

Insee

Méthodes

N° 136

Octobre 2020

La qualité des estimations de population dans le recensement

Insee Méthodes

La qualité des estimations de population dans le recensement

Ont participé à cet ouvrage : Martin Chevalier, Sébastien Durier, Lionel Espinasse, Vincent Le Palud, Stéfan Lollivier, Benjamin Marteau, Jeanne Pagès, Julie Prévot, Didier Reynaud, Valérie Roux, Gwennaël Solard, Julien Talbot, Laurent Toulemon, Lucile Vanotti, Sabrina Volant

Coordination : Gwennaël Solard

Avant-propos

Jusqu'à la fin du XX^e siècle, la France a réalisé des recensements traditionnels espacés d'une petite dizaine d'années. L'ensemble des personnes résidant en France étaient alors recensées en quelques semaines. Depuis le début des années 2000, l'Insee a modifié la méthode de recensement. Le recensement est à présent annuel et tournant. Chaque année, il concerne 5 millions de logements et 9 millions de personnes. La première enquête annuelle de recensement a eu lieu en 2004.

Cette nouvelle méthode de recensement présente deux avantages majeurs par rapport à un recensement traditionnel. Le premier avantage est de disposer de données localisées actualisées tous les ans, alors qu'il fallait attendre près de 10 ans entre les résultats de deux recensements traditionnels consécutifs. Cette amélioration de la fraîcheur des résultats était demandée par les nombreux utilisateurs et était indispensable pour suivre au mieux les évolutions toujours plus rapides de notre société et aider à la décision publique. Le deuxième avantage est le lissage des coûts du recensement. Alors qu'il est de plus en plus difficile de disposer d'un budget conséquent, la nouvelle méthode de recensement permet de lisser la charge financière, sans augmenter la facture globale. Ce lissage est aussi effectif sur les moyens humains mobilisés pour les recensements. Dans les communes de plus de 10 000 habitants, où l'enquête a lieu tous les ans sur un échantillon de logements, les mêmes acteurs peuvent ainsi participer tous les ans au recensement, ce qui permet une professionnalisation accrue.

Malgré ses avantages, la méthode française reste unique dans le monde. De fait, au niveau européen, la France est entourée de pays qui passent progressivement de recensements traditionnels à des recensements combinés mobilisant de plus en plus des données administratives. Le recensement français reste donc une exception.

Dès la mise en place du nouveau recensement, l'Insee a publié des documents de référence sur les modalités de mise en œuvre de la nouvelle formule. D'un point de vue méthodologique, les travaux ont surtout porté sur les calculs de précision des estimations de population dans les communes de plus de 10 000 habitants, associés à la mise en place d'un sondage qui constituait la principale nouveauté de la méthode. Aujourd'hui, plus de 15 ans après la mise en place de cette méthode, certains partenaires sollicitent l'Insee pour disposer de davantage de documentation sur la mesure plus globale de la qualité du recensement français.

Ce document présente de manière détaillée les méthodes utilisées à ce jour pour produire les estimations de population, chaque année. Ces méthodes sont susceptibles d'évoluer à l'avenir pour tenir compte de nouvelles opportunités ou contraintes externes ainsi que d'éventuelles avancées méthodologiques. La qualité de l'information produite à chaque phase du processus y est analysée. Il s'adresse tant aux personnes spécialisées dans les travaux démographiques (chercheurs, agences d'urbanisme, etc.) qu'à des acteurs locaux.

Ce document constitue une première brique dans l'amélioration de la documentation du recensement, en détaillant le processus de production des populations légales, les opérations qualité menées à chaque étape et les comparaisons avec d'autres sources qui ont été réalisées. D'autres documents, au stade de projet, viseront à analyser les autres productions du recensement, comme les résultats statistiques (caractéristiques des logements et des habitants), ou des modifications du protocole du recensement (introduction de la réponse par internet, tirage de l'échantillon de l'enquête annuelle de recensement à Mayotte par exemple).

