes relatifs au traitement des données à caractère personnel.....	23

Depuis sa création fin 2014 et plus particulièrement à partir de 2016 (où les données administratives s'enrichissent), le SSMSI constitue et exploite des bases de données statistiques sur la délinquance enregistrée produites à partir des données administratives issues des procédures enregistrées par les services de police et de gendarmerie nationales (saisies en continu dans les logiciels de rédaction de procédures), des procès-verbaux électroniques (y compris les amendes forfaitaires délictuelles) et des plaintes pour e-escroquerie enregistrées en ligne via la plateforme THESEE. Il s'agit de procédures relatives à des infractions pénales, avant leur transmission à l'autorité judiciaire qui est susceptible de requalifier ces infractions par la suite. Ces infractions ont pu être constatées suite à une plainte déposée par une victime, un signalement, un témoignage, un flagrant délit, une dénonciation, etc., mais aussi sur l'initiative des forces de sécurité intérieure. La comptabilisation des infractions enregistrées peut fournir une indication du volume réel des infractions commises, et donc de l'insécurité qui en découle, dans les domaines où le taux de plainte est important.

Les bases statistiques du SSMSI sur la délinquance enregistrée sont ainsi utilisées afin d'élaborer des statistiques et des analyses sur la délinquance et l'insécurité enregistrées (publication de résultats statistiques, production d'études, réponse aux demandes...). Elles permettent de caractériser les infractions (lieu des faits, date des faits, nature de l'infraction, mode opératoire...), les personnes associées comme victimes ou comme mis en cause (âge, sexe, nationalité...). Trois bases statistiques, disponibles depuis 2016, existent actuellement : la base statistique Victimes, la base statistique Mis en cause et la base statistique Infractions.

Le processus de production statistique du SSMSI se perfectionne au fur à mesure des années. Le service a notamment mené une refonte du processus de production de ses bases statistiques, qui couvrent depuis 2022 tous les crimes et délits non routiers enregistrés ou élucidés par les services de police et de gendarmerie. La base statistique Infractions couvre également tous les délits routiers et les contraventions tandis que, sur ce périmètre, les bases statistiques Victimes et Mis en cause couvrent uniquement ceux relevés par les services de police. Auparavant, ne figuraient dans ces bases que les infractions, victimes et mis en cause entrant dans le périmètre historique de l'État 4001 (voir la partie 2 et l'éclairage 1 de l'Interstats Références « Insécurité et délinquance en 2021 : bilan statistique »).

Ce document a vocation à être mis à jour au fil de l'amélioration du processus de production des bases statistiques du SSMSI, que ce soit suite au développement de nouveaux redressements, à l'évolution des sources de données administratives en entrée, ou à toute autre évolution.

1. Sources de données

Les données administratives utilisées par le SSMSI pour constituer ses bases statistiques sont les informations contenues dans les procédures pénales enregistrées par la police et la gendarmerie nationales. Ces informations sont majoritairement saisies dans les logiciels de rédaction de procédures de la police et de la gendarmerie nationales (LRPPN et LRPNG). Le SSMSI s'appuie également sur les données issues des procès-verbaux électroniques, y compris les amendes forfaitaires délictuelles – infractions relevées avec les terminaux NEO (terminal mobile police gendarmerie) – collectées par l'Agence nationale de traitement automatisé des infractions (Antai). Enfin, il intègre les données issues des plaintes pour e-escroqueries déposées en ligne via la plateforme THESEE. Chaque mois, le SSMSI extrait de ces sources de données les informations utiles à l'établissement des statistiques sur l'insécurité et la délinquance.

Le SSMSI a accès aux données détaillées saisies dans LRPPN sur les infractions¹, les victimes et les mis en cause associés, quelle que soit la classification de l’infraction (crime, délit ou contravention) et qu’elle soit volontaire ou involontaire. Il n’a en revanche pas accès aux pièces jointes des procédures (procès-verbaux, rapports de synthèse, etc.). Le SSMSI a également accès aux données détaillées sur les délits et contraventions, les mis en cause et les éventuelles victimes associés, issues des procès-verbaux électroniques (PVE) dressés depuis 2016. Les informations sur les mis en cause sont toutefois moins souvent renseignées pour les contraventions (notamment dans le cas des infractions relevées par radar). Le SSMSI a également accès aux données détaillées sur les infractions et leurs victimes, issues des plaintes déposées sur la plateforme THESEE depuis 2022 (figure 1).

Côté gendarmerie, le SSMSI a accès aux données détaillées sur les infractions², les victimes et mis en cause associés, uniquement pour les infractions correspondant à des crimes ou des délits autres que routiers³ (champ de l’État 4001 historique, cf. Partie 2). Il dispose de comptages agrégés sur les infractions correspondant à des délits routiers ou des contraventions, et ne dispose d’aucune information sur leurs victimes et leurs mis en cause. En effet côté gendarmerie, dans un premier temps, un procès-verbal est rédigé à l’aide du logiciel de rédaction de procédures de la gendarmerie nationale (LRPGN). Le gendarme enregistre une ou plusieurs infractions, et si l’une au moins de ces infractions entre dans le champ des statistiques de l’Etat 4001 (crimes et délits non routiers, cf. Partie 2), le logiciel impose la création d’un message d’information statistique (MIS). À ce stade, le SSMSI n’a accès qu’aux données détaillées renseignées dans le message d’information statistique, et non à l’exhaustivité des données saisies dans LRPGN, ces dernières n’étant pas centralisées dans une base informatique nationale (cf. Partie 6.6).

Figure 1 : Niveau de disponibilité de l’information, selon la source et le type d’infraction

	Saisies LRPPN	Saisies LRPGN	MIS (GN)	PVE	THESEE
Crimes	Données détaillées sur les infractions, les victimes et les mis en cause	Comptages agrégés sur les infractions	Données détaillées sur les infractions, les victimes et les mis en cause		
Délits non routiers				Données détaillées sur les infractions, les victimes et les mis en cause	Données détaillées sur les infractions et les victimes
Délits routiers					
Contraventions				Données détaillées sur les infractions ; données partielles sur les mis en cause	

Cf. Partie sur les textes relatifs au traitement des données à caractère personnel

¹ La nature de l’infraction détaillée (NATINF) est disponible depuis 2015.

² La nature de l’infraction détaillée (NATINF) est disponible depuis 2016.

³ Par abus de langage, le terme de délit routier intègre également les délits liés à la navigation maritime ou aérienne.

2. La statistique institutionnelle historique – « Etat 4001 »

A partir de 1972, les forces de sécurité se sont dotées d'un outil standardisé de mesure de l'activité judiciaire des services basé sur des comptages mensuels, appelé « Etat 4001 ». Ce document administratif porte sur les crimes et les délits non routiers, enregistrés pour la première fois par les services de police ou de gendarmerie nationales (afin d'éviter une double comptabilisation si une même infraction est traitée successivement par des services différents, mais les pratiques d'enregistrement ne respectent pas toujours les consignes méthodologiques) et portés à la connaissance de l'institution judiciaire (n'y sont donc retracées que les infractions suffisamment constituées juridiquement pour pouvoir être poursuivies par un tribunal). Les infractions ne sont pas toutes comptabilisées dans l'Etat 4001, des conventions spécifiques étant mises en œuvre⁴ (voir Interstats Méthode n°2). Enfin, les infractions routières et les contraventions sont exclues de ce dispositif.

Les infractions y sont regroupées en 103 catégories (sur 107 possibles), très hétérogènes par la nature et la gravité des faits, mais aussi par le nombre d'infractions constatées chaque mois. On y trouve aussi bien les « homicides commis sur des mineurs de moins de 15 ans » (catégorie qui compte environ une cinquantaine de victimes enregistrées chaque année) que les « coups et blessures volontaires criminels ou correctionnels sur personne de 15 ans et plus », l'infraction de « non versement de pension alimentaire » ou encore les « infractions relatives à la chasse et à la pêche ». Les critères de différenciation entre les postes de cette nomenclature font souvent référence à l'incrimination pénale constitutive du crime ou du délit, mais aussi parfois au type de victime (les mineurs de moins de 15 ans sont souvent spécifiés, ainsi que les particuliers, voire les femmes, ou certains groupes professionnels), au mode opératoire (le cambriolage est spécifié, ainsi que le « vol à la tire ») ou certains lieux de commission de l'infraction (lieux publics, domiciles, etc.). Numérotés de 1 à 107 (quatre positions ne sont pas utilisées, on recense donc 103 groupes d'infractions), cette nomenclature, qui n'a pas évolué depuis 1972, est appelée couramment « les 107 index de l'Etat 4001 ». Ce sont les séries suivies historiquement par le ministère de l'Intérieur. Chaque index a sa propre unité de compte : il peut s'agir d'un nombre de victimes, d'un nombre de plaignants, d'un nombre de chèvres, d'un nombre de mis en cause, etc.

