

Partie 2 : De la confection du RIL à l'élaboration des populations : des contrôles de la qualité tout au long du processus

La confection des estimations de population est un processus comportant de nombreuses étapes (préparation de la collecte, collecte, contrôles post-collecte, calculs, etc.). Pour assurer une bonne qualité des estimations de population, de nombreux contrôles sont effectués aux différentes phases du processus du recensement de la population. *Via* le GSDEM (Generic Statistical Data Editing Model), l'Unece classe les contrôles de données en trois catégories (Unece, 2019) : les contrôles de cohérence, les contrôles de vraisemblance et les contrôles des unités. Les contrôles de cohérence consistent à vérifier que les équations statistiques (égalité, infériorité, supériorité) entre variables sont bien vérifiées. Par exemple, un contrôle de cohérence est de vérifier que la population municipale est bien la somme entre la population des ménages, la population des communautés et la population des habitations mobiles, des sans-abris et des bateliers. Les contrôles de vraisemblance consistent à vérifier la plausibilité d'une donnée. S'agissant des contrôles de vraisemblance, le plus souvent, on vérifie que la donnée étudiée est vraisemblable par rapport à la même donnée sur une période passée ou par rapport à une source auxiliaire. Ainsi, dans le cadre de la préparation de la collecte du recensement de la population dans les communes de moins de 10 000 habitants, l'Insee réalise, par exemple, des contrôles en comparant le nombre de logements sur une voie donnée issue de la liste exhaustive des logements mise à jour par la commune à celui présent dans les données fiscales. Le contrôle des unités s'appuie sur le calcul de scores pour isoler les unités atypiques. Par exemple, ce type de contrôle peut consister à repérer des unités qui influencent beaucoup un agrégat (forte contribution à l'évolution par exemple), comme cela peut être le cas de certaines adresses atypiques qui sont recensées dans les communes de plus de 10 000 habitants et qui présentent un nombre moyen de personnes par logement très différent des autres adresses, et qui dès lors peuvent faire l'objet d'un traitement spécifique.

Ces trois types de contrôle sont effectués tout au long du processus d'estimation des populations légales. Cette partie passe en revue l'ensemble de ce processus, de la constitution de la base de sondage jusqu'à la validation des estimations produites, et précise les contrôles qui sont effectués à chaque phase.

A) La qualité de la base de sondage en grande commune : le répertoire des immeubles localisés

A-1. Définition du RIL

Le Répertoire d'immeubles localisés (RIL) n'existe qu'en grande commune² et contient la liste des adresses d'habitation des communes de 10 000 habitants ou plus ainsi que le nombre de logements associés à chaque adresse. Le RIL contient ainsi et localise tous les logements habitables des grandes communes, y compris les établissements touristiques et les communautés. Il est maintenu au plus près de la situation réelle, car il est utile pour trois étapes du recensement de la population des communes de 10 000 habitants ou plus :

- **avant la collecte**, il permet de constituer la base de sondage des adresses (BSA) dans laquelle sont tirés les échantillons d'adresses à recenser l'année suivante lors de l'enquête annuelle de recensement (EAR) ;

- **pendant la collecte**, le RIL va permettre de réaliser des plans de collecte qui seront remis aux agents recenseurs afin de faciliter le repérage des adresses à enquêter et de suivre l'avancement de la collecte;

- **après la collecte**, le RIL est un élément du calcul des populations légales des communes : il fournit le nombre de logements au 1^{er} janvier.

A-2. Mise à jour du RIL

Les RIL des communes de plus de 10 000 habitants ont été créés pour la première enquête de recensement par sondage en 2004. Ils ont été initialisés en partant des documents du recensement exhaustif de 1999³. Les plans communaux ont été numérisés et les adresses d'habitation ont ensuite été placées sur ceux-ci. La mise à jour de ce répertoire est réalisée depuis de manière collaborative par l'Insee et les communes ou les établissements publics de coopération intercommunale (EPCI). Au sein de chaque commune ou EPCI, un correspondant RIL (Corril) est nommé par arrêté municipal et est l'interlocuteur privilégié sur le RIL pour l'Insee.

La mise à jour du RIL est cruciale pour assurer son exhaustivité, afin de déterminer avec précision le champ des adresses à recenser lors de l'enquête annuelle de recensement et d'établir, de manière fiable, le nombre de logements habitables de la commune qui sera utilisé pour les estimations de population.

Depuis 2016, le RIL est mis à jour à partir de l'application Rorcal, outil en ligne développé par l'Insee et partagé entre l'Insee et les communes. Les outils cartographiques de Rorcal permettent de garantir une bonne qualité de localisation des adresses d'habitation. Les adresses du RIL sont ainsi calées sur le Référentiel à grande échelle (RGE) de l'IGN, une infrastructure nationale de données géographiques (parcelles cadastrales, limites administratives, etc.) qui fait référence. Le calage sur le RGE permet d'utiliser différents fonds de carte, des plans ou des vues satellites ou aériennes fournies par l'IGN, OpenStreetMap ou Google. L'interface Rorcal développée pour les communes facilite également leur gestion. Elles peuvent facilement signaler toutes modifications de caractéristiques d'une adresse (nombre de logements par exemple) ou effectuer des signalements de construction ou destruction de bâtiments. Lorsque la modification concerne plusieurs logements, l'Insee doit valider manuellement cette modification. La commune peut ainsi intervenir sur son RIL quand elle le souhaite, ce qui favorise une mise à jour en continu du RIL. Au-delà de

² Dans les DOM, les communes de moins de 10 000 habitants disposent également d'un RIL.

³ Lorsqu'une commune en croissance franchit le seuil des 10 000 habitants, un RIL est alors créé par l'Insee à l'image de la dernière collecte exhaustive. Ce RIL initial est ensuite mis à jour des permis intervenus entre la dernière collecte exhaustive et l'année courante. Le RIL passe alors en processus de mise à jour classique par l'Insee et la commune.

ces mises à jour spontanées de la part du Corril, un processus de mise à jour annuel du RIL a été mis en place par l'Insee. Celui-ci mobilise plusieurs sources (décrites ci-dessous) et aboutit à la confection d'un RIL millésimé et d'une base de sondage d'adresses (BSA) chaque année.

Les permis

Les permis (de construire, d'aménager, de démolir) sont fournis à l'Insee par le ministère chargé du logement à partir de la base Sitadel. Ils permettent à l'Insee de mettre à jour le RIL à la suite des constructions ou démolitions de logements, en créant ou en mettant à jour dans le RIL des habitations en construction. Ces permis sont livrés à l'Insee 4 fois par an (en août, novembre, février et avril).

Les nouvelles constructions repérées lors du traitement des permis, ainsi que certaines habitations dont on veut vérifier l'état d'habitabilité, sont envoyées aux Corril *via* l'application Rorcal pour traitement. Le rôle du Corril est alors de déclarer, *via* Rorcal, l'état de chaque habitation : « habitable », « périmée » (adresses ne contenant plus de logements habitables), « travaux autorisés », « travaux commencés », « non habitable travaux terminés ». Pour ce faire, le Corril peut se rendre sur le terrain ou se rapprocher du service d'urbanisme de la mairie.

Seules les adresses habitables sont utilisées pour le tirage de l'échantillon et le calcul des populations légales.

L'intégration de la collecte du recensement

Il arrive que, lors de l'enquête annuelle de recensement, l'agent recenseur s'aperçoive d'une incohérence entre les données du RIL et le terrain. Par exemple, le nombre de logements peut être différent. Cela peut s'expliquer par le décalage temporel entre la réalisation de la base de sondage d'adresses ayant permis le tirage de l'échantillon et l'enquête, mais aussi par des erreurs présentes dans le RIL. Il est alors pertinent de mettre à jour le RIL par les informations dont disposent les agents recenseurs. La commune, qui emploie les agents recenseurs, peut établir une « fiche navette » si elle observe une différence entre le terrain et le RIL, au niveau des adresses ou du nombre de logements. La commune peut aussi signaler une adresse censée être enquêtée mais qui n'est plus habitable. Dans ce cas, on change son statut dans le RIL pour qu'elle ne figure pas dans la prochaine base de sondage.

En complément des fiches navettes, les résultats de l'enquête annuelle de recensement, obtenus après les contrôles post-collecte de l'Insee, sont également utilisés pour mettre à jour le nombre de logements, le nombre d'étages et l'année de construction de chaque adresse recensée.

Les enquêtes cartographiques dans les DOM pour mettre à jour le RIL

Dans les DOM, la seule connaissance des permis de construire ne permet pas une bonne mise à jour du RIL. Des enquêtes terrain sont alors réalisées par l'Insee pour pallier ce défaut de couverture.

Aux Antilles, en Guyane et à partir de 2020 à Mayotte⁴, l'Insee effectue une enquête cartographique quasi-exhaustive sur les îlots⁵ de chaque commune de plus de 10 000 habitants et sur les communes de moins de 10 000 habitants⁶ concernées par l'enquête annuelle de recensement de l'année suivante. Cette enquête consiste à mettre à jour le RIL sur les territoires considérés. Seuls quelques îlots isolés ou très stables ne

4 La première EAR à Mayotte sera réalisée en 2021.

5 Dans les DOM, les grandes communes sont découpées en îlots pour les besoins du recensement. Ces îlots sont répartis en 5 groupes de rotation. Le tirage des adresses à recenser lors de l'EAR suivante s'effectue uniquement à l'intérieur des îlots du groupe de rotation de l'année. Ainsi, il est suffisant de réaliser l'enquête cartographique sur environ un cinquième des îlots des grandes communes pour préparer l'EAR à venir.

6 Dans les DOM, les communes de moins de 10 000 habitants disposent également d'un RIL.

font pas l'objet d'une enquête cartographique préalable au recensement. Ceux-ci sont alors enquêtés lors de la tournée de reconnaissance de la collecte (cf. partie 2-B) (enquête dite « 2 en 1 »).

Dans les communes de plus de 10 000 habitants, le RIL est donc mis à jour chaque année sur un cinquième du territoire de la commune, ce qui fait, qu'à un instant donné, il est plus ou moins à jour selon les îlots.

À La Réunion, l'Insee effectue une enquête cartographique dite « en différentiel ». Contrairement aux enquêtes cartographiques des Antilles-Guyane, cette enquête n'est pas exhaustive, car les données administratives y sont plus fiables que dans les autres DOM. Les enquêteurs examinent des habitations en construction (connues *via* des permis de construire) et leurs alentours pour voir si elles sont devenues habitables et si d'autres constructions n'ont pas vu le jour à proximité. Ils examinent également des habitations présentes dans les données fiscales, mais absentes du RIL.

L'expertise des communes

L'expertise légale des communes ou EPCI est une opération préliminaire aux enquêtes de recensement prévue dans l'article 27 du décret n° 2003-485. Elle permet de faire une ultime vérification de la qualité du RIL avant de produire la version du RIL (RIL millésimé) qui servira aux tirages d'échantillon et aux calculs de populations légales et de résultats statistiques. Cette vérification est réalisée par la commune. À l'issue de l'expertise, la commune signale les éventuels ajouts ou suppressions à faire et confirme que son RIL contient et localise l'ensemble des logements habitables sur son territoire. Le calendrier de l'expertise est fixé par arrêté : elle commence le 25 mai en métropole. Des listes d'adresses à vérifier peuvent être envoyées dans Rorcal par l'Insee à la commune pour concentrer son expertise sur les cas repérés comme fragiles. Les communes ont jusqu'au 30 juin pour effectuer des modifications. Dans les DOM, cette expertise se déroule entre fin août et fin septembre.

La livraison des RIL

Après l'expertise, les équipes recensement dans les établissements régionaux de l'Insee fournissent les RIL au responsable du tirage de l'échantillon, opération dite de « livraison des RIL ». Avant cela, elles doivent justifier les fortes évolutions du nombre de logements habitables des Iris (cf. encadré 2-A.1) (plus de + 5 % ou moins de - 5 %) et des communes (plus de + 3 % ou moins de - 1 %).

Après la fin de l'expertise légale des communes, les équipes recensement réalisent la première livraison des RIL de mi-juin à mi-juillet. Les RIL livrés sont contrôlés par la direction générale de l'Insee. On cherche principalement à éviter des doublons, des incohérences dans les données ou un manque de précision dans l'adressage, qui compliquerait la collecte du recensement. Ces contrôles peuvent conduire à de nouvelles livraisons des RIL. Quand les RIL livrés sont validés, l'Insee constitue les bases de sondages d'adresses et procède au tirage de l'échantillon des adresses à recenser pour l'enquête annuelle de recensement à venir : en août pour les communes de métropole et en octobre pour les communes des DOM.

Encadré 2-A.1 : les mailles géographiques infracommunales mobilisées dans le recensement : les IRIS et les îlots

Définition des IRIS

L'IRIS (Îlots Regroupés pour l'Information Statistique) constitue la brique de base en matière de diffusion de données infra-communales. Il doit respecter des critères géographiques et démographiques et avoir des contours identifiables sans ambiguïté et stables dans le temps. Les communes d'au moins 10 000 habitants et une forte proportion des communes de 5 000 à 10 000 habitants sont découpées en IRIS. Ce découpage constitue une partition de leur territoire.

Définition des îlots

Jusqu'au recensement de la population 1999, toutes les communes étaient découpées en îlots pour des raisons d'organisation du recensement. Cette partition des communes en îlots facilitait l'organisation et la collecte du recensement sur le terrain. L'îlot est l'unité géographique de base :

- en zone bâtie dense, il représente le plus souvent un pâté de maisons (ou petit groupe de pâtés de maisons), éventuellement scindé en cas de limite communale ou cantonale traversant le pâté de maison;
- en zone " périphérique ", l'îlot est un ensemble limité par des voies (ou autres limites visibles) découpant cette zone en plusieurs morceaux. Les îlots peuvent être vides d'habitants (par exemple une gare). Les îlots ont été définis par l'Insee en concertation avec les communes.

Les résultats du recensement de 1999 ont été diffusés à l'îlot. Or, même exhaustif, le recensement n'était pas « parfait » et pouvait souffrir de certaines incertitudes tant sur le dénombrement de la population que sur ses caractéristiques.

Utilisation de ces mailles territoriales dans le recensement rénové

Les IRIS interviennent à trois niveaux dans le recensement rénové de la population :

- dans les communes de plus de 10 000 habitants, le tirage de l'échantillon de l'EAR une année donnée fait intervenir le nombre de logements par IRIS.
- les estimations de populations dans les communes de plus de 10 000 habitants sont très majoritairement réalisées au niveau IRIS avant agrégation sur l'ensemble de la commune.
- l'IRIS est la maille infracommunale de diffusion du recensement rénové sur l'ensemble des communes de plus de 10 000 habitants et la majorité des communes de plus de 5 000 habitants. Il est plus pertinent que l'îlot dans les communes de plus de 10 000 habitants du fait du sondage.

La France compte 16 349 IRIS. Dans les communes de plus de 10 000 habitants de métropole, un IRIS contient en moyenne 1 360 logements.

Les îlots ne sont plus utilisés que pour les communes des DOM. La méthode par sondage appliquée aux communes de plus de 10 000 habitants s'appuie en métropole sur une base de sondage de qualité maintenue grâce à la mise à jour continue du RIL. Dans les communes de plus de 10 000 habitants des DOM, pour que le RIL soit de bonne qualité, on procède à une enquête cartographique qui consiste avant chaque collecte du recensement à ratisser une zone dans son intégralité et à repérer toutes les habitations. Cette opération étant lourde, elle n'est réalisée chaque année par roulement que sur 1/5 du territoire de la commune, défini à partir d'un regroupement d'îlots (qui forment ainsi un des cinq "groupes de rotation" de la commune). C'est au sein des adresses repérées lors de cette enquête cartographique qu'est tiré l'échantillon de l'EAR. Pour que cet échantillon soit représentatif de la commune, il est nécessaire que chaque groupe de rotation soit composé de beaucoup de petites zones réparties sur toute la commune pour avoir le maillage le plus fin possible, d'où l'idée de prendre le niveau îlot déjà existant comme brique élémentaire. Dans les communes de moins de 10 000 habitants des DOM, l'îlot est la brique de base pour la constitution des districts de collecte.

Les DOM (y compris Mayotte) comptent au total 17 996 îlots avec au moins un logement habitable. Dans les communes de plus de 10 000 habitants des DOM, un îlot contient en moyenne 59 logements.