La statistique institutionnelle historique (Etat 4001) est calculée en début de mois, à partir des données opérationnelles. Les données opérationnelles sont les données détaillées issues des saisies journalières des policiers et des gendarmes, elles correspondent aux informations contenues dans une procédure et évoluent constamment au fil du déroulement de l'enquête ou de l'instruction. Le SSMSI s'appuie sur ces mêmes données opérationnelles pour constituer ses bases statistiques, qui contiennent des éléments constitutifs de l'Etat 4001 (index de l'infraction, nombre de faits constatés, dates de référence, etc.).

12 compteurs sont calculés chaque mois pour chaque index : nombre de faits constatés, nombre de faits élucidés, nombre de mis en cause (femmes/hommes, français/étrangers, majeurs/mineurs, libres/écroués), nombre de gardes à vue (moins de 24 heures/plus de 24 heures). Ces indicateurs sont comptabilisés sur le mois écoulé, en cumul depuis le début de l'année, en cumul sur l'année glissante.

⁴ En théorie, seules les infractions principales doivent être comptabilisées. En pratique, la police nationale comptabilise les infractions principales et les infractions secondaires. Le guide de méthodologie statistique, qui dicte les règles statistiques pour l'établissement de l'Etat 4001, est en cours de révision pour comptabiliser les infractions principales et secondaires pour la police nationale comme pour la gendarmerie nationale.

L'Etat 4001 est figé : il s'agit d'une photographie agrégée des données opérationnelles en début de mois. Cela signifie qu'une fois que les compteurs agrégés sont calculés pour un mois, ils ne sont plus recalculés même si des faits sont requalifiés ou si le nombre d'unités de compte évolue pour une infraction.

3. Contenu et périmètre des bases statistiques

3.1. Contenu des bases statistiques

La base statistique Infractions décrit l'ensemble des infractions (principales et secondaires) commises sur le Territoire de la République française (soit France + Collectivités d'Outre-mer) et relevées lors de l'établissement du procès-verbal ou de l'enregistrement de la plainte par les services de police et de gendarmerie nationales ainsi que les infractions élucidées par les forces de sécurité intérieure. Ces infractions sont de nature criminelle, délictuelle (y compris les délits routiers) ou contraventionnelle. Elles sont caractérisées par une nature d'infraction détaillée (NATINF). La base statistique Infractions intègre la date, l'heure et l'adresse de commission de l'infraction, la date d'enregistrement de l'infraction, l'éventuelle date d'élucidation ainsi que des informations sur le lieu de commission (par exemple, les transports en commun ou la nature détaillée du lieu) et sur les caractéristiques de l'infraction (classification, index de l'Etat 4001, degré de réalisation, nature de l'infraction, etc.). Elle donne également des informations sur la procédure (service l'ayant enregistrée, date d'ouverture, date de clôture, etc.). Pour les procédures enregistrées par les services de police, on dispose d'éléments sur le mode opératoire, le mobile, le contexte, le cadre juridique de la procédure, le type de saisine ou la nature de clôture. C'est à partir de l'adresse de commission des infractions qu'est établie la base géocodée des infractions. La base statistique Infractions porte sur les infractions portées à la connaissance des forces de sécurité, ainsi que sur les infractions relevées dans le cadre de l'action des services : certaines victimes ne portent pas plainte, les infractions qu'elles ont subies ne figurent donc pas dans les données administratives enregistrées par la police ou la gendarmerie.

L'enquête Cadre de Vie et Sécurité (CVS), conduite entre 2007 et 2021 par l'Insee en partenariat avec l'Observatoire national de la délinquance et des réponses pénales (ONDRP, supprimé en 2020) et avec le SSMSI (créé en 2014), puis l'enquête Vécu et Ressenti en matière de Sécurité (VRS), conduite par le SSMSI à partir de 2022, ont pour objectif d'évaluer et de décrire l'ensemble des atteintes/infractions dont sont victimes les ménages et les individus. Elles complètent les données administratives sur les infractions enregistrées par les forces de sécurité, et permettent notamment de mesurer les taux de dépôt de plainte : la proportion de victimes ayant déclaré avoir déposé plainte dans un commissariat de police ou une brigade de gendarmerie.

La base statistique Victimes décrit l'ensemble des victimes de crimes, délits ou contraventions (hors délits routiers et contraventions pour les infractions enregistrées par la gendarmerie nationale, à l'exception des infractions relevées par PVE) commis sur le Territoire de la République française (soit France + Collectivités d'Outre-mer) et enregistrés par les services de police et de gendarmerie nationales. Ce sont des personnes enregistrées en tant que victimes dans des procédures de la police ou de la gendarmerie nationales, pour des infractions portées à la connaissance des services : certaines victimes ne portent pas plainte, d'autres sont enregistrées même si elles n'ont pas porté plainte. Par

exemple, dans le cas d'une victime mineure, c'est son représentant légal qui porte plainte ; les parents d'un mineur peuvent également déposer plainte en son nom sans son accord ; une personne tierce peut signaler une infraction dont elle est témoin (y compris dans le cas d'une personne morale) ; une victime peut être identifiée suite à l'action directe des services de sécurité, etc. La base statistique Victimes couvre l'ensemble des victimes, à la fois pour l'infraction principale qu'elles ont subie et pour les autres infractions subies (infractions secondaires). Si une même victime a subi plusieurs infractions, elle apparaîtra autant de fois dans la base statistique Victimes avec les caractéristiques de chaque infraction subie. La base statistique Victimes intègre des informations sur l'adresse de résidence, le sexe, la date de naissance, l'âge au moment des événements, la profession, la nationalité et le pays de naissance des victimes personnes physiques. Elle donne également le lien entre la victime et l'auteur de l'infraction. Elle couvre également les victimes personnes morales et contient le secteur d'activité de celles-ci. Elle donne des éléments sur les infractions subies (nature de l'infraction – NATINF –, date, lieu, caractéristiques, etc.). Pour les seules procédures enregistrées par les services de police nationale, on dispose d'éléments sur le préjudice corporel subi par la victime et sur des caractéristiques spécifiques (handicap, grossesse, personne âgée, touriste, auto-stoppeur, mineur isolé, etc.), ainsi que sur l'existence ou non d'un dépôt de plainte par la victime.

La base statistique Mis en cause décrit l'ensemble des mis en cause pour des infractions élucidées par les services de police et de gendarmerie nationales (hors délits routiers et contraventions enregistrés par les services de gendarmerie), donc commises sur le Territoire de la République française (soit France + Collectivités d'Outre-mer). Ce sont les mis en cause identifiés par les forces de sécurité : dans le cadre de l'enquête, des indices graves ou concordants rendent vraisemblable leur participation comme auteur ou complice à la commission d'une infraction. Leur identité est alors transmise à l'autorité judiciaire. La base statistique Mis en cause couvre les personnes mises en cause, à la fois pour l'infraction principale qu'elles ont commise et pour les autres infractions commises (infractions secondaires). Si une personne est mise en cause pour plusieurs infractions, elle apparaîtra autant de fois dans la base statistique Mis en cause avec les caractéristiques de chaque infraction commise. La base statistique Mis en cause intègre des informations sur l'adresse de résidence, le sexe, la date de naissance, l'âge au moment des événements, la profession, la nationalité et le pays de naissance des mis en cause personnes physiques, ainsi que l'issue pour le mis en cause (libre, déferé, en fuite, etc.). Elle couvre également les mis en cause personnes morales et contient le secteur d'activité de celles-ci. Elle donne des éléments sur les infractions commises (nature de l'infraction – NATINF –, date, lieu, caractéristiques, etc.). Pour les seules procédures enregistrées par les services de police nationale, on dispose d'éléments sur la catégorie pénale du mis en cause, les caractéristiques spécifiques de celui-ci (sans domicile fixe, mineur isolé, détenu, évadé, personne vulnérable, etc.), s'il est auteur ou complice, le fait qu'il a été entendu ou non et si non, la raison pour laquelle il n'a pas été entendu.

Les bases statistiques du SSMSI sont essentiellement constituées d'informations disponibles dans les logiciels de rédaction de procédures de la police et de la gendarmerie nationales. Certains champs disponibles sont obligatoirement saisis, d'autres sont facultatifs. Le caractère obligatoire ou non des différents champs dépend de l'origine des données (police ou gendarmerie, PVE, THESEE ou procédures classiques, ...), de la classification de l'infraction (crimes, délits ou contraventions) et de l'appartenance ou non au champ de la statistique institutionnelle historique (Etat 4001) : par exemple, la profession de la victime est obligatoirement renseignée dans LRPGN, facultative dans LRPPN ou THESEE et n'est pas collectée pour les infractions relevées via PVE (cf. partie 5.1). Elles sont complétées par des données sur les zonages géographiques (taille de l'unité urbaine de la commune de commission

des faits, compétence – police ou gendarmerie), par la nomenclature française des infractions et la classification internationale des infractions (NFI et ICCS, voir Courrier des statistiques N7 – Camus, janvier 2022) et par des indicatrices thématiques permettant de repérer certains champs spécifiques de la délinquance : infractions intrafamiliales, infractions commises dans les transports en commun, infractions liées au numérique, infractions anti-LGBT, infractions à caractère raciste, xénophobe ou antireligieux, infractions qualifiées de sexisme, etc.

3.2. Périmètre et champ géographique

Les bases statistiques du SSMSI couvrent les infractions enregistrées ou élucidées par la police et la gendarmerie nationales. Sont actuellement hors champ, les infractions relevées par les polices municipales et celles relevées par les opérateurs de transport, ainsi que la majorité des infractions de la délinquance environnementale (voir Interstats Analyse n° 46), de la délinquance douanière et de la délinquance économique et financière. Ces infractions sont en effet majoritairement constatées par des institutions autres que la police ou la gendarmerie nationales : services des douanes, services en charge de la répression des fraudes à la consommation (DGCCRF), ceux en charge des fraudes sociales (URSSAF), inspections du travail et des lois sociales, etc. Seules celles donnant lieu à investigation par les services de police ou de gendarmerie nationales font actuellement partie du champ des bases statistiques du SSMSI. Les actes de police administrative et les informations recueillies via une main courante sont également hors champ.

Les bases statistiques couvrent les infractions commises sur le Territoire de la république française (soit France + Collectivités d’Outre-mer), exceptés les enregistrements de janvier 2016 par la gendarmerie dans les DROM et les COM. Pour les faits commis dans les transports, c’est la commune d’arrivée qui est prise en compte comme « lieu de commission de l’infraction ». Les faits commis à l’étranger en sont exclus : la plupart font l’objet d’un enregistrement dans le pays où ils se sont produits, seuls des cas résiduels sont présents dans les données de la police et de la gendarmerie nationales.

3.3. Dates de référence des bases statistiques

Lors de la saisie d’une procédure, toutes les infractions ayant eu lieu sur une même période (définie par une date de début et une date de fin) et au même endroit font partie du même événement. La date de début de commission de ces infractions est la date de début des faits. Lorsque plusieurs victimes sont associées à un événement, et que l’événement s’est étalé sur une longue période, la date de début des faits peut sembler incohérente avec l’âge de certaines des victimes. Par exemple, dans une affaire de pédophilie, où toutes les victimes sont décrites sous un même événement couvrant une longue période, la date de début des faits est dans ce cas la date à partir de laquelle les événements se sont produits pour la première des victimes.

Cette date de début des faits ne peut pas être la date de référence utilisée pour constituer les bases statistiques annuelles : en effet, il peut y avoir un laps de temps plus ou moins long entre la date à laquelle s’est produit un événement et la date à laquelle cet événement sera porté à la connaissance des forces de sécurité. En raisonnant ainsi sur la date de commission des infractions, il faudrait alors construire la base statistique d’une année donnée, longtemps après l’année de début des faits, sans

jamais atteindre l'exhaustivité des faits commis cette année donnée et enregistrés par la police ou la gendarmerie. Les délais de dépôt de plainte peuvent être relativement longs pour certains types d'infractions. Par exemple en 2022, les délais médians de dépôt de plainte – au bout desquels la moitié des personnes ont déposé plainte – sont de 195 jours pour les violences sexuelles et de 17 jours pour les escroqueries ; ils sont inférieurs à 5 jours pour les vols, cambriolages, destructions et dégradations volontaires, etc. (voir l'éclairage 2 de l'Interstats Références « Insécurité et délinquance en 2022 : bilan statistique »). C'est pourquoi les bases statistiques sont constituées selon la date d'enregistrement des infractions. Il s'agit de la date à laquelle une infraction est, pour la première fois, portée à la connaissance de la police ou de la gendarmerie nationale. Ainsi, la base statistique Victimes d'une année donnée permet de décrire l'ensemble des victimes d'infractions commises sur le Territoire de la république française (soit France + Collectivités d'Outre-mer) et portée pour la première fois à la connaissance des forces de sécurité cette année-là.

Dans les données provenant des saisies de la police nationale, cette date d'enregistrement correspond à la date à laquelle l'infraction est comptabilisée pour la première fois dans les statistiques de l'Etat 4001. Cette date s'enclenche dès lors que toutes les informations nécessaires à l'établissement des statistiques pour cette infraction sont renseignées. Dans les données de la gendarmerie nationale, la date d'enregistrement est approchée par la date de première validation du message d'information statistique (MIS) associé à la procédure⁵, message servant à alimenter les statistiques de l'Etat 4001 (cf. infra). Cette date ne s'enclenche pas dès la saisie de l'infraction dans le logiciel de rédaction des procédures : elle nécessite au préalable la validation d'un commandant et d'un exécutant de la brigade départementale de renseignements et d'investigations judiciaires de la gendarmerie. Concernant les infractions provenant des procès-verbaux électroniques, cette date est approchée par la date à laquelle l'Antai transmet les données au service informatique du ministère de l'intérieur, pour intégration dans le système d'information du ministère. Enfin, concernant les e-escroqueries enregistrées via la plateforme en ligne THESEE, la date de référence utilisée est la date de validation de la plainte par le service de police gérant la plateforme.

La date de premier enregistrement est utilisée à la fois pour décrire les victimes d'infractions commises sur le Territoire de la république française (soit France + Collectivités d'Outre-mer), mais aussi pour caractériser les infractions commises et établir des statistiques sur le niveau de la délinquance enregistrée une année donnée.

Cette date de premier enregistrement ne peut pas être la date de référence pour décrire les personnes mises en cause pour des infractions commises sur le Territoire de la république française (soit France + Collectivités d'Outre-mer) : les forces de sécurité sont chargées quand elles constatent ou qu'on leur signale un crime ou un délit, d'en rechercher les auteurs sous l'autorité du Procureur de la République. Or au cours de l'enquête, de nouvelles personnes peuvent être mises en cause ou au contraire, des personnes initialement mises en cause peuvent être mises hors de cause. En prenant pour année de référence la date d'élucidation de l'infraction, on repère l'ensemble des personnes ayant été entendues par la police ou la gendarmerie et à l'encontre desquelles sont réunis, dans la procédure transmise au parquet, des éléments graves et concordants attestant qu'elles ont commis ou tenté de commettre une ou plusieurs infractions. Ainsi la base statistique Mis en cause d'une année donnée permet de décrire l'ensemble des personnes mises en cause pour des infractions commises sur le

⁵ Si une infraction est ajoutée ultérieurement à la procédure, la date d'enregistrement de cette infraction sera la date de première validation du MIS de LRPGN contenant cette infraction.

Territoire de la république française (soit France + Collectivités d’Outre-mer), quelle que soit la date de commission, et élucidées par la police ou la gendarmerie cette année-là.

Dans les données provenant des saisies de la police, cette date d’élucidation correspond à la date de rédaction du compte-rendu d’enquête initial (partiel ou définitif). Dans celles de la gendarmerie, elle est approchée par la date de clôture du message d’information statistique (MIS) associé à la procédure, message servant à alimenter les statistiques de l’État 4001. Concernant les infractions provenant des procès-verbaux électroniques, cette date est approchée par la date de transmission des données par l’Antai (la date d’élucidation est concomitante à la date d’enregistrement). Enfin concernant les e-escroqueries enregistrées via la plateforme en ligne THESEE, ces infractions ne sont que constatées par le service de police gérant la plateforme. Elles sont ensuite transférées et éventuellement élucidées ultérieurement par un autre service, dans le cadre d’une procédure enregistrée dans les logiciels de rédaction des procédures et intègrent la base statistique Mis en cause par ce biais.