Les enquêtes terrain en métropole

L'Insee mène chaque année des enquêtes terrain pour vérifier la qualité du RIL sur une zone prédéterminée. Ces enquêtes ont lieu en général entre août et décembre. Deux enquêtes ayant des finalités différentes co-existent :

- **les enquêtes de mesure de la qualité du RIL (EMQR)**, qui servent à mesurer la qualité du RIL en métropole et à La Réunion. Pour ce faire, plusieurs Iris de communes de plus de 10 000 habitants sont tirés de manière aléatoire de façon à assurer une représentativité du territoire national et font l'objet d'un ratissage complet afin de repérer l'ensemble des écarts entre le RIL et le constat du terrain : différences de nombres de logements à une adresse donnée, nouvelles adresses, adresses détruites, etc. Ces enquêtes permettent en particulier d'évaluer au niveau national l'excédent ou le déficit du RIL en nombres de logements. Les résultats de ces enquêtes sont présentés dans la partie 2-A-4. Elles permettent également de construire un modèle pour estimer quels Iris ont une probabilité élevée d'être de mauvaise qualité. Des corrections sont ensuite apportées dans les RIL des communes pour corriger les erreurs repérées.

- **les opérations de mise à niveau de la qualité du RIL**, qui ont pour objectif d'améliorer la qualité du RIL en ciblant sur des zones infra-communales où le RIL est susceptible d'être de mauvaise qualité. Elles peuvent être réalisées sur un Iris ou une zone ciblée. Les établissements régionaux de l'Insee disposent d'une liste d'Iris qu'on suppose de mauvaise qualité. Ils peuvent aussi choisir des zones à partir de leur connaissance des RIL. Au-delà de l'amélioration de la qualité sur les zones enquêtées, ces enquêtes permettent d'alerter les communes sur un défaut de mise à jour du RIL afin qu'elles soient plus impliquées.

A-3. Indicateurs de qualité du RIL

Un cadre d'assurance qualité du RIL a été défini en 2016 par l'Insee, pour vérifier que le RIL remplit certains objectifs, notamment en matière d'exactitude, de fiabilité, de cohérence et de comparabilité.

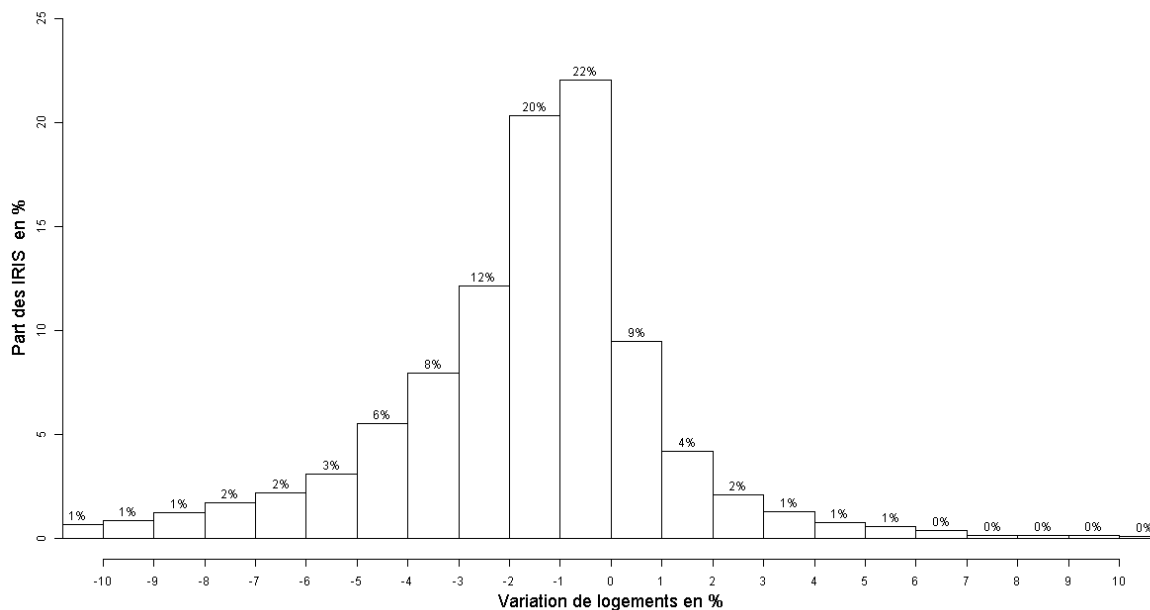
Ce cadre d'assurance qualité du RIL préconise de suivre plusieurs indicateurs qualité.

L'un de ces indicateurs compare le nombre de logements des adresses échantillonnées dans le recensement, avant et après collecte. On peut ainsi repérer les erreurs dans le RIL mises en évidence par la collecte du recensement. Il faut bien avoir à l'esprit que ces écarts ne peuvent porter que sur des adresses enquêtées lors de l'EAR et qui sont par conséquent présentes dans le RIL avec au moins un logement habitable. Lors de la collecte, on peut s'apercevoir d'une différence sur le nombre de logements (excédent ou déficit du RIL) ou d'une erreur dans le RIL sur l'habitabilité d'un logement (logement habitable à tort dans le RIL – par exemple si le logement a été détruit – qui représente un excédent du RIL). En revanche, la collecte ne permet pas de repérer des adresses d'habitation qui ne sont pas dans le RIL, car par définition elles ne peuvent pas être tirées, et donc pas enquêtées par un agent recenseur. Cet indicateur ne permet donc pas d'en tirer des conclusions sur un éventuel excédent ou déficit du RIL. Il met cependant en lumière certains défauts de qualité du RIL.

En moyenne sur 5 ans (2015-2019) et en solde, c'est-à-dire en faisant la différence entre les logements ajoutés et retirés à la suite de la collecte, l'EAR contient 1,46 % de logements de moins que l'échantillon extrait du RIL. Ce chiffre ne doit pas s'interpréter comme un excédent de 1,46 % du RIL puisque les logements habitables des adresses absentes du RIL ne sont par définition pas comptabilisés dans ce pourcentage.

Dans le détail, on constate que la différence n'est pas très élevée dans la plupart des Iris, même si elle est supérieure à 3 % pour 18 % des Iris, probablement ceux dans lesquels le travail d'expertise a été moins bien mené (cf. graphique 2-A.1).

Graphique 2-A.1 : Répartition des Iris selon la variation du nombre de logements de l'échantillon à la suite de la collecte

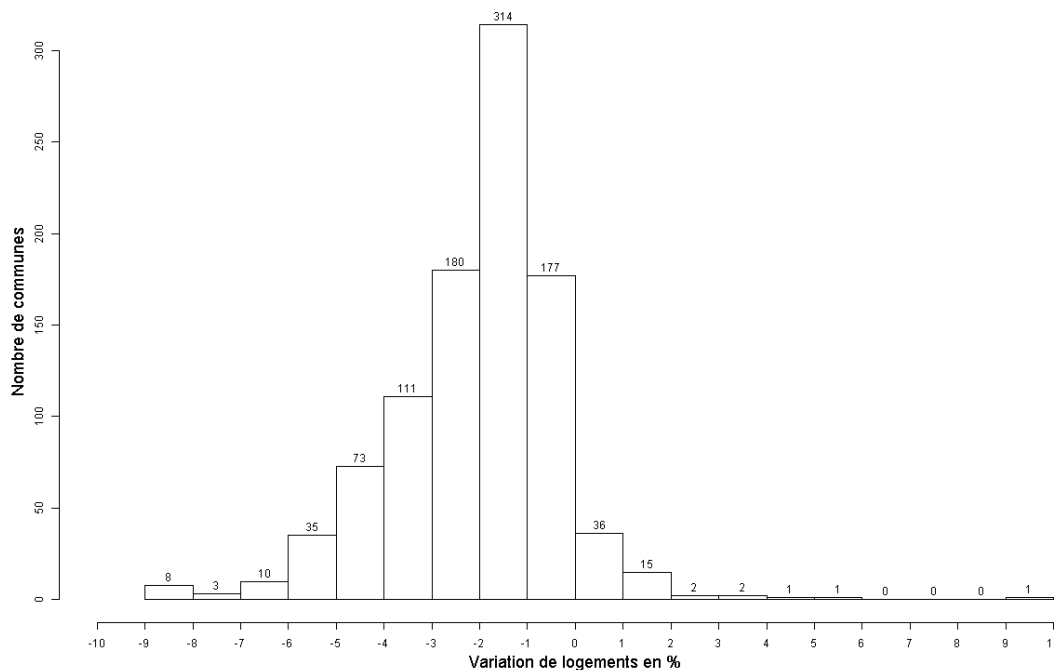


Lecture : pour 22 % des Iris, le nombre de logements de l'échantillon baisse de moins de 1 % à la suite de la collecte.

Note : la part des Iris est calculée en tenant compte du poids de chaque Iris en nombres de logements.

Source : enquêtes annuelles de recensement 2015 à 2019.

Graphique 2-A.2 : Répartition des communes selon la variation du nombre de logements de l'échantillon à la suite de la collecte



Lecture : 180 communes affichent une baisse du nombre de logements de l'échantillon comprise entre 2 % et 3 % à la suite de la collecte. Dans 15 communes, la hausse du nombre de logements de l'échantillon est comprise entre 1 % et 2 % à la suite de la collecte.

Note : en étiquette figure le nombre de communes.

Source : enquêtes annuelles de recensement 2015 à 2019.

Un autre indicateur du cadre d'assurance qualité porte sur l'évolution du nombre de logements dans le RIL. Il se décline en 2 parties.

On compare d'abord l'évolution du nombre de logements sur 5 ans dans le RIL et dans la taxe d'habitation. Une divergence peut mettre en avant un défaut de mise à jour dans le RIL ou un effet de gestion des fichiers de la taxe d'habitation. En pratique, l'écart entre l'évolution annuelle moyenne du RIL sur 5 ans et celle de la TH ne dépasse 1 % de la population que pour 23 grandes communes, avec un maximum de 2,2 %.

On compare également l'évolution du RIL sur l'année passée avec celle des 5 ans précédents. On cherche donc ici à identifier des ruptures de tendance. Ces dernières n'indiquent pas forcément une mauvaise qualité du RIL, car une évolution sur un an qui diffère de la tendance sur 5 ans peut être justifiée, par exemple quand des programmes de construction arrivent à terme, mais cet indicateur permet d'être vigilant sur certaines communes.

Les autres indicateurs sont liés à la gestion du RIL, par exemple le taux de communes actives dans l'application Rorcal, c'est-à-dire celles qui ont effectué au moins une mise à jour au cours de la campagne. Il était de 99,5 % pour la campagne 2018-2019 en métropole. Le niveau élevé de cet indicateur montre la bonne appropriation de l'outil par les communes, mais ne permet pas d'évaluer la qualité du travail effectué.

A-4. Principaux résultats de l'enquête de mesure de la qualité du RIL (EMQR)

L'EMQR 2017

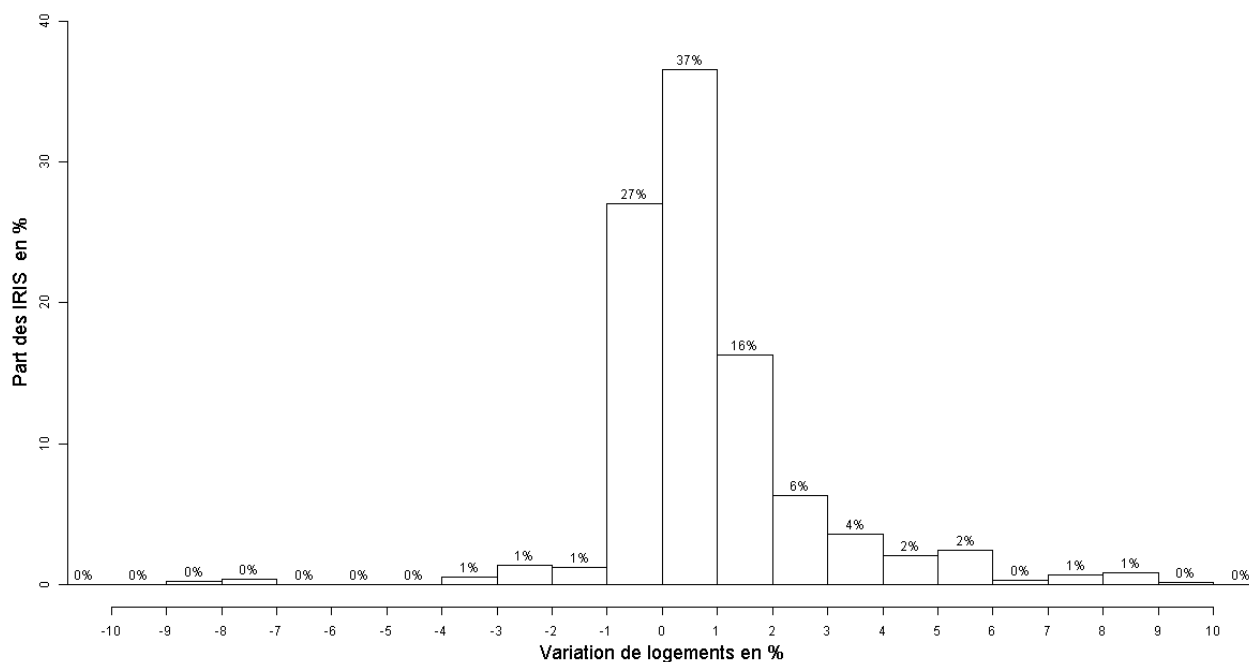
Les EMQR servent à mesurer la qualité du RIL en métropole et à La Réunion. Dans cette partie, nous nous restreignons aux résultats relatifs à la métropole, territoire sur lequel la dernière EMQR a eu lieu en 2017. Il s'agit d'effectuer un ratissage complet d'un échantillon d'Iris. Des enquêteurs de l'Insee déterminent le nombre de logements habitables de l'Iris. On estime ainsi l'écart entre le nombre de logements habitables dans le RIL et sur le terrain. Cette opération s'effectue entre les mois d'août et de décembre, soit entre 1 et 6 mois après la validation du RIL par la commune : **un écart observé entre le terrain et le RIL peut donc être la conséquence d'une modification réelle sur le terrain et être attribué à tort comme un défaut du RIL.**

Cette enquête permet de repérer des logements comptés à tort dans le RIL qui, sur le terrain, n'existent pas ou sont inhabitables et qui amènent notamment à surestimer la population, ou un déficit de logements qui amène à sous-estimer la population (cf. partie 2-F). En pratique, le déficit est plus important que l'excédent. En 2017, on avait :

- > 1,8 % de logements manquants dans le RIL ;
- > 0,9 % de logements comptés à tort dans le RIL ;
- > soit en solde un déficit de logements de 0,9 %.

La répartition pondérée par le nombre de logements du solde montre qu'on a peu d'Iris avec un déficit ou un excédent élevé (cf. ci-dessous) : 10 % d'Iris avec un déficit supérieur à 3 % et 4 % d'Iris avec un excédent supérieur à 1 %. Dans les illustrations ci-dessous, on représente la variation du nombre de logements à la suite de l'EMQR, c'est-à-dire qu'une augmentation du nombre de logements met en évidence un déficit dans le RIL.

Graphique 2-A.3 : Répartition des Iris selon la variation du nombre de logements à la suite de l'EMQR



Lecture : pour 37 % des Iris, le nombre de logements augmente de moins de 1 % à la suite de l'EMQR.

Note : la part des Iris est calculée en tenant compte du poids de chaque Iris en nombres de logements, et du taux de sondage dans chaque strate de tirage de l'EMQR.

Source : EMQR 2017.

On peut séparer le déficit de logements dans le RIL en 3 types :

1) Il manque des logements sur une adresse habitable dans le RIL. C'est le cas le moins gênant, puisqu'on a une chance que l'adresse soit tirée dans l'échantillon de l'EAR et donc que le nombre de logements soit corrigé au moment de la collecte⁷. Sur 5 ans, on enquête 40 % des logements connus dans le RIL. Sur les 1,84 % de logements manquants dans le RIL, d'après l'EMQR, 1,03 % le sont sur une adresse habitable dans le RIL.