La date d’élucidation est utilisée à la fois pour décrire les mis en cause pour des infractions commises sur le Territoire de la république française (soit France + Collectivités d’Outre-mer) et élucidées par les forces de sécurité, mais aussi pour caractériser les infractions élucidées et établir des statistiques sur les taux d’élucidation (voir Interstats Méthode n° 18).

La base statistique Infractions contient les infractions commises sur le Territoire de la république française (soit France + Collectivités d’Outre-mer), en date d’enregistrement comme en date d’élucidation, afin de pouvoir fournir de l’information sur l’ensemble des infractions qu’il y ait ou non une victime identifiée, et qu’il y ait ou non une personne identifiée comme mise en cause.

Le champ temporel de chaque base statistique est présenté dans la **figure 2**.

Figure 2 : champ temporel retenu pour chaque base statistique

Base statistique Infractions de l’année N	Ensemble des infractions enregistrées l’année N ou élucidées l’année N
Base statistique Victimes de l’année N	Ensemble des victimes d’infractions enregistrées l’année N
Base statistique Mis en cause de l’année N	Ensemble des mis en cause pour des infractions élucidées l’année N

4. Calendrier de constitution

Chaque début de mois, des bases statistiques « cumul » mensuelles sont constituées à partir des infractions enregistrées ou élucidées depuis le début de l’année en cours (**figure 3**). Toutes les données de l’année courante sont ré-extraites chaque mois. Lors de l’enregistrement des infractions par les services de sécurité, il n’est pas rare qu’une infraction soit requalifiée dans le cadre de l’avancement de l’enquête ou de l’instruction. En effet suite aux investigations, les informations initiales peuvent évoluer, auquel cas certaines infractions changent de nature d’infraction, d’autres peuvent être annulées ou des éléments constitutifs peuvent être modifiés (par exemple, le nombre de victimes de l’infraction). Dans les bases statistiques du SSMSI, les infractions sont donc requalifiées jusqu’à la date de l’extraction. La date d’extraction des bases statistiques est une information essentielle : en effet, elle donne une indication sur la durée pendant laquelle les données ont pu être requalifiées par les forces de sécurité dans le cadre de l’avancement de l’enquête.

Des bases statistiques annuelles provisoires sont constituées au début du mois de janvier de l'année suivante. L'effet des requalifications sur les indicateurs suivis par le SSMSI s'amenuise au fil du temps, plus ou moins rapidement selon le type d'infractions considéré, et à partir d'un certain délai les données ne sont quasiment plus requalifiées et peuvent être considérées comme définitives. C'est pourquoi depuis le millésime 2022 des bases statistiques, le SSMSI réalise une deuxième extraction annuelle des données début mai de l'année suivante pour constituer les bases statistiques annuelles définitives. La date de cette extraction résulte d'un compromis entre fraîcheur et qualité des données (voir Interstats Méthode n° 23).

Figure 3 : Calendrier de constitution des bases statistiques du SSMSI

Date de l'extraction	Extraction des données concernant le mois de ...						
	Janvier N	Février N	Mars N	...	Novembre N	Décembre N	
Début février N							Bases « cumul » mensuelles
Début mars N							
Début avril N							
...							
Début décembre N							
Début janvier N + 1							Bases annuelles provisoires
Début mai N + 1							Bases annuelles définitives

Dans les bases statistiques, les requalifications sont différenciées selon le mois d'enregistrement : par exemple dans les bases annuelles provisoires, les données enregistrées en janvier N ont été requalifiées pendant 11 à 12 mois, celles enregistrées en février l'ont été pendant 10 à 11 mois, ..., les données enregistrées en décembre ont été requalifiées au plus pendant un mois (figure 4).

Figure 4 : Requalifications différenciées dans les bases statistiques annuelles provisoires

Mois d'enregistrement	Nombre de mois de requalifications potentielles
Janvier N	11 à 12 mois
Février N	10 à 11 mois
Mars N	9 à 10 mois
Avril N	8 à 9 mois
Mai N	7 à 8 mois
Juin N	6 à 7 mois
Juillet N	5 à 6 mois
Août N	4 à 5 mois
Septembre N	3 à 4 mois
Octobre N	2 à 3 mois
Novembre N	1 à 2 mois
Décembre N	< 1 mois

Un nouveau processus de production des bases statistiques a été mis en place à partir de 2022. Sa reproduction sur les millésimes passés (2016 à 2021) a nécessité de refaire une extraction des données issues des logiciels de rédaction de procédures, car les données extraites au cours de l'ancien processus de production des bases statistiques n'étaient pas suffisantes pour la mise en œuvre du nouveau processus. Cette extraction a eu lieu en avril 2022. En faisant une réextraction complète sur toute la période, on récupère les données dans l'état où elles sont au moment de l'extraction, donc intégrant les requalifications intervenues jusqu'au moment de l'extraction. Ainsi, les données de l'année 2016 sont celles mises à jour à la date de cette extraction (avril 2022), et non telles qu'elles étaient qualifiées

début 2017. Le niveau de requalification n'est donc pas homogène pour les différentes années : les années anciennes ont eu davantage le temps d'être requalifiées que les années récentes. Or l'effet des requalifications n'est pas sans biais. Pour ne pas biaiser les résultats, il faudrait donc neutraliser cet effet « requalifications différenciées » (pour 2021, extraction en avril 2022 au lieu de mai 2022 ; pour 2020, extraction en avril 2022 au lieu de mai 2021 ; pour 2019, extraction en avril 2022 au lieu de mai 2020, etc.). Cette neutralisation n'a pu être réalisée que sur le champ des homicides, pour les années 2020 et 2021. Ce biais est cependant très limité pour la plupart des natures d'infraction, dont les requalifications se stabilisent au bout de quelques mois (voir Interstats Méthode n° 23).

5. Limites et précautions d'usage

5.1. Données administratives

Toutes les informations liées aux procédures ne sont pas toujours renseignées dans les logiciels d'enregistrement de plainte, faute d'être connues ou nécessaires aux procédures judiciaires. Ceci induit dans les chiffres produits à partir de ces données une certaine imprécision et des traitements statistiques sont nécessaires afin de fournir une information fiabilisée. La fiabilité des données contenues dans les bases statistiques dépend nécessairement de la qualité de l'information initiale.

La qualité de l'information initiale dépend de l'origine des données (police ou gendarmerie), de la source (logiciels de rédaction des procédures, PVE, plateformes), de la classification de l'infraction (crime, délit, contravention), de l'appartenance ou non de l'infraction au champ de l'Etat 4001, du caractère obligatoire ou non de la saisie, et de la présence ou non de valeurs par défaut. Par exemple, la profession des victimes et des mis en cause est une information obligatoire dans le message d'information statistique (MIS) de LRPGN, facultative dans LRPPN, et ne fait pas partie des champs de saisie de l'application PVE. Ainsi dans les bases statistiques 2023, la profession est renseignée pour 62 % des victimes et 87 % des mis en cause enregistrés par les services de police, 100 % de ceux enregistrés par les services de gendarmerie⁶ (voir Interstats Méthode n° 20).