2) Les logements manquants correspondent à une adresse en construction dans le RIL. Tant qu'elle est en construction, une adresse ne peut pas être tirée dans l'échantillon du recensement, donc ces cas ne peuvent pas être corrigés lors de la collecte. Toutefois, ces adresses sont soumises pour expertise au moins une fois par an aux communes (cf. plus haut), donc elles peuvent être corrigées si la commune les déclare habitables. On a 0,22 % de logements manquants dans le RIL associés à ces cas. Ces cas peuvent par ailleurs s'expliquer en partie par un décalage entre la date de validation des RIL et la date de réalisation de l'enquête de mesure de la qualité.

3) Les logements manquants correspondent à une adresse manquante dans le RIL. C'est le cas le plus gênant, puisqu'on a moins de chances de la corriger que dans les 2 premiers cas. Deux sources peuvent alors mettre à jour le RIL : un signalement spontané de la commune ou une opération de mise à jour du RIL sur le terrain. 0,59 % des logements manquants sont sur des adresses manquantes dans le RIL.

⁷ Cependant, cette correction ne concerne que les adresses recensées. Les estimations de population se basant sur l'ensemble des logements du RIL, des logements manquants sur une adresse habitable dans le RIL mais non recensée ne sont pas redressés. Le redressement n'est donc que partiel.

Tableau 2-A.1 : Comparaison de la mesure de la variation de logements du RIL à la suite de l'EMQR 2017 et des EAR

		EAR	EMQR
Variation du nombre de logements sur des adresses existantes dans le RIL (avec un nombre de logements strictement positif avant et après l'enquête)	Augmentation du nombre de logements sur des adresses déjà habitables	2,27 %	1,03 %
	Diminution du nombre de logements (hors adresses supprimées)	-2,16 %	-0,39 %
	Variation du nombre de logements (hors adresses supprimées)	0,11 %	0,64 %
Variation du nombre de logements sur des adresses qui apparaissent ou sont rendues habitables à la suite de l'enquête	Variation du nombre de logements sur des adresses qui apparaissent à la suite de l'enquête	0 %	0,59 %
	Variation du nombre de logements sur des adresses qui sont rendues habitables à la suite de l'enquête		0,22 %
Variation du nombre de logements sur des adresses qui disparaissent à la suite de l'enquête		-1,58 %	-0,54 %
Variation totale du nombre de logements		-1,46 %	0,90 %

*Champ : les pourcentages expriment une variation sur le champ de l'enquête, c'est-à-dire les logements dans les échantillons des 5 EAR prises en compte ou l'ensemble des logements habitables de l'Iris enquêté pour l'EMQR.
Sources : enquêtes annuelles de recensement 2015 à 2019, EMQR 2017.*

La variation du nombre de logements à la suite de la collecte des EAR et celle à la suite de l'EMQR ne sont pas directement comparables du fait du champ sur lequel porte chaque enquête (échantillon de l'EAR versus ensemble d'Iris). Il est possible toutefois de décomposer ces variations de nombres de logements pour mieux comparer les résultats de ces deux types d'enquêtes (cf. tableau 2-A.1). Par nature, pour les EAR, la variation du nombre de logements sur des adresses qui apparaissent ou sont rendues habitables à la suite de l'enquête est nulle. En effet, les adresses manquantes ou non habitables ne figurent pas dans la base de sondage issue du RIL. Il est donc normal que les résultats sur ce champ et donc la variation totale ne soient pas les mêmes entre les EAR et l'EMQR. Concernant les variations du nombre de logements sur des adresses habitables, hors suppressions, on peut penser que les EAR sont plus fiables que l'EMQR. En effet, les agents recenseurs accèdent obligatoirement aux immeubles qu'ils enquêtent, alors que, pendant l'EMQR, les enquêteurs de l'Insee n'ont pas forcément accès aux immeubles. Les résultats diffèrent sensiblement entre les deux enquêtes : + 0,11 % pour l'EAR *versus* + 0,64 % pour l'EMQR, avec des écarts encore plus marqués lorsque l'on sépare l'augmentation du nombre de logements et la diminution du nombre de logements (cf. les deux premières lignes du tableau). En revanche, on pourrait s'attendre à ce que les nombres de logements liés à des adresses qui disparaissent soient plus proches : - 1,58 % pour l'EAR *versus* - 0,54 % pour l'EMQR. Il y a peut-être moins d'adresses qui disparaissent à la suite de l'EMQR parce qu'elles peuvent déjà avoir été supprimées lors d'une enquête de recensement précédente. En effet, l'EAR porte sur des adresses qui n'ont pas été enquêtées depuis au moins 5 ans, puisque les adresses sont tirées dans un groupe de rotation différent chacune des 5 années d'un cycle (cf. parties 1 et 2-F). Au contraire, lors de l'EMQR, on enquête les adresses de tous les groupes de rotation, donc celles des groupes de rotation enquêtées dans les années les plus récentes ont plus de chances d'avoir été retirées *a priori*.

Corrélations entre les résultats de l'EMQR et les caractéristiques des Iris enquêtés

Le déficit et l'excédent dans le RIL, mis en évidence par l'EMQR, peuvent être corrélés à des caractéristiques de l'Iris (cf. tableau 2-A.2). On a notamment une corrélation positive avec le taux d'adresses non normalisées, c'est-à-dire les adresses qui ne donnent pas précisément un numéro et un nom de voie (exemple : lotissement XXX). Ces adresses sont plus fréquentes dans le sud de la France. En effet, les adresses non normalisées sont plus difficiles à repérer et à distinguer entre elles. On a donc plus de risques de mettre à jour la mauvaise adresse ou de ne pas repérer et donc ne pas inclure certaines adresses dans le RIL. On a aussi une corrélation positive avec le taux d'adresses non enquêtées lors du recensement (FANE), qui correspondent à des adresses qu'on considérerait comme habitables dans le RIL mais qui ne l'étaient pas lors de la collecte du recensement. Par construction, cet indicateur est corrélé à un excédent dans le RIL, mais il est également corrélé avec le déficit. L'excédent et le déficit sont d'ailleurs très corrélés entre eux, c'est-à-dire qu'un Iris de mauvaise qualité au niveau de l'excédent le sera aussi souvent pour le déficit. C'est en général le signe d'une expertise moins fine menée dans les communes concernées. Lorsque les communes réalisent une expertise soignée du RIL, il y a à la fois peu de déficit mais aussi peu d'excédent.

En revanche, on trouve peu de corrélations avec le solde. En effet, certains Iris avec un RIL de mauvaise qualité ont un déficit et un excédent élevés, mais qui s'annulent. Un solde proche de zéro peut donc correspondre à un RIL de bonne ou de mauvaise qualité, ce qui fausse l'étude des corrélations.

Tableau 2-A.2 : Corrélation de Spearman entre les résultats de l'EMQR et les caractéristiques des Iris

	Déficit		Excédent		Solde	
	Coefficient de corrélation	P-value	Coefficient de corrélation	P-value	Coefficient de corrélation	P-value
Taux d'adresses non normalisées	0,16	0,01	0,16	0,00	-0,09	0,09
Surface de l'Iris	0,11	0,05	0,09	0,10	-0,07	0,19
Taux de FANE	0,26	0,00	0,40	0,00	-0,06	0,32
Variabilité du nombre de logements	0,18	0,00	0,09	0,11	-0,16	0,00

Lecture : la p-value du test de corrélation entre la surface et le déficit est de 0,05 et le coefficient de corrélation est de 0,11. Les coefficients et les p-values sont en vert pour les variables dont la corrélation est significative et en rouge pour celles dont la corrélation ne l'est pas.

Note : les corrélations de Spearman mettent en évidence une relation monotone entre 2 variables, sans que cette relation soit obligatoirement linéaire, comme c'est le cas pour les corrélations de Pearson. On a une relation monotone entre la variable A et la variable B si quand A croît, B croît ; ou quand A croît, B décroît. Par exemple, on a une corrélation significative (0,16) entre le déficit de logements dans le RIL et le taux d'adresses non normalisées.

Sources : EMQR 2017, enquêtes annuelles de recensement 2014 à 2018.

Comparaison des EMQR dans le temps

Entre 2003 et 2017, à chaque EMQR réalisée, le déficit brut et l'excédent brut dans le RIL sont toujours compris entre 0,9 % et 2,6 % (cf. tableau 2-A.3). Le solde est toujours compris entre - 1 % et 0,7 %.

Le plan de sondage de l'EMQR a évolué. De 2003 à 2007, il était à 2 degrés : on tirait des communes et ensuite 8 % des îlots de ces communes (cf. encadré 2-A.1). À partir de 2010, on utilise un sondage par strate et on tire des Iris, indépendamment de leur commune d'appartenance. Actuellement, il n'existe plus de contours d'îlots en métropole, l'Iris est donc le plus petit zonage infra-communal disponible. *A priori*, il était

plus intéressant de tirer des îlots, puisqu'on pouvait enquêter plus de zones. Par contre, on peut plus facilement extrapoler les résultats sur des Iris puisqu'on dispose de plus de données issues du recensement sur ce zonage.

Tableau 2-A.3 : Résultats et taille d'échantillon des EMQR 2003 à 2017

Année	Déficit de logements dans le RIL	Excédent de logements dans le RIL	Solde	Nombre de zones enquêtées	Nombre d'adresses enquêtées
2003	2,3 %	2,1 %	-0,2 %	8 % des îlots de 225 communes	138 500
2004	1,9 %	2,3 %	0,4 %	16 % des îlots de 166 communes	140 000
2005	1,1 %	1,8 %	0,7 %	16 % des îlots de 149 communes	143 200
2006	0,9 %	1,4 %	0,5 %	ND*	ND*
2007	1 %	1 %	0 %	ND*	340 000
2010-2011	2,6 %	1,6 %	-1 %	716 Iris	353 000
2013	ND*	ND*	-0,6 %	291 Iris	127 960
2017	1,8 %	0,9 %	-0,9 %	330 Iris	143 000

*ND : non disponible.

Source : EMQR 2003 à 2017.

B) L'exhaustivité de la collecte

Disposer d'abord d'une liste exhaustive et à jour de logements

Dans les communes de moins de 10 000 habitants de métropole où la collecte est exhaustive tous les cinq ans, la qualité du dénombrement de la population sur le terrain dépend en premier lieu de la qualité du repérage des logements. Il convient d'abord de s'assurer que les agents recenseurs disposent d'une liste exhaustive et à jour des logements, habités ou non habités de la commune. Pour ce faire, l'Insee transmet aux responsables communaux, 6 mois avant l'enquête (au mois de juin), la liste des adresses recensées cinq ans plus tôt. Les communes ont alors pour mission d'expertiser et de mettre à jour cette liste. À chaque adresse doit être associé un nombre de logements. Il s'agit ainsi d'ajouter les nouvelles constructions et d'ôter celles qui ont été détruites. Toutes les constructions habitables doivent être répertoriées, qu'elles soient habitées ou non. En effet, un logement vide en septembre pourrait devenir habité en janvier et l'agent recenseur doit se rendre dans tous les logements afin de vérifier leur catégorie (résidence principale, résidence secondaire, logement vacant, logement occasionnel⁸). Parallèlement à cette expertise menée en commune, l'Insee exerce sa mission de contrôle en comparant le nombre de logements par commune, voie ou parfois adresse avec des informations d'origine fiscale. En cas d'écarts, un dialogue est engagé avec la commune pour les expliquer ou les corriger.

Dans les communes de plus de 10 000 habitants de métropole, la question de l'exhaustivité et de la bonne mise à jour de la liste des adresses de la commune relève des travaux préalables sur le RIL (cf. partie 2-A). Lors de la livraison de l'échantillon de l'année, le responsable en commune est chargé de vérifier si certaines adresses pourraient poser des difficultés particulières et si le nombre de logements est cohérent. L'Insee réalise également des opérations de contrôles par comparaison avec des données fiscales.

Dans les départements d'outre-mer, la liste exhaustive de logements est également issue du RIL, dont la mise à jour passe notamment par la réalisation d'enquêtes cartographiques préalables (cf. partie 2-A). À partir de cette liste, les travaux de vérification réalisés par les communes en amont du recensement sont les mêmes qu'en métropole.

Puis, bien identifier la zone de travail de chaque agent recenseur ou l'échantillon d'adresses

Une fois la liste des adresses et leur nombre de logements stabilisés par l'expertise de la commune et de l'Insee, une dernière phase de vérification est menée par l'agent recenseur dans la semaine qui précède l'interrogation des habitants. Celui-ci, muni d'une liste pré-établie, réalise une « tournée de reconnaissance ». Cette tournée de reconnaissance est essentielle pour la qualité de la collecte et est explicitement mentionnée dans le décret n° 2003-485 du 5 juin 2003 modifié relatif au recensement de la population : « Au plus tard avant le premier jour de la collecte d'informations, les communes ou établissements publics de coopération intercommunale concernés effectuent une tournée de reconnaissance, dans les conditions fixées par un arrêté du ministre chargé de l'économie, pour vérifier les informations figurant sur la liste d'adresses à recenser et informent l'Institut national de la statistique et des études économiques des modifications à apporter. »

Dans les communes de moins de 10 000 habitants, l'agent recenseur vérifie qu'il n'y aurait pas d'adresse oubliée ou en trop en parcourant son secteur. Il vérifie également s'il n'y a pas d'écart entre le nombre de logements prévus et celui qu'il constate sur le terrain. Cette tournée de reconnaissance permet aussi de s'assurer que l'agent recenseur a bien identifié les limites de son secteur, notamment qu'il a bien repéré les

8 Pour le recensement, un logement occasionnel est un logement occupé une partie de l'année pour des raisons professionnelles par une personne qui dispose d'un autre logement familial. Ce dernier est alors considéré comme sa résidence principale.

limites de la commune et, s'il y a plusieurs agents recenseurs, qu'il a bien repéré les limites des zones de ses voisins. Il s'agit en effet de se prémunir contre le risque qu'une partie de territoire ne soit pas recensée (chaque agent recenseur pensant qu'un collègue s'en charge) ou au contraire que les habitants reçoivent la visite de deux agents recenseurs différents.

Ces secteurs de travail, appelés « zones de collecte », ont été déterminés par les communes lors de la phase préparatoire. Il s'agit d'une partition du territoire qui n'omet aucun espace. L'Insee a pour mission de valider ce découpage, notamment en vérifiant qu'il respecte un certain nombre de règles facilitant les repérages sur le terrain.

Dans les communes de plus de 10 000 habitants, l'agent recenseur vérifie qu'il parvient à bien repérer les adresses de l'échantillon. Dans une enquête par sondage, il est indispensable que les agents recenseurs identifient sans erreur les adresses à recenser (et n'enquêtent pas une adresse non tirée au sort). En complément de sa liste d'adresses, l'agent recenseur dispose de plans de collecte pour faciliter le repérage. Une procédure de vérification est également prévue en cas de doute sur la détermination des bonnes adresses à recenser (cf. partie 2-C). Il vérifie aussi s'il n'y a pas d'écart entre le nombre de logements prévus et celui qu'il constate sur le terrain.

La tournée de reconnaissance permet également de mener une communication locale (dépôt de courriers annonçant l'enquête, pose d'affiches dans les halls d'immeubles...). Elle permet enfin à l'agent recenseur d'anticiper d'éventuelles difficultés.

Avec une liste d'adresses et de logements à jour, des zones de collecte bien découpées et des adresses bien identifiées par les agents recenseurs, une grande partie de l'exhaustivité ou du respect de l'échantillon est assurée. Les outils de suivi intègrent ces informations et permettent de connaître à tout moment le degré d'avancement des opérations de collecte.

Enfin, dénombrer les habitants de chaque logement

La dernière étape consiste alors à déterminer la catégorie de logement et à dénombrer les habitants des résidences principales. En effet, les définitions relatives à la population stipulent que les habitants doivent être pris en compte uniquement dans la commune de leur résidence principale. Aucun habitant n'est comptabilisé dans les résidences secondaires, les logements vacants et les logements occasionnels.

L'agent recenseur doit ainsi se rendre dans tous les logements de sa liste et déterminer en premier lieu leur catégorie. S'il s'agit d'une résidence non principale, il recueille quelques informations simples sur le logement et ne fait pas remplir de questionnaires. Dans la grande majorité des cas, ces logements sont vides au moment du passage de l'agent recenseur. Celui-ci obtient alors les informations par le voisinage.

Dans les résidences principales, l'agent recenseur rencontre les habitants. Il explique l'utilité du recensement et les règles de confidentialité qui y sont associées ; puis il propose de répondre par Internet en remettant une notice comprenant des codes de connexion sécurisés. Si les personnes ne peuvent pas répondre par Internet, il remet alors des questionnaires papier (cf. partie 1-C) et prend rendez-vous pour venir les récupérer remplis quelques jours plus tard. En 2019, 52 % des résidences principales ont été recensées par Internet, représentant ainsi 60 % des habitants⁹.

Des tests sont menés depuis 2018 pour simplifier ce protocole. Il s'agit, dans un premier temps, de déposer la notice permettant de se connecter au questionnaire en ligne directement dans la boîte aux lettres, puis d'attendre 4 à 5 jours que les personnes souhaitant répondre spontanément et rapidement le fassent. Ensuite, le protocole de visite reprend uniquement auprès des habitants qui n'ont pas déjà répondu. Le

⁹ Le taux en résidences principales est calculé comme « nombre de résidences principales répondant par Internet / nombre total de résidences principales (répondant sur papier, sur Internet et non-répondantes) ». Le taux en habitants est calculé comme « nombre d'habitants des logements ayant répondu par Internet / nombre d'habitants des logements ayant répondu (par papier et par internet) ».

résultat de ces tests valide une généralisation de ce protocole sur les adresses individuelles, qui sera effective pour l'EAR 2021.

Des critères pour être comptabilisé en tant qu'habitant permanent et un questionnaire auto-administré

La comptabilisation des habitants suppose de mettre en œuvre deux critères successifs. D'abord déterminer si la personne est éligible à être recensée en France, puis déterminer dans quelle commune elle doit être comptabilisée. La méthode utilisée en France suit les recommandations des Nations unies (cf. encadré 1-C.1 et [Unece, 2015]).

La notion clé est celle de « *population usuelle résidente* ». La « *population de résidents habituels* » d'un pays se compose des personnes qui ont leur lieu de résidence habituelle dans le pays au moment de référence du recensement et qui y ont résidé ou ont l'intention d'y résider pendant une période continue d'au moins 12 mois. L'expression « période continue » signifie que les absences (du pays de résidence habituelle) dont la durée est inférieure à 12 mois ne modifient pas le lieu de résidence habituelle. Les mêmes critères s'appliquent à toutes les divisions territoriales du pays.

Cela implique notamment qu'une personne de nationalité étrangère qui vit en France depuis au moins 12 mois ou qui a l'intention d'y vivre au moins 12 mois doit être recensée. Inversement, une personne de nationalité française qui vit toute l'année dans un autre pays et qui revient en France uniquement pour les vacances ne doit pas être recensée.

Les personnes qui n'ont pas de résidence habituelle dans un pays précis sont comptabilisées sur le lieu où elles se trouvent au moment de référence du recensement.

Ces règles pour la détermination de la commune de résidence habituelle sont traduites en France par le décret n° 2003-485 du 5 juin 2003 (cf. encadré 2-B.1). Il précise la définition des habitants permanents à comptabiliser dans les résidences principales. Dans l'immense majorité des cas, les habitants du logement y vivent toute l'année et le remplissage du questionnaire est simple.

Toutefois, quelques situations méritent d'être précisées. Il s'agit principalement des étudiants qui vivent dans un logement sur le lieu de leurs études et qui reviennent le week-end et les vacances dans le logement de leurs parents et des enfants en résidence partagée à l'issue d'une séparation ou d'un divorce. On peut également citer le cas des conjoints éloignés pour raisons professionnelles qui vivent la semaine sur leur lieu de travail et qui reviennent le week-end dans leur résidence familiale ou des personnes qui sont hébergées dans une communauté (centre pour handicapés, hôpital...).

Encadré 2-B.1 : Les cas de multi-résidence

Le décret n° 2003-485 du 5 juin 2003 précise où comptabiliser les personnes qui vivent dans plusieurs logements. Parmi les situations les plus fréquentes, on trouve les élèves ou étudiants qui vivent la semaine dans un internat ou un logement proche de leur lieu d'études et qui reviennent le week-end chez leurs parents. S'ils sont majeurs, ils sont comptabilisés dans la population usuelle résidente de la commune de lieu d'étude ; s'ils sont mineurs, ils sont comptabilisés dans la commune de leurs parents. En complément, un dispositif de « double-compte » administratif (population comptée à part – cf. encadré 1-A.1) est prévu pour que chaque commune « bénéficie » de cette population pour l'application de certains textes juridiques, notamment pour le calcul des dotations financières versées par l'État.

Les enfants en résidence partagée à la suite d'un divorce ou d'une séparation sont comptabilisés dans la commune où ils passent le plus de temps. En cas d'égalité parfaite, ils sont comptabilisés dans la commune où ils ont dormi à la date de référence du recensement.

Les conjoints éloignés de leur famille pour des raisons professionnelles et qui reviennent dans leur domicile familial le week-end sont comptabilisés dans la commune de ce domicile familial.

Quel que soit le mode de réponse, internet ou papier, le questionnaire est auto-administré¹⁰. Ce sont les habitants eux-mêmes qui renseignent seuls leurs réponses. L'agent recenseur peut informer les personnes enquêtées si elles s'interrogent sur les critères à remplir pour pouvoir être considéré comme un habitant du logement, mais il ne peut pas corriger les choix des habitants. Dans ce contexte, une erreur de mesure peut intervenir *via* deux mécanismes. Le premier est la fausse déclaration volontaire de la part des personnes recensées. Le second est la mauvaise compréhension des questions et des consignes sous-jacentes.

Pour limiter cette erreur de mesure, il est important de bien informer les personnes enquêtées. Cette information passe tout d'abord par un rappel du contexte juridique et des finalités du recensement. En particulier, il est largement indiqué que l'anonymat des réponses est garanti et que les informations individuelles collectées lors du recensement ne sont jamais transmises à une autre administration. Le recensement de la population est une enquête couverte par la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistique et poursuit une finalité uniquement statistique. Répondre est obligatoire et en contrepartie les réponses restent confidentielles. Concrètement, cela se traduit par le fait que les questionnaires papier ou internet sont physiquement détruits avant le 31 décembre de l'année d'enquête et que les noms et prénoms des personnes ne sont pas intégrés dans les fichiers de données. Ils sont néanmoins nécessaires au moment de la collecte pour les contrôles d'exhaustivité et après la collecte pour les contrôles post-collecte.

En second lieu, pour limiter les erreurs de mesure, les questions et les consignes sont formulées de manière à ce que les personnes recensées puissent être comptabilisées au plus proche de leur situation réelle, sous contrainte d'ergonomie et de place sur le questionnaire. Sur le questionnaire internet, il est proposé une liste de questions de situation qui permettent ensuite automatiquement de filtrer le reste du questionnaire selon que la personne réside ou pas de manière permanente dans le logement.

Par exemple, pour un étudiant :

Si Sandrine MAURIN vit aussi dans un autre logement, merci de compléter la partie suivante :

En raison de ses ÉTUDES, Sandrine MAURIN :

- est logée ailleurs (internat, logement indépendant, chambre en ville, etc.) et revient vivre ici pour les week-ends ou les vacances.
- vit ici et retourne dans son logement familial pour les week-ends ou les vacances.

ou pour un enfant en résidence alternée :

Si Arthur MAURIN vit aussi dans un autre logement, merci de compléter la partie suivante :

Suite à une SÉPARATION OU UN DIVORCE, Arthur MAURIN habite également chez son autre parent (père ou mère) et :

- vit ici plus de la moitié du temps.
- vit ici la moitié du temps.
- A-t-il dormi ici la nuit du début du recensement, soit du lundi 17 au mardi 18 avril: Oui. Non.
- vit ici moins de la moitié du temps.

¹⁰ Des exceptions à cette règle peuvent se produire si la personne a des difficultés pour répondre seul au questionnaire

Sur le questionnaire papier, des consignes sont indiquées pour définir les catégories de population à partir de cas concrets.

Par exemple, pour les habitants permanents :

Inscrivez les personnes qui vivent dans ce logement la plus grande partie de l'année, y compris :

- les nourrissons même encore à la maternité ;
- les personnes temporairement absentes ;
- les sous-locataires et colocataires ;
- les enfants qui habitent également chez un autre parent à la suite d'une séparation ou d'un divorce et qui vivent dans ce logement :
 - plus de la moitié du temps,
 - ou la moitié du temps et y ont passé la nuit du début du recensement, soit du mercredi 15 au jeudi 16 janvier.

Ne vous oubliez pas vous-même.

Inscrivez également les :

- enfants mineurs logés ailleurs pour leurs études dont ce logement est la résidence familiale ;
- conjoints éloignés pour raisons professionnelles ;
- personnes majeures qui habitent ici pour leurs études ;
- personnes présentes dans ce logement et qui n'ont pas de résidence habituelle ailleurs ;
- employés de maison, salariés et jeunes au pair qui habitent ici.

Malgré les précautions prises pour bien informer les personnes enquêtées des règles permettant de déterminer qui est un habitant permanent du logement et qui ne l'est pas, il peut rester une marge d'appréciation de la part des répondants.

Pour les personnes ne vivant pas en logement ordinaire (personnes ne disposant pas d'un logement fixe ou personnes vivant dans une communauté), la procédure de collecte est adaptée : cf. encadré 2-B.2.

Que se passe-t-il en cas de refus de répondre ?

La non-réponse à l'enquête regroupe trois types de situation : les personnes absentes de longue durée, les personnes impossibles à joindre et les refus. Au total, la part de la non-réponse est très faible ; elle s'est élevée à 3,9 % en 2019, dont 36 % de refus explicites. Ce taux de non-réponse est extrêmement faible au regard des taux de non-réponse observés pour les autres enquêtes statistiques, même obligatoires. Ce très bon résultat est en grande partie dû à la proximité entre les habitants et les organisateurs sur le terrain (la commune qui organise et l'agent recenseur qui habite la commune ou le quartier). Elle s'explique sans doute aussi par les arguments mis en avant pour convaincre les personnes, fondés sur l'utilité sociale de l'opération et sur son caractère obligatoire. Enfin, les questionnaires sont relativement courts et les thèmes abordés simples par rapport aux enquêtes thématiques.

Il reste que ce bon taux global masque une grande diversité de situations. Dans les communes de plus de 10 000 habitants, la non-réponse est en moyenne de 5,6 %. Dans 9 % de ces communes (soit environ 90 communes), elle est supérieure à 10 %.

En cas de refus, les communes mettent en œuvre différents moyens : courrier du maire, appel des services communaux, changement d'agent recenseur. Le recensement étant une enquête obligatoire, une amende de 38 € peut être infligée aux personnes qui refusent de répondre, conformément au dernier alinéa de l'article 7 de la loi du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. Ce moyen n'est que très rarement employé, le principe général étant avant tout de convaincre les personnes de l'utilité du recensement.

En cas d'impossibilité de réaliser l'enquête, il est demandé à l'agent recenseur d'estimer, s'il le peut, le nombre d'habitants du logement en recueillant les informations auprès du voisinage. Au niveau national en 2019, parmi les 3,9 % de non-réponses, cette information sur le nombre de personnes résidant habituellement dans le logement a pu être obtenue dans 74 % des cas. Si ce n'est pas possible, il indique qu'il ne connaît pas cette information et une procédure statistique d'estimation sera appliquée pour le logement concerné (cf. partie 2-D).

Encadré 2-B.2 : Les personnes ne vivant pas en logement ordinaire

Les personnes qui ne disposent pas d'un logement fixe

Pour les personnes qui ne disposent pas d'un logement fixe (habitations mobiles, personnes sans abri, bateliers), la procédure est adaptée. Les habitations mobiles (caravanes) et les personnes sans abri (vivant dans la rue, dans des tentes ou d'autres types d'abris précaires) sont recensées dans la commune où elles se trouvent durant les deux premiers jours de l'enquête. Cette durée courte permet d'éviter que ces personnes puissent être recensées dans plusieurs communes différentes au cours des 4 ou 5 semaines d'enquête. L'opération a lieu une fois tous les 5 ans dans les communes de plus de 10 000 habitants (les années en 1 et en 6) et l'année de recensement dans les communes de moins de 10 000 habitants.

Pour les habitations mobiles, la commune repère au préalable si les aires d'accueil de gens du voyage sont occupées ou si des caravanes sont installées en d'autres lieux. Puis, pendant les deux premiers jours de l'enquête, les agents recenseurs procèdent, comme pour les logements ordinaires, en rencontrant les personnes et en faisant renseigner des questionnaires papier (la collecte par Internet n'est pas proposée dans ces cas).

Pour les personnes sans abri, il est conseillé aux communes de réaliser l'opération en collaboration avec les associations venant en aide à ce public. Là encore, le repérage préalable des lieux où sont installées ces personnes est essentiel. Dans les communes de plus de 10 000 habitants qui comprennent de nombreux sans-abri, la procédure de repérage des lieux fait l'objet d'instructions associées au remplissage de documents spécifiques. Puis, pendant les deux premiers jours de l'enquête, les agents recenseurs vont à la rencontre des personnes pour leur faire remplir uniquement un bulletin individuel papier. Les conditions de dialogue n'étant pas toujours aisées, il est autorisé de ne remplir que très partiellement le questionnaire.

Le recensement des bateliers concerne les personnes exerçant une activité professionnelle de navigation fluviale (les péniches fixes à quai servant de logement ne sont pas dans ce cas). L'enquête a lieu une fois tous les 5 ans (les années en 1 et en 6). En 2016, elle a été réalisée par voie postale.

Certains lieux d'habitation se situent à la frontière entre la définition d'un logement et la définition des personnes sans abri. Il s'agit notamment des immeubles squattés ou des caves habitées. Dans les communes de moins de 10 000 habitants où l'enquête est exhaustive, la consigne est de tous les recenser. Dans les communes de plus de 10 000 habitants où l'enquête est réalisée sur un échantillon d'adresses, les consignes dépendent de la manière dont ces habitations figurent dans la base de sondage (le répertoire d'immeubles localisés - RIL). Si l'immeuble squatté (ou la cave habitée) figure dans la base de sondage et est tiré au sort, il doit être recensé. En revanche, s'il n'est pas tiré au sort, il ne sera pas enquêté. Dans le cas où l'immeuble ne figure pas dans le RIL, ses habitants devraient être recensés dans le cadre du recensement des personnes sans abri.

Les personnes qui vivent en communauté

L'enquête de recensement dans les communautés (cf. annexe 1) est effectuée par les enquêteurs de l'Insee et non par des agents recenseurs des communes. En petites communes, les communautés sont recensées exhaustivement la même année que les logements ordinaires. En grandes communes, si le nombre de

communautés est élevé, il arrive que les communautés soient réparties dans différents groupes de rotation (de 2 à 5 selon la commune). À défaut, toutes les communautés sont dans le même groupe. Les communautés sont alors recensées de manière exhaustive l'année correspondant à leur groupe de rotation.

Certaines communautés peuvent être constituées en partie de logements (cf. annexe 1). C'est par le cas de certaines cités universitaires où certains étudiants vivent dans des chambres et d'autres dans des logements qui disposent d'un endroit pour cuisiner et de sanitaires privatifs. Pour des raisons pratiques, l'enquête de recensement dans les communautés recense l'ensemble de la population y résidant : la population de la communauté et la population des logements ordinaires situés dans ces communautés.

La population des communautés est recensée avec un bulletin individuel communauté (BIC) qui ne se différencie du bulletin individuel utilisé dans les ménages que par une question supplémentaire : « Avez-vous une résidence personnelle dans une autre commune [...] si oui, précisez où » (utilisée pour le calcul de la population comptée à part (PCAP)). La population détenue dans des établissements pénitentiaires est recensée avec un bulletin individuel établissement pénitentiaire (BIEP) qui ne comprend pas la question sur la commune de résidence personnelle (les détenus ne comptent pas dans la PCAP) ni les questions liées à l'emploi. La population des logements ordinaires situés au sein de la communauté est recensée avec des bulletins de type ménages (feuille de logement et bulletins individuels).

Les personnes vivant en communauté (ou dans ses logements) n'ont, pour le moment, pas la possibilité de répondre aux questionnaires du recensement par Internet.