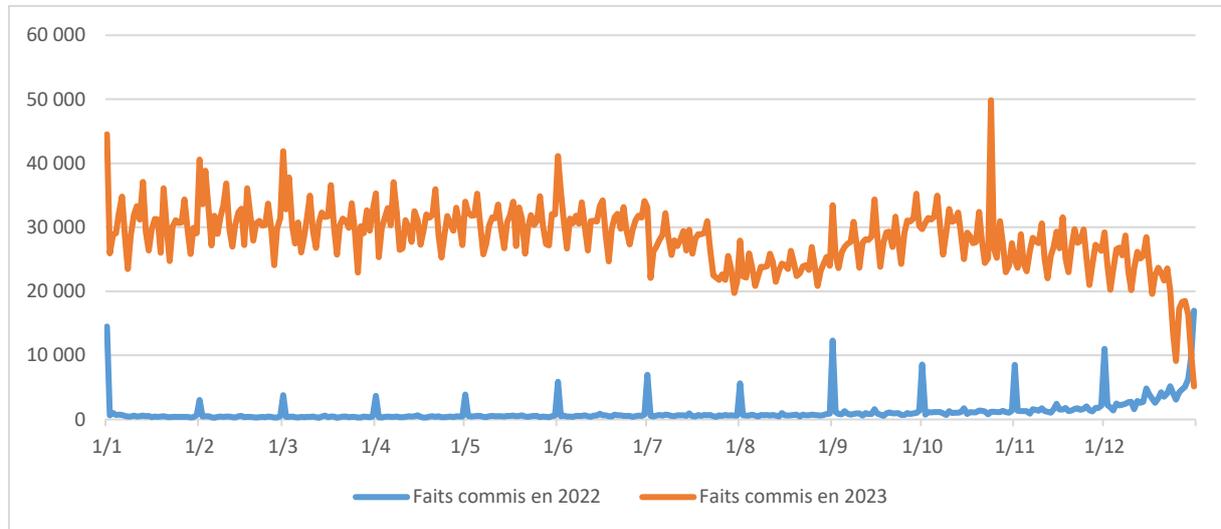
La plupart des champs de type « Date » des logiciels d'enregistrement des plaintes sont saisis manuellement par les utilisateurs, et sont donc sujets aux approximations et erreurs de saisie. Dans LRPPN, seules les dates de création ou de dernière modification d'un enregistrement sont générées automatiquement. La date d'enregistrement d'une infraction est calculée automatiquement, une fois que les informations nécessaires à l'établissement des statistiques pour cette infraction sont renseignées. Ces informations nécessaires varient selon la nature de l'infraction : par exemple pour une atteinte aux personnes, une victime doit obligatoirement être saisie. Dans LRPGN, les dates de validation ou de clôture d'un enregistrement sont générées automatiquement. Les autres champs de type « Date » de LRPPN comme de LRPGN sont libres. La date de commission des faits est particulièrement sujette aux approximations et erreurs de saisie. Ainsi, on observe :

- des pics du nombre de faits qui ont été commis le 1^e de chaque mois. En effet, lorsque la personne qui dépose plainte indique un événement qui s'est produit en août 2023 sans en préciser le jour, le plaignant aura tendance à renseigner la date des faits au 1^e août 2023.

⁶ L'exhaustivité n'est néanmoins pas une garantie de qualité de l'information renseignée.

- des erreurs de saisie sur l'année des faits, en début d'année : en janvier 2024, saisie d'un fait au 2 janvier 2023 au lieu du 2 janvier 2024 (figure 5).

Figure 5 : Nombre d'infractions commises chaque jour de 2022 et 2023, parmi celles enregistrées en 2023.



Source : SSMSI – base statistique des infractions enregistrées par la police et la gendarmerie en 2023.

Par ailleurs, les enregistrements peuvent être sujets à des erreurs de saisie de la part des policiers et des gendarmes et font l'objet de travaux de contrôle qualité par le SSMSI (cf. Partie 6). Par exemple, le pic observé le 24 octobre 2023 dans la figure ci-dessus correspond à une erreur d'une brigade de gendarmerie, qui a enregistré près de 23 000 infractions de circulation avec un véhicule terrestre à moteur sans assurance ce jour-là (contre environ 600 par jour habituellement, tous services confondus). Cette erreur a ensuite été corrigée dans les bases statistiques du SSMSI.

Certaines informations sur les personnes mises en cause doivent être maniées avec précaution : l'agent qui enregistre la plainte n'a la possibilité de renseigner qu'une seule nationalité dans les logiciels de rédaction de procédures. Il n'est donc pas possible de savoir si les mis en cause disposent de plusieurs nationalités (il en est de même pour les victimes). Ainsi, parmi les mis en cause identifiés comme étant de nationalités étrangères, certains peuvent aussi disposer de la nationalité française et inversement. Par ailleurs, l'identité, la nationalité et l'âge des mis en cause ne peuvent pas être systématiquement vérifiés par les services de police et de gendarmerie, faute de disposer de pièces d'identité sécurisées ou en l'absence de toute pièce d'identité. Or, les mis en cause peuvent avoir intérêt, par exemple, à déclarer aux services de sécurité un âge inférieur à leur âge réel afin d'échapper à une sanction pénale.

5.2. Données enregistrées

Les bases statistiques s'appuient sur les enregistrements d'infractions par les forces de sécurité, or les taux de plainte varient selon les champs infractionnels. Par exemple, ils sont plus faibles pour les violences à la personne et plus élevés pour les cambriolages « aboutis ». Ces bases statistiques ne permettent donc pas de mesurer exhaustivement le niveau de la délinquance subie par la population. Par ailleurs, les enregistrements des forces de sécurité dépendent des pratiques et des consignes de gestion ou de saisie.

Les taux de plainte pour certains groupes d'infractions peuvent évoluer dans le temps et influencer sur le nombre d'infractions enregistrées par les forces de sécurité : par exemple, la hausse du nombre de coups et blessures volontaires s'exerçant dans un contexte intrafamilial peut refléter une déclaration plus importante des faits à la police et à la gendarmerie des violences physiques subies, ainsi qu'une meilleure prise en compte de la parole des victimes au moment de l'accueil dans les services.

Les effets d'actualité impactent également le volume des déclarations aux forces de sécurité. Par exemple, la hausse des violences sexuelles depuis 2018, dans un contexte de libération de la parole et de prise de conscience collective des violences faites aux femmes nées de l'affaire Weinstein et de la vague #MeToo qui s'en est suivie, procède sans doute à la fois d'une plus grande révélation des faits et d'un abaissement du seuil de tolérance face à ce type de violences.

Enfin, dans les données enregistrées par la police et la gendarmerie, il existe un décalage temporel entre la survenue des événements et le moment où ils sont portés à la connaissance des forces de sécurité. Dans le cas d'un dépôt de plainte par la victime ou son représentant légal, ce décalage correspond au délai que met la personne pour se rendre dans un commissariat de police ou une brigade de gendarmerie après l'infraction ; ce qui peut dépendre de contraintes ou de choix personnels (opportunité de porter plainte, disponibilité, etc.) ou d'accessibilité aux services de sécurité (voir Interstats Analyse n° 61). Dans le cas où l'infraction est directement constatée par les services de sécurité, à leur initiative ou après le signalement d'un tiers, il peut s'écouler un délai entre le moment où l'infraction est commise et celui où les services en ont connaissance et la constatent. Il peut également inclure un délai de contrôle qualité de l'enregistrement statistique de l'infraction au sein des services de sécurité, notamment en gendarmerie, qui est au plus de quelques jours en moyenne. Ce contrôle qualité est effectué principalement par le secrétariat judiciaire dans les services de police, par la brigade départementale de renseignements et d'investigations judiciaire pour les unités de gendarmerie.

5.3. Suivi des victimes et des mis en cause

A l'heure actuelle, les données administratives enregistrées par les forces de sécurité ne permettent pas de suivre de façon certaine les victimes et les mis en cause dans le temps, par exemple pour étudier la multi-victimation ou la réitération : l'identifiant des personnes est propre à la procédure, on ne peut donc pas repérer si une personne est victime dans plusieurs procédures distinctes. Concernant les données saisies par la police nationale, l'utilisation des données nominatives⁷ pourrait permettre de suivre les victimes et les mis en cause dans le temps. Cependant, un même événement peut faire l'objet de plusieurs procédures distinctes : par exemple, les affaires d'homicides sont souvent transférées d'un service de proximité vers un service spécialisé. Par ailleurs, certains événements font l'objet de saisines successives, entraînant à chaque fois l'ouverture d'une nouvelle procédure (initiative, instruction parquet, commission rogatoire pour investigations, pour écoute, etc.). Ainsi, certaines précautions sont à prendre pour distinguer la multi-victimation et la réitération des multi-procédures liées à un même événement.

Afin d'éviter les doubles-comptes, les bases statistiques sont constituées uniquement des primo-constatations et des élucidations (cf. supra).

⁷ Les données nominatives des infractions enregistrées par la gendarmerie ne sont à l'heure actuelle pas disponibles pour le SSMSI.

6. Contrôles de cohérence, redressements et apurements

Pour améliorer la qualité des données contenues dans les bases statistiques sur la délinquance enregistrée, le SSMSI procède à de nombreux contrôles de cohérence et est amené à corriger certaines données.

6.1. Communes

Les bases statistiques intègrent les nomenclatures officielles les plus récentes : le code officiel géographique (mis à jour annuellement), la taille de l'unité urbaine, le type de commune, la zone de compétence, etc.

Dans les données enregistrées par la police nationale, le département est incohérent avec la commune (de commission des faits, de résidence de la victime ou de résidence du mis en cause) dans quelques rares cas : 0,1 % pour le lieu de commission des faits et pour le lieu de résidence de la victime, 0,4 % pour le lieu de résidence du mis en cause. Dans le logiciel de saisie des plaintes, le policier renseigne indépendamment la commune et le département (de commission des faits ou de résidence). L'incohérence est systématiquement liée au fait qu'il existe deux communes portant le même nom (homonymes), appartenant à deux départements différents. Le SSMSI corrige le code commune en récupérant le code de la commune portant le même nom et se situant dans le département indiqué (Par exemple : Saint-Denis (93) et Saint-Denis (974)).

6.2. Dates incohérentes

Pour quelques observations, la date de début des faits est postérieure à la date d'enregistrement ou la date d'élucidation de l'infraction. Ces cas particuliers correspondent à des procédures ouvertes à l'avance (numéros réservés, en cas de déplacement sur le terrain et d'impossibilité d'utiliser les logiciels d'enregistrement des plaintes au moment des premières investigations). Le SSMSI redresse la date d'enregistrement et la date d'élucidation à la date de début des faits.

Lorsque la procédure contient plusieurs infractions d'une même nature, la date d'enregistrement est souvent renseignée pour une seule de ces infractions. Le SSMSI impute la date d'enregistrement des autres infractions en fonction de la date d'enregistrement de l'infraction principale.

En tout, 0,01% des infractions enregistrées par la gendarmerie nationale et 0,06% de celles enregistrées par la police nationale sont corrigées.

6.3. Homicides et tentatives d'homicides

Les données sur les homicides font l'objet d'un traitement spécifique. Il existe de nombreux doublons sur ce sous-champ, les procédures d'homicides étant souvent transférées d'un service à un autre (notamment, d'un service de sécurité publique à un service de police judiciaire chargé de mener des investigations approfondies lors de la poursuite de l'enquête). En théorie, les homicides ne doivent

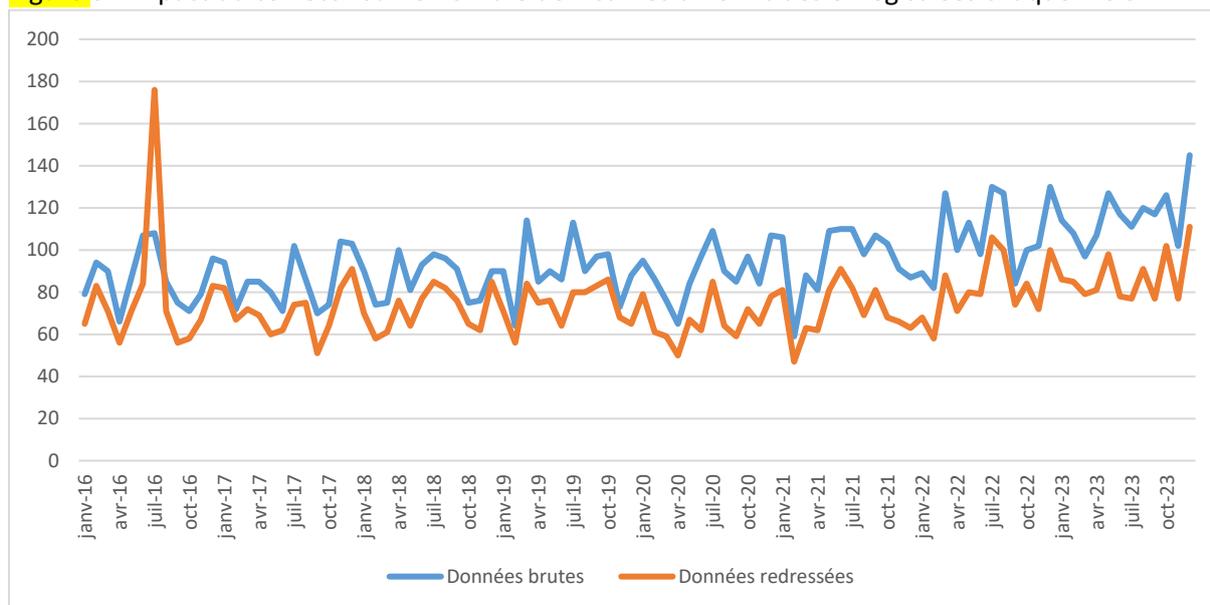
être comptabilisés que par le premier service, mais dans la pratique, certains homicides sont comptabilisés plusieurs fois : par le service primo-intervenant, puis par le nouveau service en charge. Les doublons d'homicides sont détectés via un processus sécurisé s'appuyant sur les noms et prénoms des victimes (actuellement disponibles uniquement dans les données de la police nationale, un chantier est en cours pour l'accès par le SSMSI aux données nominatives des victimes et des mis en cause pour les infractions enregistrées par la gendarmerie nationale), mais aussi sur le lieu et la date de commission des faits, la date de naissance et le sexe de la victime, le résumé des procédures et les informations sur les numéros de procédure d'origine – partiellement renseignés en cas de transfert de procédure).

Le SSMSI a entrepris depuis 2020 un chantier méthodologique de grande ampleur, visant à réaliser une expertise poussée de la qualité des données sur les homicides. Outre le problème des doublons, ces travaux ont permis de mettre en évidence un certain nombre d'anomalies, ayant un impact sur le nombre d'homicides comptabilisés, notamment :

- certaines tentatives d'homicides sont enregistrées en tant qu'homicides ;
- certains homicides ne sont pas enregistrés en tant qu'homicides ;
- dans certaines procédures comportant plusieurs victimes, des personnes blessées mais non tuées sont enregistrées comme victimes de l'homicide commis ;
- certaines années, des personnes morales sont enregistrées comme victimes d'homicide ;
- certaines victimes d'homicides sont animales et non humaines ;
- certaines procédures sont fictives (formation de nouveaux policiers, test de maintenance du logiciel d'enregistrement des procédures, ...) ;
- certains faits d'homicides sont absents des enregistrements, du fait d'erreurs de saisie ;
- les victimes d'attentats sont parfois manquantes.

Le processus de production des bases statistiques par le SSMSI corrige ces anomalies (voir Interstats Méthode n° 25) : une expertise qualitative exhaustive des enregistrements d'homicides est effectuée chaque mois, ainsi qu'un suivi quotidien de la presse permettant de détecter les homicides absents des enregistrements (figure 6). Les victimes d'homicides qui s'avèrent être vivantes et non décédées sont réintégrées aux données sur les tentatives d'homicides. Les homicides absents car mal enregistrés sont réintégrés aux données après correction par le service de police ou de gendarmerie ou par le SSMSI. L'ensemble de ces corrections est intégré dans les bases statistiques le mois suivant. Au total, le nombre d'homicides commis en France est révisé à la baisse d'environ 30 % chaque année.

Figure 6 : Impact du correctif sur le nombre de victimes d’homicides enregistrées chaque mois



Note : les données de décembre 2023 ne sont pas corrigées de l’expertise qualitative exhaustive.

Source : SSMSI – extraction alimentant la base statistique des victimes enregistrées par la police et la gendarmerie en 2023.