Comme pour l'enquête auprès des logements ordinaires, la garantie d'exhaustivité est d'abord assurée par l'identification préalable de toutes les communautés concernées une année donnée. La liste est mise à jour en continu avec des sources administratives et est expertisée par les communes. Par ailleurs, avant l'enquête, chaque communauté reçoit un courrier de l'Insee destiné à confirmer son existence et indiquer une première estimation du nombre de chambres et du nombre de logements.

Malgré toutes ces précautions, il peut arriver qu'une communauté se révèle fermée pendant l'enquête ou, au contraire, soit découverte pendant les opérations de terrain. Ces cas sont rares (0,6 % de fermetures et 1,2 % de découvertes en 2019) et corrigés immédiatement.

Les enquêteurs de l'Insee mobilisés pour le recensement des communautés suivent une formation d'une journée en janvier. Ils prennent contact avec la communauté dans la semaine qui précède la collecte, afin de prendre rendez-vous pour le dépôt des imprimés. Lors du premier rendez-vous, les enquêteurs sensibilisent leur interlocuteur au recensement, expliquent la procédure et déposent les imprimés de collecte. C'est le correspondant de la communauté qui va distribuer les imprimés aux résidents et aux occupants des logements et qui va les récupérer ensuite. L'enquêteur prend rendez-vous pour venir récupérer plus tard les imprimés collectés. L'enquêteur ne se déplace jamais dans les chambres et les logements et n'entre pas en contact avec la population de la communauté.

Lorsque des résidents de la communauté ne sont pas aptes à répondre (personnes très âgées, handicapées, ayant des difficultés cognitives ou ne maîtrisant pas la langue française...) ou que certains refusent de répondre, l'enquêteur se tourne vers le correspondant de la communauté afin d'obtenir le minimum d'informations permettant de garantir qu'aucun individu ne sera omis dans la population. Les informations obtenues sont plus ou moins complètes, allant d'un simple comptage jusqu'à une liste contenant des éléments socio-démographiques (sexe, âge, nationalité, situation matrimoniale, profession...). À partir de ces informations, des BIC sont renseignés à la place des personnes.

Au total, environ 55 % des résidents des communautés ne remplissent pas eux-mêmes de questionnaire. Ils sont recensés au moyen d'informations administratives détenues par les structures. Les communautés peuvent en effet transmettre ces données administratives à l'Insee grâce à la loi de 1951 sur l'obligation, la coordination et le secret en matière de statistiques et aux dispositions du Règlement général sur la protection des données (RGPD) qui autorisent les réutilisations de données administratives à des fins

statistiques, sous réserve des garanties des droits et libertés des personnes concernées. Cette proportion s'élève à 84 % dans les Ehpad, mais n'est que de 22 % dans les internats d'établissements d'enseignement (cf. tableau E2-B.2).

Tableau E2-B.2 : Pourcentage de bulletins établis à partir d'une liste, par sous-catégorie de communauté

Sous-catégories	2015	2016	2017	2018
11- Maison de retraite, Ehpad ¹¹	74 %	84 %	84 %	84 %
12- Foyer, résidence sociale, Cada et Auda gérés par Adoma	89 %	87 %	87 %	85 %
13- Autre foyer de travailleurs	54 %	65 %	54 %	53 %
14- Autre sanitaire ou social de moyen ou long séjour	63 %	76 %	75 %	70 %
21- Communauté religieuse	16 %	19 %	22 %	17 %
31- Gendarmerie	9 %	2 %	7 %	3 %
32- Autre établissement militaire	39 %	32 %	36 %	30 %
41- Cité universitaire	68 %	46 %	52 %	55 %
42- Autre internat	20 %	21 %	23 %	22 %
51- Établissement pénitentiaire	58 %	66 %	65 %	79 %
61- Établissement social de court séjour	53 %	62 %	62 %	73 %
71- Autre catégorie de communauté	80 %	76 %	88 %	90 %
Total	52 %	57 %	60 %	57 %

Source : feuille de communauté renseignée par l'enquêteur.

¹¹ Les taux sont peut-être sous-estimés, car tous les enquêteurs ne pensent pas à préciser le remplissage sur liste dans les Ehpad, ce mode de collecte étant en général considéré comme incontournable dans ces établissements.

C) Les contrôles de l'enquête

Les contrôles de la bonne exécution de l'enquête sont menés à la fois par les communes et par l'Insee.

Le coordonnateur communal est chargé d'encadrer au quotidien les agents recenseurs. Le superviseur de l'Insee est chargé de vérifier la bonne réalisation de l'enquête par les communes.

C-1. Les contrôles effectués par le coordonnateur communal

En premier lieu, le coordonnateur communal (cf. encadré 2-C.1) doit rencontrer au minimum une fois par semaine ses agents recenseurs individuellement. Lors de cette rencontre, ils font le point sur l'avancement de la collecte et le coordonnateur examine les documents remis par l'agent recenseur. En particulier, il examine le document en retour de la tournée de reconnaissance ; celui-ci doit porter des annotations sur les adresses. Il examine également quelques questionnaires papier pour s'assurer qu'ils sont correctement remplis, notamment sur les identifiants du logement qui sont essentiels pour un bon suivi des réponses. Il s'assure également que les carnets de tournée des agents recenseurs sont renseignés avec des informations cohérentes (le carnet de tournée est un document de travail de l'agent recenseur. Il comporte la liste des adresses et logements à recenser).

Lors de ces rencontres, l'agent recenseur fait part des difficultés auxquelles il est confronté et le coordonnateur étudie avec lui les solutions qui peuvent être apportées, notamment la mise en place de relance auprès des habitants refusant de répondre.

Par ailleurs, le coordonnateur dispose, dans l'application de gestion de la collecte fournie par l'Insee (Omer), des informations sur le nombre de logements recensés à une adresse et sur le nombre d'habitants de chaque logement. Par comparaison avec les données issues de la phase de préparation de l'enquête, il peut repérer d'éventuels écarts à expliquer.

Le coordonnateur est également informé par l'application de cas de réponses internet à vérifier. Il s'agit notamment de situations où le ménage enquêté a modifié l'adresse de son logement, ce qui constitue un signal d'alerte sur le fait que l'agent recenseur ne serait peut-être pas allé à la bonne adresse. Dans les communes de plus de 10 000 habitants, le respect de l'échantillon est essentiel pour la qualité des résultats et il n'est pas admis qu'un logement non tiré au sort soit recensé.

À la fin de l'enquête, la commune est responsable de la vérification de son exhaustivité. Le coordonnateur communal dispose dans l'application de gestion de l'état des lieux des adresses et logements recensés et peut voir les éventuels manques. À l'issue des opérations de contrôle, le maire signe un document attestant du nombre de logements et d'habitants recensés sur sa commune.

Encadré 2-C.1 : Le coordonnateur communal : acteur clé de l'enquête de recensement

Dans le cadre du partenariat entre l'Insee et les communes institué pour le recensement par la loi n° 2002- 276 du 27 février 2002 relative à la démocratie de proximité, les communes doivent nommer un coordonnateur communal. Le coordonnateur communal est l'interlocuteur principal de l'Insee pour toutes les opérations de préparation et de réalisation de l'enquête sur le territoire de la commune. En amont de la collecte, il vérifie les listes d'adresses, détermine les secteurs de travail des agents recenseurs, prépare la communication locale et gère la logistique des imprimés et des moyens de formation. La commune est responsable du recrutement des agents recenseurs et le coordonnateur les encadre au quotidien pendant l'enquête. Il vérifie leur travail et les aide à trouver des solutions pour les cas où la collecte présente des difficultés.

C-2. Les contrôles effectués par l'Insee

Les contrôles effectués par l'Insee sont réalisés par le superviseur sur le terrain pendant la phase de collecte, puis par les établissements régionaux de l'Insee après l'enquête.

Plusieurs moyens sont mis en œuvre pour cela :

- la formation des acteurs et l'assistance-conseil auprès des communes ;
- le contrôle sur pièce d'un échantillon de documents pendant l'enquête ;
- le contrôle du respect de l'échantillon dans les communes de plus de 10 000 habitants ;
- le contrôle de faux questionnaires sur Internet ;
- la comparaison du nombre et de la catégorie de logements collectés avec des informations fiscales ou des résultats des recensements précédents, pendant et après l'enquête ;
- les contrôles sur le terrain par des enquêteurs de l'Insee après l'enquête ;
- les contrôles à l'occasion de la phase de lecture optique des questionnaires.

Dans tous les cas, l'Insee contrôle uniquement le nombre de logements, la catégorie des logements et le nombre de personnes comptabilisées dans les logements. Lors des phases de collecte et de contrôle post-collecte, l'Insee ne contrôle jamais les réponses aux questions socio-démographiques (date de naissance, nationalité, statut conjugal, formation, emploi...) qui relèvent de la seule déclaration et responsabilité des personnes enquêtées. Cependant, certains contrôles de cohérence sont effectués au moment du redressement des variables, qui peuvent amener à modifier certains éléments déclarés.

La formation des acteurs et l'assistance-conseil auprès des communes

La formation des acteurs et leur bonne connaissance du protocole de l'enquête constituent un levier majeur pour garantir la qualité des résultats. Chaque année, les 10 000 coordonnateurs communaux reçoivent une journée de formation dispensée par l'Insee. De même, les 24 000 agents recenseurs sont formés en deux demi-journées. Un manuel du recensement ainsi que de nombreuses fiches thématiques sont mises à disposition des coordonnateurs communaux et chaque agent recenseur se voit remettre un livret d'instructions.

Ces séances de formation ne suffisent pas pour maîtriser complètement le processus. Elles sont complétées par une assistance pendant la préparation de l'enquête et pendant la collecte. 400 agents de l'Insee, dénommés « superviseurs du recensement » se rendent régulièrement en commune pour faire le point sur le déroulement de l'enquête et apporter des compléments de formation. Selon la taille des communes, un superviseur peut être en charge de 5 à 30 communes (5, en métropole, s'il s'agit de communes de plus de 100 000 habitants, 30 s'il s'agit de communes de moins de 1 000 habitants). Il se déplace physiquement dans les locaux des communes pour s'entretenir avec le coordonnateur communal et est joignable en permanence.

Le contrôle sur pièce d'un échantillon de documents

Lors de ses visites en commune, le superviseur doit avoir accès aux documents d'enquête et procède à quelques vérifications. Il ne s'agit pas de vérifier tous les documents, mais de s'assurer que les différents acteurs de terrain ont bien compris les consignes. En particulier, le superviseur doit s'assurer que le coordonnateur communal a bien vérifié les documents remplis par les agents recenseurs. Pour cela, il

examine également les documents en retour de la tournée de reconnaissance, quelques questionnaires papier et les carnets de tournée.

Sur les questionnaires papier, le superviseur vérifie notamment que les écritures manuscrites sont différentes et que les réponses ne sont pas répétitives (pour repérer éventuellement un agent recenseur qui renseignerait lui-même des bulletins sans visiter les personnes).

Le superviseur s'assure également que le coordonnateur communal traite bien les cas de questionnaires internet dont l'adresse a été changée par les habitants.

Le contrôle du respect de l'échantillon dans les communes de plus de 10 000 habitants

Le respect de l'échantillon est essentiel pour garantir la pertinence de l'élaboration des résultats. Il s'agit donc d'un point de contrôle important pour les superviseurs. Il peut y avoir des écarts entre la dénomination d'une adresse issue du RIL et la réalité sur le terrain. Il existe également des configurations de bâtis complexes (cour intérieure, plusieurs bâtiments à une même adresse, plusieurs accès sur des voies différentes...). Par ailleurs, il a pu y avoir des évolutions entre la phase de préparation du RIL et l'enquête sur le terrain.

Pour gérer ces situations où l'agent recenseur peut douter de la bonne adresse à recenser, un dispositif de « fiche-navette » a été mis en place. En cas de question ou de doute, le coordonnateur communal transmet à l'Insee cette fiche qui explique la configuration du terrain et les interrogations qu'elle suscite. L'Insee, à l'aide de divers outils cartographiques, analyse la situation et informe la commune de l'action à mener (recenser ou pas). Dans la majorité des cas, ces signalements portent sur des écarts de nombre de logements (22 %) ou de modifications dans l'usage du bâtiment (17 %). 7 % des signalements concernent des adresses détruites ou murées qui ne sont alors pas recensées et sont supprimées du RIL.

Ce dispositif complète le contrôle des adresses déclarées par les habitants, sur Internet ou sur papier, qui permet de signaler des cas où l'agent recenseur se serait trompé d'adresse.

Le contrôle de faux questionnaires sur Internet

Ce contrôle a pour objectif de repérer d'éventuels cas où un agent recenseur peu scrupuleux renseignerait lui-même des questionnaires internet fictifs sans rencontrer les habitants. Pour cela, des vérifications de concentration massive de réponses à partir d'une même adresse IP sont effectuées.

Ce contrôle complète celui mené sur les questionnaires papier en vérifiant la diversité des écritures manuscrites.

La comparaison du nombre et de la catégorie de logements collectés avec des informations fiscales ou des résultats des recensements précédents, pendant et après l'enquête

Ce contrôle a pour objectif de repérer des évolutions atypiques par rapport au passé récent ou des écarts avec des sources d'informations fiscales. L'Insee dispose, pour le recensement, d'informations issues des fichiers fiscaux, notamment de la taxe d'habitation. Celles-ci ne portent que sur le nombre de logements et le nom des occupants à chaque adresse, mais ne portent pas sur les revenus ou le montant des taxes.

Pendant la phase de préparation de l'enquête et pendant l'enquête, le superviseur de l'Insee dispose du nombre de logements par adresse ou par voie connu par la taxe d'habitation. Cette information lui permet de repérer les écarts importants qui nécessitent une explication. Celle-ci est obtenue par le dialogue avec la commune. Pendant la collecte, l'information fiscale disponible date de 2 ans, ce qui explique une grande partie des écarts. Le superviseur dispose aussi, pour les communes de moins de 10 000 habitants, des

résultats du recensement précédent, ce qui lui permet également de repérer des évolutions atypiques à expliquer.

Après la collecte, l'Insee dispose d'informations fiscales plus fraîches (datant d'un an). S'enclenche alors un dispositif de contrôle plus systématique et automatisé. Pour la totalité des communes recensées, sont calculés plusieurs indicateurs pertinents pour estimer la qualité de la collecte. Il s'agit principalement :

- des écarts de nombre de logements avec la source fiscale ou avec le RIL ;
- des écarts dans les catégories de logements avec la source fiscale (résidence principale, secondaire, logement vacant) ;
- d'un nombre moyen de personnes par logement atypique (en niveau ou en évolution) ;
- de l'importance du nombre de non-réponses (en niveau ou en évolution) ;
- d'un nombre estimé de personnes par logement non recensé atypique ou trop souvent inconnu.

Ces indicateurs sont calculés par commune ou pour des zones infra-communales. Les zones où ces indicateurs dépassent certains seuils sont alors vérifiées d'abord en bureau. Il s'agit d'examiner, logement par logement, la cohérence des informations. Les noms des personnes habitant les logements sont utiles pour repérer d'éventuelles anomalies. Une fois les contrôles terminés (et au plus tard au 31 décembre de l'année), les noms et prénoms sont définitivement effacés des fichiers, conformément à l'arrêté de traitement du recensement du 4 février 2016.

Si les écarts ne peuvent pas s'expliquer avec les informations disponibles en bureau, des contrôles terrain peuvent être déclenchés.

Les contrôles sur le terrain par des enquêteurs de l'Insee après l'enquête

Les enquêtes de contrôle terrain après le recensement sont déclenchées après la phase de contrôles systématiques et automatisés de l'ensemble des communes, sur une sélection de cas présentant des anomalies significatives. Ces enquêtes peuvent avoir plusieurs objectifs selon les anomalies constatées. Il peut s'agir de :

- vérifier l'existence réelle d'une adresse ou d'un logement (ajout frauduleux de logements) ;
- vérifier l'absence réelle d'une adresse ou d'un logement (oubli de l'agent recenseur) ;
- vérifier la catégorie de logement (résidence principale, secondaire, logement vacant) ;
- récupérer des réponses dans les logements non répondants.

La réalisation de contrôles sur le terrain est soumise à des contraintes de moyens. De ce fait, ils ne sont effectués que dans des zones qui présentent une concentration suffisante d'anomalies pour justifier le déplacement d'un enquêteur.