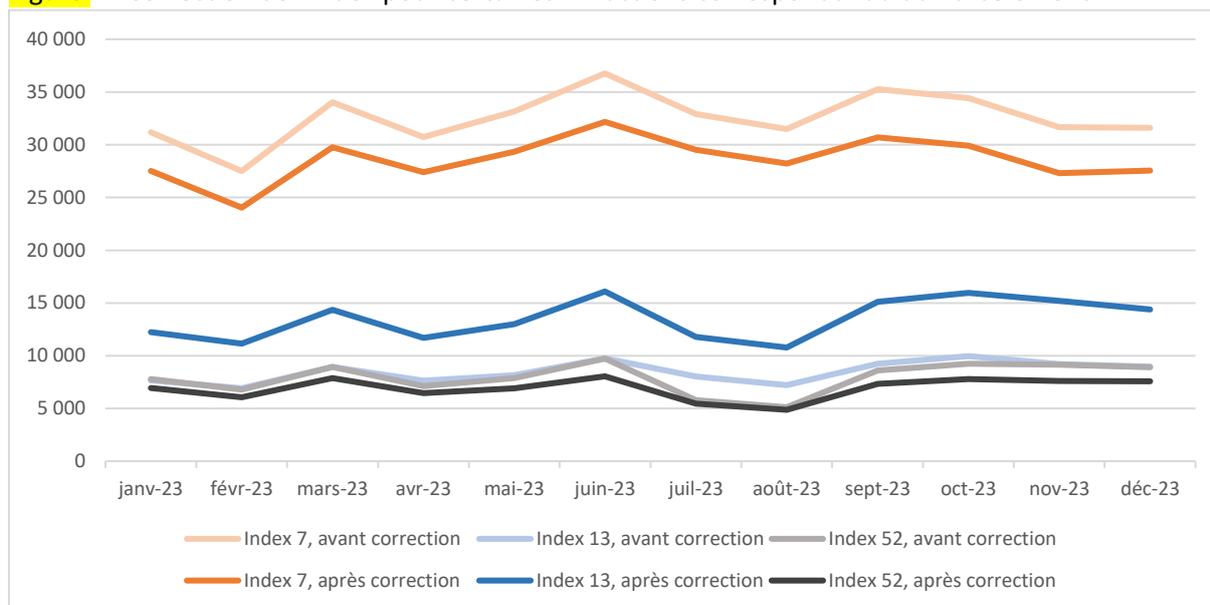
Les travaux se sont poursuivis en 2022, par l’expertise des données sur les tentatives d’homicides. Comme pour les homicides aboutis, les données sur les tentatives d’homicide nécessitent un traitement particulier car elles comportent de nombreux doublons. Le processus de production des bases statistiques par le SSMSI corrige ces doublons (voir Interstats Méthode n° 25). Au total, le nombre de tentatives d’homicides commises en France entre 2016 et 2022 est révisé à la baisse d’environ 6 % chaque année.

6.4. Mauvaises indexations

L’index est une catégorie de classification des infractions utilisée par les services de police et de gendarmerie nationales pour établir les statistiques de l’État 4001. Pour chaque nature d’infraction, la liste des index possibles (107 en tout) est déterminée dans le thésaurus des infractions, mis à jour par le SSMSI en collaboration avec la Direction des affaires criminelles et des grâces du ministère de la Justice, le Bureau des systèmes d’information opérationnels de la gendarmerie nationale et le Département des technologies appliquées à l’investigation de la police nationale. Par exemple, l’infraction « vol de carburant » doit nécessairement être associée à l’index 38 (vol d’accessoires sur véhicules à moteur immatriculés), alors que l’infraction « violence suivie de mutilation ou infirmité permanente » peut être associée à l’index 7 (autres coups et blessures volontaires criminels ou correctionnels) ou à l’index 73 (violences à dépositaires de l’autorité).

Or certaines infractions correspondant à du harcèlement sont mal classées par les services de police et de gendarmerie nationales, la consigne d’indexation ayant été mal spécifiée lors de la création de ces natures d’infraction : initialement classées en index 7 (autres coups et blessures volontaires criminels ou correctionnels), 52 (violences, mauvais traitements et abandons d’enfants), voire 73 (violences à dépositaires de l’autorité), 12 (menaces ou chantage dans un autre but) ou 48 (harcèlements sexuels et autres agressions contre des majeur(e)s), elles sont reclassées en index 13 (atteintes à la dignité et à la personnalité) dans les bases statistiques du SSMSI (figure 7). En tout, environ 60 000 infractions par an sont reclassées dans l’index 13 (dont 80% provenant de l’index 7).

Figure 7 : Correction de l'index pour certaines infractions correspondant à du harcèlement



Note : Index 7 : autres coups et blessures volontaires ; Index 13 : Atteintes à la dignité ; Index 52 : Violences, mauvais traitements et abandons d'enfants.

Source : SSMSI – base statistique des victimes enregistrées par la police et la gendarmerie en 2023.

6.5. Imputation de valeurs manquantes

Lorsqu'ils sont manquants dans les enregistrements de la police ou de la gendarmerie nationales, le sexe et l'âge de la victime sont imputés. La méthode d'imputation choisie est le Hot Deck par classe⁸. Elle est basée sur le concept de « donneur » : la donnée manquante est remplacée par la valeur observée pour une autre victime, choisie au hasard parmi des personnes qui lui ressemblent a priori. On construit des classes d'imputation selon un certain nombre de caractéristiques : on constitue des groupes de victimes homogènes, puis on recherche un donneur parmi celles appartenant au même groupe que la victime pour laquelle la donnée est manquante.

Pour définir les classes d'imputation, on procède de façon séquentielle : si on n'arrive à trouver aucun donneur ayant les mêmes caractéristiques que la victime pour laquelle l'information (sexe ou âge) est manquante, on élargit les classes d'imputation.

On impute :

- l'âge de la victime lorsqu'il est manquant (0,5 % des victimes personnes physiques)
- le sexe de la victime lorsqu'il est manquant (0,2 % des victimes personnes physiques)

De même, on impute :

- l'âge du mis en cause lorsqu'il est manquant (0,02 % des mis en cause personnes physiques)
- le sexe du mis en cause lorsqu'il est manquant (0,01 % des mis en cause personnes physiques)
- la nationalité du mis en cause lorsqu'elle est manquante (0,1 % des mis en cause personnes physiques)

⁸ Cf. « La correction de la non-réponse par repondération et par imputation », N. Caron, Documents de travail Insee n° M0502, 2005.

6.6. Délits routiers et contraventions enregistrés par la gendarmerie nationale

Concernant les délits routiers et les contraventions enregistrées par la gendarmerie nationale (infractions ne relevant pas de l'un des 107 index de l'état 4001), les données disponibles dans la base statistique Infractions sont reconstituées à partir de comptages agrégés par source des données (saisies journalières LRPGN, PVE, verbalisations par carnet à souches – timbre amende, quittance, consignation –, saisies dans la base nationale des statistiques des infractions – BNSI – à Saint-Pierre-et-Miquelon), service, commune de la personne concernée par la procédure (commune de résidence de la victime ou du mis en cause ou commune de commission des faits...), date d'attribution d'un numéro de procédure et nature d'infraction. Ces comptages incluent les infractions éligibles aux statistiques de l'état 4001 (infractions appartenant à l'un des 107 index de l'état 4001), pour lesquelles le SSMSI a accès aux données détaillées contenues dans le message d'information statistique. Cependant, il n'existe pas de marqueur dans les comptages agrégés permettant de repérer directement les infractions ayant fait l'objet d'un MIS car éligibles au champ du 4001, et le contour du champ du 4001 évolue dans le temps.

Le SSMSI a actuellement accès aux données détaillées concernant les infractions générant des messages d'information statistique (MIS), pas aux données détaillées des autres infractions qui ne sont pas centralisées dans une base informatique nationale. La gendarmerie nationale livre chaque mois au SSMSI deux listes : la liste des infractions actives éligibles aux statistiques de l'état 4001 et la liste des infractions abrogées qui étaient éligibles aux statistiques de l'état 4001. Le SSMSI considère ces infractions comme déjà présentes dans les données détaillées ayant fait l'objet d'un MIS (données 4001) et les supprime des données agrégées pour ne conserver que les données hors 4001, afin d'éviter une double comptabilisation.

Pour les natures d'infraction considérées comme hors champ de l'État 4001, le SSMSI génère autant de lignes que d'infractions enregistrées par chaque source des données x service x commune de la personne concernée par la procédure x date d'attribution d'un numéro de procédure x nature d'infraction. Faute d'autres informations, le SSMSI approxime la date des faits par la date d'attribution d'un numéro de procédure, et la commune de commission des faits par la commune de la personne concernée par la procédure. Sur le champ des données de l'État 4001, cette dernière correspond dans 50% des cas à la commune de commission des faits, dans 28% des cas à la commune de résidence de la victime ou du mis en cause, dans 15% des cas elle est non renseignée et dans 7%, elle ne correspond ni à la commune de commission des faits, ni à la commune de résidence de la victime ou du mis en cause.

La commune de commission des faits est par défaut égale à la commune de la personne concernée par la procédure (seule commune disponible dans les données agrégées). La date d'élucidation, la date d'unité de compte, la date d'ouverture et de clôture de la procédure, la date de début des faits et la date de fin des faits sont par défaut égales à la date d'attribution d'un numéro de procédure (seule date disponible pour ces données agrégées).

Un chantier est en cours concernant la transmission au SSMSI des données détaillées sur les infractions hors champ de l'Etat 4001 (contraventions, délits routiers) enregistrées par la gendarmerie nationale (caractéristiques détaillées des infractions, y compris sur les victimes et les mis en cause associés à l'infraction).

7. Avantages des bases statistiques par rapport aux données administratives d'origine

Les bases statistiques du SSMSI permettent d'effectuer une photographie à une date figée de la délinquance commise sur le Territoire de la république française (soit France + Collectivités d'Outre-mer) sur une période de référence bien définie. Elles regroupent des données provenant de sources différentes, qui sont homogénéisées pour faciliter les analyses tous services confondus (police + gendarmerie). Elles sont également harmonisées quel que soit le vecteur d'enregistrement (logiciels de rédaction de la police et de la gendarmerie, procès-verbaux électroniques, plateformes de dépôt de plainte en ligne, etc.). Contrairement aux statistiques de l'Etat 4001, les bases statistiques du SSMSI donnent une vision la plus exhaustive possible de la délinquance enregistrée, en intégrant à la fois les infractions principales et les infractions secondaires commises sur le Territoire de la république française (soit France + Collectivités d'Outre-mer), en ajoutant les délits routiers et les contraventions⁹ et en proposant des données détaillées plutôt que des agrégats. Les données des bases statistiques sur la délinquance enregistrée sont également requalifiées. Le SSMSI met en œuvre un certain nombre d'expertises et de redressements des données (mise en place d'une chaîne de contrôle, correction d'anomalies, nettoyage des données, détection et suppression de doublons, imputation de valeurs manquantes, intégration des nomenclatures à jour, documentation associée, etc.), les rendant plus robustes et de meilleure qualité. Ces redressements ne sont pas réalisés sur les comptages agrégés de l'Etat 4001. Le SSMSI définit et construit des indicateurs permettant de suivre certains champs infractionnels spécifiques (par exemple, les violences conjugales ou la cybercriminalité), non disponibles via la nomenclature des index de l'Etat 4001. Il s'assure de la cohérence des informations dans le temps et s'adapte aux nouveaux modes de collecte des plaintes pour maintenir l'exhaustivité du champ de la délinquance enregistrée.

Les bases statistiques du SSMSI permettent d'établir des résultats selon un maillage géographique fin. Elles permettent de suivre des infractions spécifiques ou rares (effet événementiel, homicides, etc.) et d'effectuer un suivi infra-annuel de la délinquance enregistrée par les forces de sécurité.

8. Perspectives

8.1. Expertises complémentaires

Le processus de production des bases statistiques du SSMSI a été entièrement refondu en 2022, notamment pour intégrer les infractions hors champ des statistiques de l'Etat 4001. D'autres améliorations sont prévues, comme par exemple l'imputation de certaines valeurs manquantes dans les données de la police et de la gendarmerie nationales pour des variables descriptives clés : la commune de commission des faits (non renseignée pour 1,1 % des infractions), la commune de résidence des victimes (non renseignée pour 0,6 % des victimes) ou la commune de résidence des mis en cause (non renseignée pour 3,0 % des mis en cause).

⁹ Dans le champ de la base statistique Infractions : relevés de la police et de la gendarmerie. Dans le champ des bases statistiques Victimes et Mis en cause : uniquement les relevés de la police.

Certaines expertises de données seront menées, par exemple sur l'existence ou non de doublons dans les enregistrements concernant d'autres champs infractionnels que les homicides ou sur les erreurs de qualifications réservées à des populations spécifiques (par exemple, les mineurs de 15 ans). La construction d'un processus de détection des valeurs aberrantes est également en projet pour améliorer la qualité des bases statistiques.

8.2. Appariements et enrichissements

Dans le cadre de futurs travaux statistiques, notamment relatifs à des appariements entre différentes bases de données, le SSMSI sollicitera l'Insee pour disposer du code statistique non signifiant (CSNS) pour les victimes et les mis en cause pour des infractions enregistrées par les services de police et de gendarmerie. Ce CSNS a été défini par la loi pour une République numérique de 2016 afin de faciliter les appariements de fichiers tout en garantissant un niveau de protection élevé des données à caractère personnel (voir l'article n° 4 du Courrier des statistiques N9). Il consiste en une clé d'appariement calculée à partir d'un chiffrage irréversible du NIR (numéro d'identification au répertoire des personnes physiques dénommé communément « numéro de sécurité sociale »).

Les appariements entre différentes bases de données permettront d'accroître les possibilités d'études sur la délinquance et l'insécurité en France : analyse des revenus ou de la profession des personnes victimes ou mises en cause via un appariement avec les sources Insee, analyse comparative des déclarations des victimes et des enregistrements par les forces de sécurité via appariement entre les sources administratives et l'enquête sur le Vécu et le ressenti en matière de sécurité (VRS), devenir des plaintes et analyse du parcours pénal des personnes via un appariement entre les données enregistrées par les forces de sécurité et celles enregistrées par le ministère de la Justice, etc.

8.3. Modalités d'accès pour les chercheurs

Enfin, les bases statistiques du SSMSI, lorsqu'elles ne contiennent aucune information directement identifiante, seront mises à disposition des chercheurs, sous réserve de l'accord du comité du secret statistique et selon des modalités d'accès sécurisées.

Bibliographie

Bénichou Y.-L., Espinasse L. et Gilles S., « Le code statistique non signifiant (CSNS) : un service pour faciliter les appariements de fichier », Courrier des statistiques N9, Insee, juin 2023.

Bernard M., « L'effet des requalifications des infractions sur les statistiques mensuelles et annuelles de la délinquance enregistrée », Interstats Méthode n° 23, SSMSI, novembre 2023.

Bernard M., « Elucidation des faits de délinquance par la police et la gendarmerie nationales : une nouvelle méthode de calcul », Interstats Méthode n° 18, SSMSI, décembre 2021.

Camus B., « Le défi de l'élaboration d'une nomenclature statistique des infractions », Courrier des statistiques N7, Insee, janvier 2022.

Chenu B., « 95% de la population vit à moins de 14 minutes d'un lieu d'accueil de la police ou de la gendarmerie nationales », Interstats Analyse n° 61, SSMSI, juin 2023.

Frattini F., « Les atteintes à l'environnement enregistrées par la police et la gendarmerie entre 2016 et 2021 », Interstats Analyse n° 46, SSMSI, mai 2022.

Juillard M., « Analyse de la délinquance enregistrée selon la profession des victimes : un difficile état des lieux », Interstats Méthode n° 20, juillet 2022.

Salembier L., « La mesure du nombre d'homicides et de tentatives d'homicide enregistrés par la police et la gendarmerie depuis 2016 », Interstats Méthode n° 25, SSMSI, janvier 2024.

SSMSI, « Insécurité et délinquance en 2022 : bilan statistique complet », Interstats Références, SSMSI, septembre 2023.

SSMSI, « Insécurité et délinquance en 2021 : bilan statistique », Interstats Références, SSMSI, septembre 2022.

SSMSI, « L'enregistrement des crimes et délits non routiers par la police et la gendarmerie », Interstats Méthode n° 2, SSMSI, octobre 2015.

Textes relatifs au traitement des données à caractère personnel

THESEE : Arrêté du 26 juin 2020 portant création d'un traitement automatisé de données à caractère personnel dénommé « traitement harmonisé des enquêtes et des signalements pour les e-escroqueries (THESEE).

MIS : Arrêté du 2 décembre 2010 autorisant la mise en œuvre d'un traitement automatisé de données à caractère personnel pour la gestion des messages d'information statistique et des bulletins d'analyse des accidents des unités élémentaires de la gendarmerie départementale et des gendarmeries spécialisées dénommé « Gestion des MIS et des BAA ».

LRPGN : Décret n° 2011-111 du 27 janvier 2011 autorisant la mise en œuvre par le ministère de l'intérieur (direction générale de la gendarmerie nationale) d'un traitement automatisé de données à caractère personnel d'aide à la rédaction des procédures (LRPGN).

LRPPN : Décret n° 2011-110 du 27 janvier 2011 autorisant la création d'un traitement automatisé de données à caractère personnel dénommé Logiciel de rédaction des procédures de la police nationale (LRPPN).

PVE : Arrêté du 18 décembre 2023 modifiant l'arrêté du 20 mai 2009 portant création d'un traitement automatisé dénommé « Application de gestion centrale ».