Au total, peu d'anomalies sont constatées sur les 8 000 communes recensées chaque année. Par exemple, en 2019, 486 communes ont fait l'objet d'un contrôle terrain. Pour 418 d'entre elles, les résultats d'enquête ont été modifiés suite à ces contrôles :

Population modifiée entre...	Nombre de communes
0 à moins de 1 %	201
1 à moins de 2 %	105
2 à moins de 3 %	41
3 à moins de 10 %	65
Plus de 10 %	6

Le maire de la commune est informé au préalable des contrôles qui peuvent être réalisés sur le territoire de sa commune. Dans les cas où ces contrôles ont conduit à une modification des résultats de l'enquête, le maire en est informé par courrier.

Les contrôles à l'occasion de la phase de lecture optique des questionnaires

Les données figurant sur les questionnaires papier sont acquises par lecture optique. Cette phase du processus comprend également la réunion des questionnaires papier et internet par adresse (par exemple dans les cas où certains habitants d'un même immeuble auraient répondu par Internet et d'autres par papier).

Ce processus bénéficie d'un dispositif spécifique de mesure de la qualité de la lecture optique qui a pour objectif principal de garantir la qualité des données socio-démographiques. Toutefois, il produit aussi des effets sur le contrôle du dénombrement des personnes. En particulier dans deux situations :

- lorsque des réponses internes au questionnaire peuvent changer la catégorie du logement (par exemple, un logement se révèle en fait être une résidence secondaire et non une résidence principale, car personne ne s'y déclare comme résident permanent) ;
- lorsqu'on constate un doublon de questionnaire entre Internet et papier.

Ces cas sont rares, mais procèdent aussi du contrôle qualité du dénombrement de la population.

D) Le traitement de la non-réponse

Une Fiche de logement non enquêtée (FLNE) est renseignée en général en fin de collecte, lorsque l'agent recenseur n'a pas pu obtenir de réponse des habitants d'un logement, que les habitants refusent de répondre ou qu'ils sont absents pour une longue durée. Cette fiche informatisée est établie par l'agent recenseur en lieu et place d'une feuille de logement. Elle doit être renseignée uniquement pour les résidences principales.

L'**information essentielle** d'une FLNE est le nombre supposé de personnes résidant dans le logement. On dit qu'une FLNE est :

- **renseignée** dans le cas où l'agent recenseur a indiqué le nombre supposé de personnes dans le logement ;
- **non renseignée** dans le cas où l'agent recenseur n'a pas indiqué le nombre supposé de personnes dans le logement.

Pour renseigner une FLNE, l'agent recenseur interroge le voisinage pour déterminer, d'une part, si le logement non répondant est une résidence principale ou non et, d'autre part, le nombre de personnes qui y résident. Ces informations sont indispensables pour dénombrer de manière la plus précise possible la population.

L'agent recenseur n'ayant pu contacter les habitants du logement, il n'est jamais certain qu'il s'agisse bien d'une résidence principale, en particulier lorsqu'il n'a pu obtenir avec certitude cette information auprès du voisinage. **Ainsi, des FLNE peuvent être remplies à tort pour des résidences non principales et cela concerne essentiellement les FLNE non renseignées.**

Dans le cadre du dénombrement de la population, seules les FLNE non renseignées correspondent à de la vraie non-réponse. Elles font alors l'objet de redressements pour déterminer la part de résidences principales au sein de ces FLNE et aussi pour déterminer un nombre de personnes par FLNE. Les informations fournies dans les FLNE renseignées sont en revanche directement intégrées dans le dénombrement de la population.

D-1. Quelques statistiques sur les FLNE pour la collecte 2019

Le taux de non-réponse totale à l'enquête annuelle de recensement est très faible, comparé aux taux de non-réponse aux enquêtes auprès des ménages. En 2019, il s'établit à 3,9 % (cf. tableau 2-D.1). Cette faible proportion montre que **le redressement du nombre de personnes dans les logements non enquêtés n'engendre pas de forte imprécision sur l'estimation de la population**. De plus, parmi les FLNE, seules 26 % sont non renseignées, soit 1 % des logements recensés (cf. tableau 2-D.2). Cela représente 43 000 logements pour l'EAR 2019.

Le taux de FLNE croît avec la taille des communes : de 2 % en moyenne pour les communes de moins de 500 habitants à 6 % pour les communes de plus de 100 000 habitants. De plus, la part de FLNE non renseignées parmi les FLNE est plus importante dans les grandes communes.

Tableau 2-D.1 : Taux de FLNE selon la taille des communes

Tranche (population au RP2016)	Nombre de communes	Taux de FLNE	Nombre de FLNE
<500	3 766	2,0	7 338
500-999	1 340	2,2	8 935
1000-4999	1 545	2,9	40 503
5000-9999	253	4,6	33 163
10000-49999	853	5,2	35 341
50000-99999	99	6,2	14 954
>100000	50	6	20 805
Total	7 906	3,9	161 039

Lecture : le taux de FLNE d'une tranche est défini comme le rapport entre le nombre de FLNE de l'ensemble des communes de la tranche et la somme des résidences principales et des FLNE (RP + FLNE).

Source : EAR 2019.

Tableau 2-D.2 : Part de FLNE non renseignées

Tranche (population au RP2016)	Nombre de FLNE non renseignées	en % des FLNE
Moins de 10 000	17 338	19,3
Plus de 10 000	25 229	35,5
Total Résultat	42 567	26,4

Source : EAR 2019.

D-2. Les différentes étapes du redressement des FLNE non renseignées

Le redressement des FLNE non renseignées se décompose en deux étapes :

- détermination de la part de résidences principales parmi les FLNE non renseignées ;
- imputation du nombre de personnes (en tranche) pour les résidences principales faisant l'objet d'une FLNE non renseignée.

Une fois que le nombre d'habitants est connu pour toutes les FLNE (par la collecte pour les FLNE renseignées ou par imputation pour les FLNE non renseignées), les caractéristiques des logements et des individus sont redressées par imputation¹². Ce redressement n'a toutefois pas d'impact sur le nombre de personnes.

12 Pour en savoir plus : <https://insee.fr/fr/information/2526415>

Détermination de la part de résidences principales parmi les FLNE non renseignées

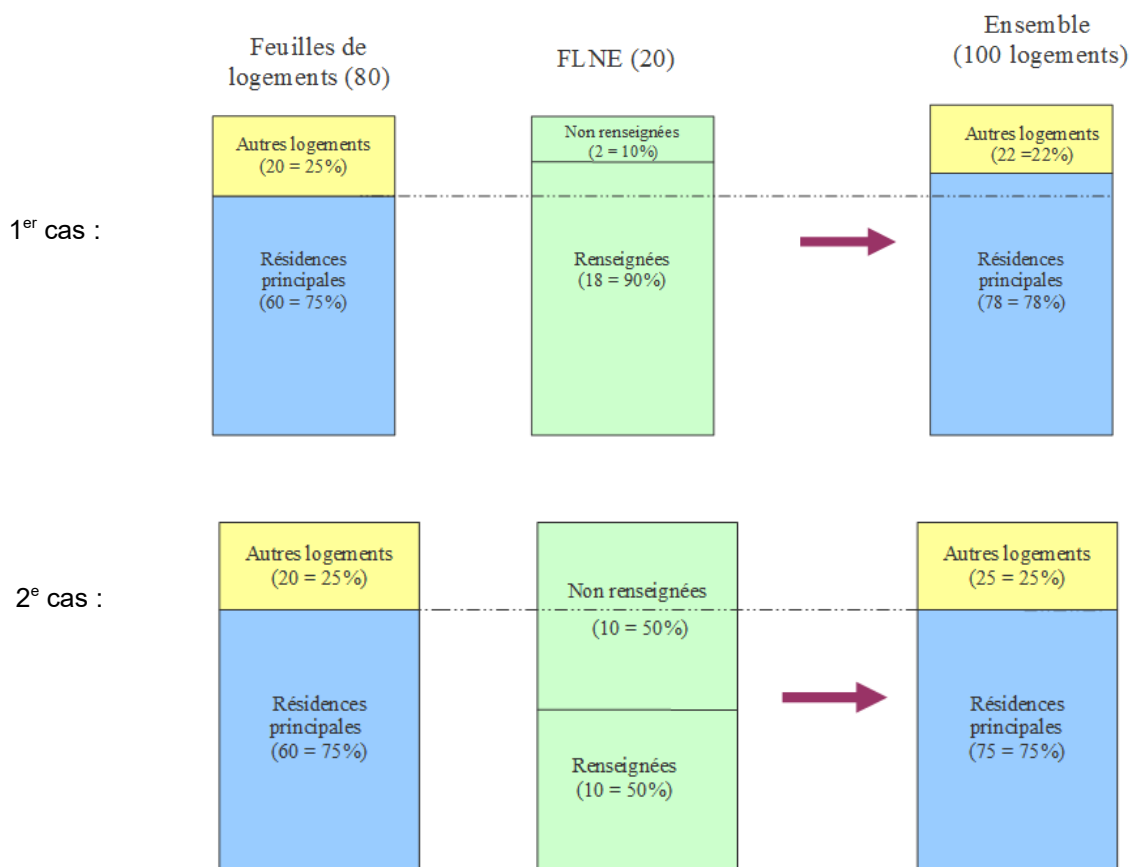
Une FLNE doit normalement être remplie lorsque le logement est une résidence principale non répondante. Néanmoins, dans de nombreux cas, les agents recenseurs ne sont pas en mesure de déterminer si un logement vide est ou non une résidence principale. Une FLNE peut donc dans certains cas signaler qu'un logement n'a pas été enquêté, mais sans savoir réellement s'il s'agit d'une résidence principale. Afin de ne pas surestimer la population, certaines FLNE non renseignées sont considérées *a posteriori* comme des résidences non principales. En pratique, une part de résidences principales parmi les FLNE non renseignées est déterminée pour chaque commune, de manière à ce que, si possible, le taux de résidences principales parmi les FLNE (renseignées ou non) soit égal au taux de résidences principales sur les logements répondants.

Deux cas se rencontrent alors (cf. schéma 2-D.1) :

- 1^{er} cas : le taux de FLNE renseignées est supérieur au taux de résidences principales observé pour les logements répondants de la commune. Dans ce cas, toutes les FLNE non renseignées sont considérées comme des logements non principaux ;

- 2^e cas : le taux de FLNE renseignées est inférieur au taux de résidences principales observé pour les logements répondants. Dans ce cas, un certain nombre de FLNE non renseignées sont considérées comme résidences non principales, à concurrence du taux observé sur les logements répondants. Le reliquat, une fois le taux atteint, constitue des résidences principales.

Schéma 2-D.1 : Détermination de la part de résidences principales parmi les FLNE non renseignées



Les FLNE non renseignées qui ne sont pas classées en résidences principales sont réparties en résidences secondaires, logements occasionnels et logements vacants dans les mêmes proportions que l'ensemble des logements de la commune. Sur l'EAR 2019, 72,3 % des FLNE non renseignées sont considérées comme des résidences principales, soit un peu plus de 30 000 logements (cf. tableau 2-D.3).

Tableau 2-D.3 : Part de résidences principales parmi les FLNE non renseignées

Tranche (population au RP2016)	Nombre de FLNE NR principales	% parmi les FLNE NR
Moins de 10 000	12 219	70,5
Plus de 10 000	18 543	73,5
Total Résultat	30 762	72,3

Note : FLNE NR : fiche de logement non enquêtée, non renseignée.

Source : EAR 2019.

Imputation du nombre de personnes (en tranche) pour les résidences principales faisant l'objet d'une FLNE non renseignée

Pour imputer un nombre d'habitants dans les FLNE non renseignées, on utilise une distribution du nombre de personnes par FLNE renseignée, stratifiée suivant la taille des communes (moins ou plus de 10 000 habitants) et suivant leur localisation (métropole/DOM). On considère que la distribution des individus dans les FLNE non renseignées est plus proche de celle des FLNE renseignées que de celle des individus dans toutes les résidences principales. En effet, les enquêteurs peuvent avoir plus de difficultés à contacter les logements de petite taille. La distribution utilisée a été calculée à partir des FLNE renseignées issues du cumul des EAR 2006-2010 (cf. tableau 2-D.4). Cette distribution n'a que très peu évolué pour les EAR plus récentes: le nombre moyen de personnes par logement dans les FLNE renseignées est passée de 1,69 sur la période 2006-2010 à 1,74 sur 2015-2019 pour les grandes communes de métropole, et de 1,84 à 1,89 pour les petites communes de métropole. Une mise à jour de la distribution pour les imputations sera réalisée pour les prochaines EAR.

Pour les FLNE non renseignées dont le nombre d'habitants imputé est égal à 5 ou plus, le nombre de personnes effectif est déterminé par une procédure d'imputation statistique, le hot deck.

En 2019, le nombre moyen de personnes dans les FLNE non renseignées qui sont considérées comme résidences principales est de 1,7 contre 1,8 dans les FLNE renseignées¹³ et 2,2 dans les résidences principales répondantes.

13 La différence du nombre moyen de personnes par résidence principale entre les FLNE non renseignées et les FLNE renseignées provient du fait que les FLNE sont moins souvent renseignées dans les grandes communes que dans les petites et que le nombre moyen de personnes par résidence principale pour les FLNE est plus faible en grande commune qu'en petite commune.

Tableau 2-D.4 : Distribution du nombre de personnes imputées dans le FLNE non renseignées en fonction de la commune (en %)

	<u>1 ind.</u>	<u>2 ind.</u>	<u>3 ind.</u>	<u>4 ind.</u>	<u>5 ind. ou plus</u>
METRO-GC	61,08	23,02	7,04	5,49	3,37
METRO-PC	50,03	30,76	9,05	7,02	3,14
DOM-GC	40,58	34,7	12,53	8,37	3,82
DOM-PC	40,65	30,39	12,95	9,68	6,33

Note :

PC : communes de moins de 10 000 habitants.

GC : communes de 10 000 habitants ou plus.

Ind. : individu.

Lecture : dans les communes de 10 000 habitants ou plus de métropole, les résidences principales faisant l'objet d'une FLNE non renseignée ont une probabilité de 61,08 % d'avoir un nombre d'habitants imputé égal à 1.

D-3. Résultats sur le RP 2016

Sur le RP 2016, 370 200 personnes (en pondéré) ont été imputées dans les FLNE non renseignées, ce qui représente environ 0,6 % de la population totale. Finalement, même si la correction de la non-réponse totale est réalisée sous certaines hypothèses (part de résidences principales, imputation du nombre de personnes par logement), et donc entourée d'une marge d'incertitude, celle-ci n'a que peu d'influence sur le niveau de la population du fait du faible taux de non-réponse et de la prépondérance des FLNE renseignées.

E) La méthode d'estimation de population en petites communes

Les communes de moins de 10 000 habitants sont recensées exhaustivement, par roulement, tous les cinq ans, à raison d'une commune sur cinq chaque année.

On explicite ci-après la méthode d'estimation de la population des ménages dits ordinaires, c'est-à-dire hors habitations mobiles, sans-abri et bateliers, hors résidant en communauté (foyers de travailleurs, maisons de retraite, résidences universitaires, maisons de détention, internats et militaires vivant en caserne sans leur famille) et hors résidant en hôtel¹⁴. On traite donc ici le cas de 97 % de la population française. Les estimations de population vivant en communauté sont présentées dans l'encadré 2-E.1.

E-1. Présentation de la méthode

En fin d'année N, la population est estimée en référence au 01/01/N-2 pour chacune des communes de moins de 10 000 habitants (cf. schéma 2-E.1).

Pour les communes recensées en N-2, on retient le résultat de l'enquête de recensement.

Pour les autres, il s'agit d'actualiser la dernière population légale de la commune publiée l'année précédente.

Pour celles qui ont été recensées en N-1 ou N, on procède à cette actualisation par **interpolation** avec le chiffre de population recensée lors de l'enquête de N-1 ou N qui est déjà disponible. Cela consiste à calculer l'écart entre le dernier chiffre publié (pour le 01/01/N-3) et celui de l'enquête de recensement et à le répartir à parts égales sur les deux ou trois années d'écart.

Par exemple, fin 2019, l'Insee a diffusé les populations légales de toutes les communes avec comme date de référence le 1^{er} janvier 2017. Pour les communes recensées en 2018 ou 2019, l'interpolation a consisté à mesurer la différence entre la population légale au 1^{er} janvier 2016 et la collecte de 2018 ou 2019 et à retenir respectivement 1/2 ou 1/3 de cette différence. Le résultat obtenu a alors été additionné à la population légale 2016 pour aboutir à la population légale 2017.

Pour celles qui ont été recensées en N-3, on actualise le chiffre tiré de l'enquête de recensement N-3 en s'aidant d'une information supplémentaire sur l'évolution du parc de logements : cette information est tirée du fichier de la taxe d'habitation. On parle alors d'**extrapolation**.

Il faut ensuite relier l'évolution du parc de logements à celle du nombre d'habitants et tenir compte du fait que, souvent, la population et le nombre de ménages ne croissent pas exactement au même rythme. En moyenne et, dans la plupart des communes, la taille moyenne des ménages diminuant, le nombre de ménages augmente plus vite que celui des habitants. Pour passer d'une évolution du parc de logements à une évolution du nombre d'habitants, on estime, à l'aide des dernières enquêtes de recensement disponibles, un coefficient égal au rapport du rythme d'évolution de la population des ménages à celui du nombre de résidences principales. On applique ensuite ce coefficient, dit « de décohabitation », au taux d'évolution du parc calculé sur le fichier de taxe d'habitation pour avoir une estimation du taux de croissance de la population.

Remarque : on fait l'hypothèse implicite que la proportion de résidences principales dans le parc de logements est constante sur la période, car son évolution est en général lente et irrégulière.

Les données de la TH n'étant pas assez robustes pour la Guadeloupe, la Martinique et la Guyane, celles-ci ne sont pas mobilisées pour l'extrapolation. La population est alors estimée par prolongement de l'évolution entre les deux derniers recensements. Il est procédé de même pour les trois communes de France métropolitaine exonérées de taxe d'habitation (Île de Sein, Île de Molène et Suzan).

14 Pour plus d'informations sur les estimations de population, le lecteur pourra se référer à la documentation disponible sur le site internet de l'Insee : <https://insee.fr/fr/information/2553979>

Lorsque les données de la taxe d'habitation ne seront plus disponibles, l'Insee mobilisera d'autres données provenant de l'administration fiscale, qui devraient fournir des informations équivalentes.

Pour les communes recensées en N-4, on procède de la même façon que pour les communes recensées en N-3, en partant du dernier chiffre de population publié, c'est-à-dire celui du 01/01/N-3. Cela correspond à une **extrapolation** sur deux ans du résultat de la dernière enquête de recensement disponible.

Au cours du cycle quinquennal, chaque commune passe donc par les cinq « états » suivants : interpolation sur deux ans, interpolation sur un an, résultat de l'enquête de recensement, extrapolation sur un an, extrapolation sur deux ans. Le chiffre de population est toujours calculé sur la base d'une enquête de recensement distante d'au plus deux ans.

Schéma 2-E.1 : Méthode d'estimation de la population en petite commune

Année de collecte	Année de référence				
	N-4	N-3	N-2	N-1	N
N-4	collecte		extrapolation		
N-3		collecte	extrapolation		
N-2			collecte		
N-1			interpolation	collecte	
N			interpolation		collecte

E-2. Exemple d'application de la méthode

Exemple chiffré :

Considérons la commune de Z, recensée en 2014 et 2019.

Le recensement de 2009 donnait 1 600 habitants et 600 résidences principales.

L'enquête de recensement de 2014 donne 2 000 habitants et 800 résidences principales.

Le fichier de taxe d'habitation donne 900 logements en 2014, 930 en 2015, 950 en 2016, 970 en 2017.

Fin 2016, la population publiée sera celle au 1^{er} janvier 2014 directement issue de l'enquête de recensement 2014, soit 2 000 habitants.

Fin 2017, il s'agit de calculer une population au 01/01/2015.

L'évolution du parc de logements entre le 01/01/2014 et le 01/01/2015 est estimée à : $930/900=1,03333$ (soit : 3,3 %). Entre les deux derniers recensements connus (en l'occurrence 2009 et 2014), le nombre de résidences principales n'a pas crû au même rythme que le nombre d'habitants. Pour le nombre de résidences principales, ce rythme est égal à : $(800/600)^{(1/5)} = 1,05922$ (soit : 5,9 % par an) ; pour le nombre d'habitants, le rythme est égal à $(2000/1600)^{(1/5)} = 1,04564$ (soit : 4,6 % par an). Le différentiel de rythme d'évolution est donc de : $(1,04564 / 1,05922)$, soit : 0,98718. En appliquant ce différentiel, on estime la progression du nombre d'habitants à : $1,03333*0,98718=1,02008$ (soit : 2,0 %).

La population estimée au 01/01/2015 sera donc de 2 000*1,02008, soit 2 040 habitants.

Fin 2018, il s'agit d'actualiser la population au 01/01/2016.

Entre le 01/01/2015 et le 01/01/2016, l'évolution du parc de logements est estimée à : $950/930=1,02151$ (soit : 2,2 %). Le différentiel est toujours de 0,98718. Cela correspond à une évolution du nombre d'habitants de : $1,02151*0,98718=1,00840$ (soit : 0,8 %).

La population au 01/01/2016 est donc estimée à : $2040*1,00840$, soit 2 057 habitants.

Fin 2019, il s'agit d'actualiser la population au 01/01/2017.

On dispose du résultat de l'enquête de recensement de 2019, qui donne une population de 2 200 habitants.

Cela représente une progression de 143 habitants par rapport à 2016. L'interpolation consiste à répartir cette progression à parts égales sur chaque année du 01/01/2016 au 01/01/2019, soit : 2 fois 48 et 1 fois 47 habitants.

La population au 01/01/2017 est donc estimée à : $2\ 057 + 48$ soit : 2 105 habitants.

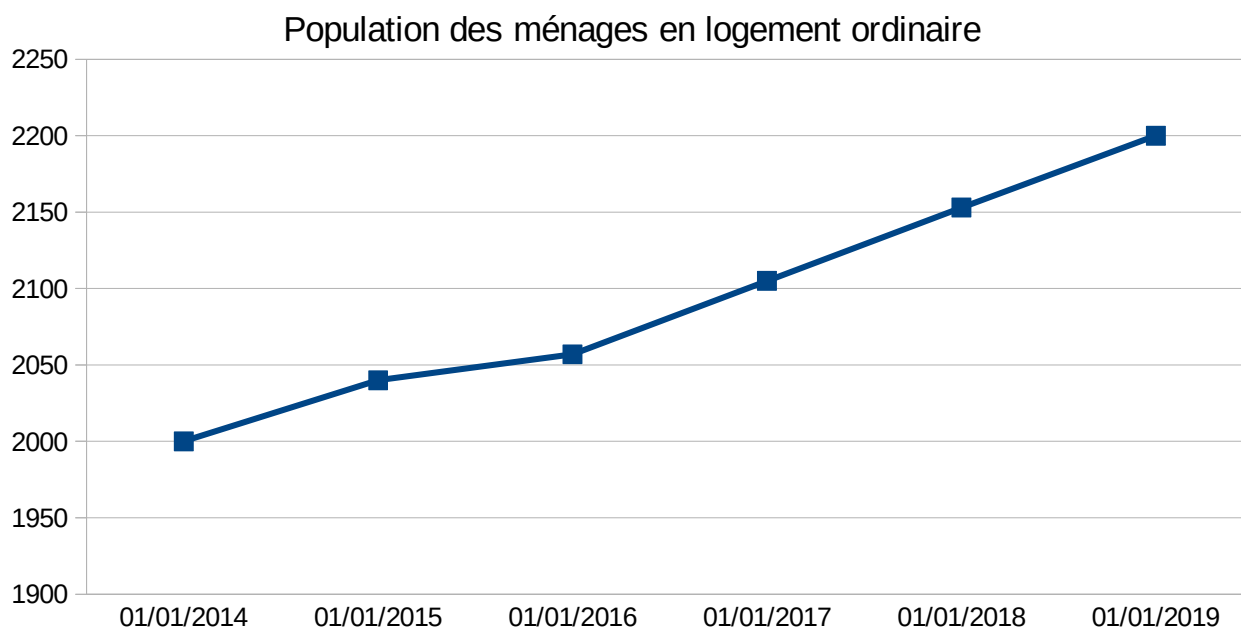
Fin 2020, il s'agit d'actualiser la population au 01/01/2018.

On poursuit l'interpolation, en ajoutant 48 habitants.

La population estimée au 01/01/2018 est donc de 2 153 habitants.

Fin 2021, il s'agit d'actualiser la population au 01/01/2019. On retrouve alors le résultat de la collecte 2019 soit 2 200 habitants.

On retient alors le chiffre du recensement de 2019, soit 2 200 habitants.



E-3. Éléments de discussion de la méthode

L'utilisation du fichier de taxe d'habitation requiert la compatibilité de la donnée auxiliaire issue de ce fichier avec le concept de logement au sens du recensement. L'Insee opère ainsi un filtre sur les locaux présents dans le fichier de taxe d'habitation afin de déterminer ceux qui doivent être pris en compte comme des logements dans le cadre des estimations du recensement (c'est le cas de 75 % des locaux présents dans le fichier ; les 25 % de locaux non considérés comme des logements sont principalement des places de parking ou des garages). Ce filtre, basé sur les variables de gestion de l'impôt, s'avère globalement efficace (particulièrement efficace pour les résidences principales au sens large : en 2017, on décompte environ 29 378 000 résidences principales selon le recensement contre 29 251 000 selon la taxe d'habitation - soit un écart de 0,4%).

Il existe toutefois des différences de concept entre les deux sources : par exemple, un ménage peut déclarer de manière différente sa résidence principale/secondaire au recensement et pour la taxe d'habitation, en lien avec un objectif d'optimisation fiscale notamment. La notion de logement peut différer également pour des structures particulières, notamment à la limite de la définition du logement au sens du recensement (ex :

local du syndicat de copropriété), mais surtout pour les logements vacants qui sont largement surestimés avec la taxe d'habitation.

La source taxe d'habitation présente au moins trois avantages dans le cadre du recensement :

- son exhaustivité
- son caractère régulier (annuel) et pérenne (bien qu'aujourd'hui la fin de cet impôt et donc de la source soit programmée pour 2023)

Les données de flux issues de la taxe d'habitation sont alors très pertinentes (ce qui permet de s'extraire du problème connu relatif au stock trop important de logements vacants).

- son actualité

En effet, au moment où l'on réalise les estimations des populations légales N-2, le fichier de la taxe d'habitation N-2 est lui-même disponible.

On connaît aujourd'hui trois faiblesses à la méthode actuelle d'estimation :

- on utilise une évolution auxiliaire basée sur l'ensemble des logements pour estimer une évolution de la population

La méthode d'extrapolation repose sur l'évolution auxiliaire basée sur l'ensemble du parc de logements. Il serait en réalité plus pertinent de retenir l'évolution basée sur les résidences principales. En effet, l'évolution du parc total de logements ne reflète pas nécessairement l'évolution des logements qui portent la population. Le coefficient de décohabitation, quant à lui, compare l'évolution de la population par rapport à celles des seules résidences principales. Une augmentation du nombre de logements vacants conduit donc à un accroissement de la population à coefficient de décohabitation constant.

- le retard de prise en compte d'un événement par la taxe d'habitation (dû au délai de déclaration)

On peut ainsi avoir des décalages entre la collecte et la taxe d'habitation l'année de collecte, ou avoir un retard de prise en compte d'une démolition ou d'une construction lorsque celle-ci intervient au quatrième trimestre de l'année. Une expertise des données de la taxe d'habitation est réalisée afin d'éviter d'éventuels doubles comptes (de création ou de disparition de logements) occasionnés par ces décalages de prise en compte des événements.

- le coefficient permettant de prendre en compte l'aspect de décohabitation repose sur une hypothèse relativement forte

Il suppose que la structure du parc de logements est la même sur une période de sept ans à compter de la collecte antérieure. L'émergence de programmes immobiliers vient parfois contredire cette hypothèse, le calcul d'estimation peut alors être faussé. L'expertise de ce coefficient permet d'identifier et de corriger les cas les plus flagrants. En outre, l'utilisation du coefficient de décohabitation n'est pas adaptée aux communes de très petite taille. En effet, pour celles-ci, les trop petits nombres de logements et d'habitants ne permettent pas d'obtenir un coefficient de décohabitation pertinent, et l'hypothèse sur laquelle repose son utilisation n'est alors pas validée. Seule l'évolution du nombre de logements à la TH est utilisée pour ces communes..

Encadré 2-E.1 : Estimation de la population vivant en communauté

Estimations de population vivant dans des chambres en communauté

Les estimations de population sont effectuées communauté par communauté. Elles sont réalisées en tenant compte de la situation de la communauté (ouverte ou fermée) au 1^{er} janvier de l'année.

Pour l'année de la collecte (N), la population des communautés est estimée à hauteur du nombre de bulletins individuels de communauté (BIC) collectés.

Durant les deux années qui suivent (N+1 et N+2), les seules informations dont dispose l'Insee concernent les ouvertures et fermetures de communautés. Ainsi, durant cette période :

- si la communauté est toujours active, le nombre de BIC collectés est reporté à l'identique ;
- si la communauté ferme, la population de la communauté est ramenée à zéro ;
- si une nouvelle communauté est créée, on estime sa population à hauteur de sa capacité d'accueil (nombre de lits).

À partir des estimations de la population N+3, on dispose d'une nouvelle collecte (réalisée en N+5), en raison du décalage de trois ans entre le recensement et la diffusion des populations légales. Pour estimer les populations N+3 et N+4, on effectue une interpolation linéaire entre le point N+2 et la nouvelle collecte (N+5), comme pour les populations vivant en logement ordinaire dans les petites communes.

Il en va de même en cas de fermeture (interpolation entre le point N+2 et zéro) et d'ouverture (interpolation entre zéro et le point N+5).

Estimations de population des logements de fonction¹⁵ en période intercensitaire

En commune de plus de 10 000 habitants, la population des logements de fonction est estimée à hauteur du nombre de bulletins individuels de la collecte la plus récente. Ce nombre est maintenu 5 ans.

En commune de moins de 10 000 habitants, la population des logements de fonction est estimée avec la population des logements ordinaires : l'extrapolation avec la taxe d'habitation et l'interpolation linéaire sont appliquées à l'ensemble des logements de la commune. La ventilation entre les deux types de populations est faite au *pro rata* de la collecte la plus récente.

Les procédures de calculs exceptionnelles

Si l'Insee a connaissance de très fortes évolutions dans des établissements pénitentiaires (ouverture ou fermeture de prison, transferts des détenus dans une autre commune) ou dans les casernes militaires (restructurations décidées par le gouvernement, transferts de troupes), il les prend en compte à la date exacte de l'événement et directement avec l'effectif connu à cette date (en interrogeant le responsable de la structure), quelle que soit la phase d'estimation dans laquelle se trouve la communauté (déterminée par son groupe de rotation) afin de ne pas léser les communes concernées par un afflux de population. Il s'agit de calculs dérogatoires et ils ne s'appliquent qu'à ces catégories de communautés.

15 La définition d'un logement de fonction au sens du recensement est présentée dans l'annexe 1.

F) Les estimations de population en grandes communes : éléments sur la précision

Dans les communes de 10 000 habitants ou plus de métropole, le recensement des ménages est réalisé par sondage sur un échantillon d'environ 40 % des logements sur cinq ans. La base de sondage pour la population des ménages est constituée à partir du Répertoire d'immeubles localisés (RIL, cf. partie 2-A).

On explicite ci-après la méthode d'estimation de la population des ménages dits ordinaires, c'est-à-dire hors habitations mobiles, sans-abri et bateliers, et hors résidant en communauté (foyers de travailleurs, maisons de retraite, résidences universitaires, maisons de détention, internats et militaires vivant en caserne sans leur famille). On traite donc ici le cas de 97 % de la population française. Les estimations de population vivant en communauté sont présentées dans l'encadré 2-E.1.

F-1. Présentation de la méthode

Pour les communes de 10 000 habitants ou plus de métropole

Tirage de l'échantillon

Les adresses de chaque commune de 10 000 habitants ou plus sont réparties en cinq groupes. Les cinq groupes d'adresses appelés « groupes de rotation » ont été initialisés en amont de la première enquête annuelle de 2004 par un tirage équilibré donnant l'image la plus fidèle possible de chaque Iris de la commune en matière de population par âge, sexe, type de logement (individuel ou collectif) et nombre de logements. Ils sont depuis mis à jour chaque année pour intégrer notamment les créations et les disparitions d'adresses. Chaque année, un groupe de rotation est concerné par le recensement.

Chaque année, les adresses appartenant au groupe de rotation concerné par la prochaine enquête de recensement font l'objet d'un tirage stratifié, afin d'accroître la précision des estimations. Les adresses sont donc réparties en quatre strates de tirage : les résidences touristiques, les adresses de grande taille, les adresses nouvelles et les autres adresses :

- Les **résidences touristiques** correspondent aux hôtels, campings et résidences hôtelières. Toutes les résidences touristiques du groupe de rotation sont enquêtées (strate exhaustive).
- Les **adresses de grande taille** (ou « grandes adresses ») sont celles dont le nombre de logements dépasse un certain seuil, propre à chaque commune : ce sont les adresses d'au moins 60 logements chacune et cumulant au maximum 10 % des logements de la commune. Toutes les grandes adresses du groupe de rotation sont enquêtées (strate exhaustive).
- Les **adresses neuves** (ou « adresses nouvelles ») sont celles qui sont apparues au cours des 5 dernières années et qui ont été affectées au groupe de rotation de l'année. Toutes les adresses neuves du groupe de rotation sont enquêtées (strate exhaustive).
- Les **autres adresses** (ou « petites adresses connues ») font l'objet d'un sondage équilibré en termes de nombre de logements, de nombre de logements collectifs et de nombre de logements par IRIS¹⁶ de la commune. Le taux de sondage est déterminé en fonction du nombre de logements dans la strate des grandes adresses et des adresses neuves, de sorte que le taux de sondage sur l'ensemble des strates hors résidences touristiques soit de 40 % dans le groupe de rotation considéré.

Estimations de population

En grande commune de métropole, **la population des ménages en logement ordinaire** de l'année N est calculée au niveau de chacun des Iris (éventuellement regroupés) en multipliant le nombre de logements au

16 Îlots regroupés pour l'information statistique.

1^{er} janvier de l'année N issu du RIL par le nombre moyen de personnes par logement estimé à partir des EAR :

- Le **nombre de logements au 1^{er} janvier de l'année N** est obtenu en faisant la moyenne du nombre de logements du répertoire des immeubles localisés (RIL) de juillet N-1 et de juillet N. Cette composante n'est pas issue de la collecte du recensement et n'est donc pas affectée par l'aléa de sondage. Il est ainsi essentiel que le RIL soit exhaustif pour que le niveau de la population soit correctement estimé.
- Le **nombre moyen de personnes par logement** est obtenu à partir des données collectées lors des cinq enquêtes annuelles de recensement de N-2 à N+2. C'est cette composante qui est affectée par l'aléa de sondage. Dans le cadre des estimations de population, la variable d'intérêt est donc le nombre moyen de personnes par logement et non directement le nombre de personnes. Cette variable d'intérêt peut se décomposer en deux facteurs : le nombre de personnes par résidence principale et la part de résidences principales.

Pour les communes de 10 000 habitants ou plus des DOM

Dans les DOM¹⁷, les RIL ne sont pas mis à jour en continu sur les 5 groupes de rotation, mais plus spécifiquement sur le groupe de rotation concerné par la prochaine enquête de recensement (cf. partie 2-A). À ce titre, le RIL ne peut pas fournir une estimation précise du nombre de logements sur l'ensemble de la commune. La méthodologie de tirage et d'estimation dans les grandes communes des DOM diffère sensiblement de la métropole.

Tirage de l'échantillon

La base de sondage des EAR dans les DOM est constituée à partir du RIL qui est mis à jour sur le groupe de rotation de l'année par des enquêtes terrain (enquête cartographique notamment) et par les communes (cf. partie 2-A). Dans les grandes communes des DOM, les groupes de rotation sont affectés aux îlots, et toutes les adresses d'un îlot ont alors le même groupe de rotation. Un échantillon de 40 % d'adresses est alors tiré dans cette base de sondage. À partir de l'EAR 2020, la stratification utilisée pour ce tirage est très proche de celle mobilisée en métropole, sinon que le seuil pour les grandes adresses est fixé à 10 logements. Auparavant, il n'y avait pas de strate d'adresses neuves comme en métropole.

Estimations de population

Pour les grandes communes des DOM, chaque EAR permet d'estimer la population de l'ensemble des îlots du groupe de rotation considéré. La population des ménages en logement ordinaire de l'année N est alors calculée en sommant les estimations de population pour chaque groupe de rotation obtenues à partir des données collectées lors des cinq enquêtes annuelles de recensement de N-2 à N+2. On fait ainsi l'hypothèse que la population ainsi obtenue est une bonne estimation de la population de l'année médiane N du cycle d'estimation, c'est-à-dire que les variations de populations entre N et N+1 ou N+2 des îlots recensés en N+1 ou N+2 compensent les variations de populations entre N-2 ou N-1 et N des îlots recensés en N-2 ou N-1.

F-2. Éléments sur la précision

La qualité des estimations de population en grande commune dépend de multiples facteurs et en premier lieu de la qualité du RIL et de celle de la collecte. Les inputs des estimations de population bénéficient donc de tous les traitements antérieurs développés dans les parties précédentes : mises-à-jour du RIL, saisie, contrôles de collecte, contrôles post-collecte et correction de la non-réponse. **Cette partie s'attache plus particulièrement à l'analyse de la variance des estimations de population consécutives à l'application du sondage.**

17 Le département de Mayotte n'a pas encore intégré le processus des enquêtes annuelles de recensement. La première EAR à Mayotte est prévue pour 2021.

Estimation de la précision

En grande commune, l'estimation de la population par sondage entraîne une marge d'incertitude sur les résultats (Insee, 2017). Cette marge est mesurée par le **coefficient de variation**, noté **CV**. Il renseigne sur l'écart moyen entre la valeur estimée par le recensement et la « vraie » valeur. En termes statistiques, il correspond au rapport de l'écart-type à la moyenne. Ce coefficient de variation permet de construire un intervalle de confiance de l'estimation. La vraie valeur est comprise avec 95 % de chances dans la plage de valeurs possibles suivante :

[valeur au recensement x (1 - 2CV) ; valeur au recensement x (1 + 2CV)]

Cette imprécision varie selon l'effectif de la commune (plus l'effectif obtenu est réduit, plus l'imprécision relative risque d'être grande, car l'estimation de cet effectif repose sur peu d'observations).

Ce coefficient de variation a pu être appréhendé par estimation sur des cycles de cinq ans du recensement renouvelé. Réalisés depuis la publication des résultats du RP 2006, ces calculs tiennent compte des gains en précision liés au tirage équilibré des adresses au sein des groupes de rotation (Deville, 2000) ainsi que de la méthode d'estimation faisant intervenir le RIL médian.

Ces estimations se limitent aux communes de 10 000 habitants ou plus de métropole.

Principaux résultats

D'après les calculs d'estimation de la variance issus du recensement de 2006, l'incertitude introduite par le sondage dans les grandes communes conduit à une imprécision de 0,02 % au niveau national. Par conséquent, au niveau de l'ensemble de la population française, l'imprécision liée au sondage est de plus ou moins 15 800 personnes. Ce chiffre illustre le fait que, au niveau national, l'impact du sondage sur le résultat final est négligeable. Ces estimations restent valables dans la mesure où la méthode de tirage de l'échantillon et la méthode d'estimation des populations n'ont été modifiées qu'à la marge depuis 2006.

Le tableau 2-F.1, indique la distribution du coefficient de variation associé à la variable population aux niveaux régional (anciennes régions), départemental et communal.

Tableau 2-F.1 : Distribution des coefficients de variation (en %) de la variable population aux niveaux régional, départemental, communal

	Région	Département	Commune
Maximum	0,45	1,15	6,62
99 %	0,45	1,15	2,07
95 %	0,26	0,94	1,42
90 %	0,25	0,70	1,22
75 % Q3	0,20	0,53	1,05
50 % Médiane	0,16	0,35	0,88
25 % Q1	0,12	0,25	0,71
10 %	0,09	0,15	0,55
5 %	0,05	0,13	0,46
1 %	0,03	0,03	0,32
Minimum	0,03	0,03	0,23
Moyenne	0,17	0,40	0,91

Champ : communes de 10 000 habitants ou plus de France métropolitaine.
Source : recensement de la population 2006.

Le coefficient de variation diffère selon les régions. Pour la moitié des régions, il est inférieur à 0,16 %. Pour un quart d'entre elles, il est inférieur à 0,12 %. Seules un quart des régions (les moins peuplées) ont un CV supérieur à 0,20 %.

L'imprécision introduite par le sondage sur la variable population diffère de façon plus importante selon la commune considérée.

Pour la moitié des grandes communes, le CV est inférieur à 0,88 %. Pour un quart d'entre elles, il est inférieur à 0,71 %. Seules un quart des grandes communes ont un CV supérieur à 1,05 %.

D'une part, l'imprécision varie avec la taille de la commune (cf. tableau 2-F.2, tiré de Brillhault, 2016) : exprimée en pourcentage de la population, elle est plus faible dans les plus grandes communes. À taille identique, d'autres facteurs expliquent une précision différente selon la commune : part de grandes adresses, part d'adresses nouvelles, etc.

Tableau 2-F.2 : Distribution des coefficients de variation (en %) de la population des ménages par tranche de taille de commune

	10 000 – 19 999 habitants (452 communes)	20 000 – 49 999 habitants (318 communes)	50 000 – 99 999 habitants (79 communes)	Au moins 100 000 habitants (43 communes)
75 % Q3	1,16	0,87	0,71	0,43
50 % Médiane	1,02	0,78	0,56	0,39
25 % Q1	0,90	0,68	0,50	0,34

Champ : communes de 10 000 habitants ou plus de France métropolitaine.

Source : recensement de la population de 2006.

Pour la moitié des communes de 10 000 à 19 999 habitants, le coefficient de variation est inférieur à 1,02 %. Pour un quart d'entre elles, il est supérieur à 1,16 %, alors que pour le quart des communes où ce coefficient est le plus faible, il est inférieur à 0,90 %.

G) Le processus de validation des estimations de population communale

Des premières estimations sont établies dès le mois de juillet avant même la fin des contrôles post-collecte et de la saisie optique des questionnaires de l'enquête annuelle de recensement du début d'année. Plusieurs contrôles sont réalisés sur ces premières estimations. Ces contrôles permettent d'assurer la qualité des données produites, par la détection des éventuelles erreurs pouvant intervenir au cours des différentes phases du processus de production (collecte, RIL, estimations des populations, etc.). Ces contrôles sont de nouveau réalisés sur les estimations produites à la fin de l'année, qui prennent en compte l'ensemble des contrôles et redressements effectués après la collecte. Les populations sont ensuite diffusées sur le site insee.fr et authentifiées par un décret.

G-1. Les contrôles automatiques de cohérence et de vraisemblance

Après chaque production d'estimations de population – même provisoire –, une batterie de contrôles automatiques est mise en œuvre sur les données. Ces contrôles (de cohérence et de vraisemblance) ont deux objectifs principaux. D'une part, vérifier que les données produites sont bien cohérentes entre elles, en testant par exemple que la population municipale est bien la somme de la population des ménages, de la population vivant en communauté, de la population des habitations mobiles et des sans-abris. D'autre part, des contrôles de vraisemblance visent à détecter des évolutions impossibles de certaines catégories de population, ou sur certaines zones géographiques. L'idée de ces contrôles n'est pas de sortir une liste de communes présentant une évolution importante de la population (cf. partie 2-G-2 sur l'expertise menée par les directions régionales de l'Insee), mais de détecter des évolutions qui ne sont pas crédibles. L'idée est de passer un contrôle automatique qui puisse être analysé rapidement (et non de regarder des cas particuliers) avant la mise à disposition des données aux directions régionales de l'Insee pour une expertise plus poussée. Lors de cette phase, on contrôle, par exemple, que la taille moyenne des ménages de chaque commune est comprise entre 1 et 6¹⁸.

Ces contrôles sont récents et ont vocation à être enrichis au fil du temps pour consolider la production des estimations de population.

G-2. L'expertise des populations menées par les directions régionales de l'Insee

Les estimations de population sont transmises aux directions régionales de l'Insee afin qu'elles les expertisent finement. Contrairement aux contrôles automatiques précédents, l'objectif est d'analyser au cas par cas les estimations de population, afin de repérer les évolutions atypiques par rapport au contexte de chaque commune et corriger d'éventuelles erreurs.

Pour faciliter l'analyse, une application informatique interne à l'Insee (Apolline) a été créée en 2015. Outre les estimations de population, cette application contient de nombreuses données de contexte pour chaque commune : données historiques de population de la commune, données sur les nombres de logements, résidences principales, nombre moyen de personnes par logement, population des communautés, nombre d'habitations mobiles et sans-abris, etc. Ces données proviennent de la taxe d'habitation (TH) ou des EAR. Les données sont présentées sous forme de tableaux, de graphiques ou de cartes.

Comme il n'est pas possible d'analyser au cas par cas la situation des 35 000 communes, des outils d'aide à la priorisation ont été établis. Une variable (appelée "score") évalue notamment si l'évolution de la population présente un changement de tendance récent (par exemple, une croissance suivie d'une baisse). Les directions régionales sont alors invitées à analyser de manière plus approfondie les communes qui présentent une variable "score" élevée. Les grandes communes dont l'évolution de la population n'est pas

18 Une taille moyenne des ménages de 6 personnes est volontairement élevée, car comme expliqué dans le texte, l'objectif est de détecter des résultats aberrants, qui ne peuvent provenir que d'une erreur importante.

comprise dans l'intervalle de confiance de la population de l'année précédente sont également identifiées. Des variables de contribution ont été créées afin de mettre en évidence les composantes de la population qui contribuent le plus à l'évolution de la population de chaque commune.

L'expertise menée par les directions régionales est indispensable à plusieurs points. Elle permet en premier lieu de repérer et d'éventuellement corriger des problèmes résiduels intervenus lors des différentes phases amont du processus de production des estimations de population : problèmes de collecte qui n'ont pas été traités lors des phases de contrôles de collecte ou post-collecte, erreurs dans le RIL d'une commune, effet d'un transfert entre le champ des communautés et le champ des ménages ayant un effet sur les estimations de population, mauvaise prise en compte de l'ouverture ou de la fermeture d'une communauté, etc. Elle permet par ailleurs de valider les estimations de population qui n'entrent pas dans le champ de la procédure classique d'estimation. Les communes concernées par ces estimations spécifiques sont les communes en changement de géographie (principalement les fusions de communes) et les communes qui franchissent le seuil des 10 000 habitants – à la hausse ou à la baisse – qui pendant plusieurs années sont dans un régime transitoire entre une estimation en tant que petite commune et une estimation en tant que grande commune. Ces estimations spécifiques plus complexes et ne pouvant pas être automatisées facilement nécessitent une intervention humaine plus conséquente. Il convient donc d'y apporter une attention particulière lors de l'expertise.

L'expertise des estimations en grande commune permet également de détecter des adresses recensées qui influencent fortement les estimations de population. En effet, du fait du tirage de l'échantillon, certaines adresses atypiques (en matière de nombre de logements ou de nombre moyen de personnes par logement) peuvent avoir un effet significatif sur la mesure de la population selon qu'elle est tirée ou non dans l'échantillon de l'EAR. Des procédures de winsorisation peuvent alors être mises en œuvre afin de réduire l'influence de ces adresses et la volatilité des estimations.

G-3. L'analyse des estimations au niveau agrégé

Des contrôles de vraisemblance sont réalisés à des niveaux plus agrégés, par exemple au niveau départemental, régional ou national. Ces contrôles sont plus faciles que ceux au niveau communal, car les populations y évoluent de manière beaucoup plus régulière. Il est donc plus aisé d'y déceler une éventuelle erreur. Les contrôles effectués portent principalement sur les différentes composantes de la population (population en logement ordinaire, population des communautés, population des habitations mobiles et des sans-abris), à différents niveaux géographiques (niveau national, ensemble des grandes communes, ensemble des petites communes par groupe de rotation, etc.).