Christel Colin
Directrice des statistiques
démographiques et sociales

La qualité des estimations de population dans le recensement

Table des matières

Introduction.....	4
Partie 1 : Présentation du recensement de la population.....	5
A) Objectifs et fondements juridiques du recensement.....	5
B) Méthode de recensement.....	6
C) Déroulement de la collecte.....	8
D) Estimations de population.....	9
Partie 2 : De la confection du RIL à l'élaboration des populations : des contrôles de la qualité tout au long du processus.....	10
A) La qualité de la base de sondage en grande commune : le répertoire des immeubles localisés.....	11
B) L'exhaustivité de la collecte.....	22
C) Les contrôles de l'enquête.....	30
D) Le traitement de la non-réponse.....	35
E) La méthode d'estimation de population en petites communes.....	40
F) Les estimations de population en grandes communes : éléments sur la précision.....	45
G) Le processus de validation des estimations de population communale.....	49
Partie 3 : Des analyses quantitatives de la qualité des estimations de population.....	51
A) Estimation des doubles comptes statistiques au recensement à partir de l'Échantillon démographique permanent.....	52
B) Omissions possibles de jeunes enfants au recensement : que nous apprennent les comparaisons avec l'état civil ?.....	66
C) Comparaison entre le recensement de la population et les données fiscales : Quelles différences ? Quels enseignements ?.....	75
Conclusion.....	86
Bibliographie.....	88
Annexe 1 : définition d'une communauté et gestion du répertoire.....	90
Annexe 2 : une méthode possible pour un correctif macro de la pyramide des âges.....	92
Annexe 3 : principaux textes juridiques encadrant le recensement de la population.....	99

Introduction

La méthode de recensement de la population a évolué en France au début des années 2000. Après un dernier recensement exhaustif du territoire effectué en 1999, l'Insee a mis en place un recensement tournant qui consiste en la réalisation d'une enquête annuelle de recensement sur une partie du territoire. La diffusion des résultats du recensement repose sur 5 enquêtes annuelles.

Cette nouvelle méthode de recensement a fait l'objet au démarrage d'une documentation détaillée (*cf.* par exemple Godinot, 2015). L'Insee ne disposait à l'époque que de peu d'éléments sur la qualité des résultats produits. L'Insee a néanmoins publié en décembre 2008 un document général sur la qualité du recensement de la population (Cézar, 2008). Celui-ci documentait notamment l'adéquation aux critères internationaux de qualité (ONU et Eurostat) et présentait la façon dont la qualité était gérée tout au long du processus. Depuis, l'Insee a publié à plusieurs reprises des travaux sur la précision des estimations pour les communes de 10 000 habitants ou plus (Brilhaut, 2016). Le recensement par sondage constituait en effet la principale nouveauté du recensement tournant. Pour autant, la qualité des estimations ne s'arrête pas à ces communes, ni à la seule variance des estimations du fait des aléas de sondage.

Le Code de bonnes pratiques de la statistique européenne (Eurostat, 2017) constitue « la pierre angulaire du cadre qualité commun du système statistique européen ». Il définit 16 principes portant sur l'environnement institutionnel, les processus statistiques et les résultats statistiques. Ce code porte donc autant sur la qualité des résultats produits (pertinence, exactitude, actualité, accessibilité, etc.) que sur la manière de les obtenir (méthodologie adaptée, maîtrise de la charge pour les déclarants, rapport coût-efficacité).

Cet Insee Méthodes n'a pas l'ambition d'étudier la qualité du recensement au regard de l'ensemble des principes définis par le Code de bonnes pratiques de la statistique européenne. Il vise à documenter la robustesse, la précision et l'exactitude des résultats produits à partir du recensement au travers de la description du processus mis en œuvre, des contrôles qualité réalisés au cours de ce processus et de la comparaison avec d'autres sources de données. Il questionne ainsi 3 principes : le principe d'exactitude et de fiabilité (principe 12), le principe de méthodologie solide (principe 7) et, dans une moindre mesure, le principe de procédures statistiques adaptées (principe 8).

Le recensement de la population vise bien évidemment à être en adéquation avec l'ensemble des 16 principes. Par exemple, le recensement a évolué à deux reprises ces dernières années afin de mieux répondre aux attentes des utilisations (principe 11 de pertinence). Des modifications du questionnaire ont ainsi eu lieu en 2015 et 2018, en suivant des préconisations du Conseil national de l'information statistique exprimées en 2012 dans un rapport sur l'évolution du questionnaire du recensement de la population (Cases, 2012).

Le champ de cet Insee Méthodes est également restreint aux seules populations municipales légales afin de ne pas complexifier le document. Des publications ultérieures pourront aborder d'autres aspects de la qualité ou d'autres champs d'analyse, comme la qualité des estimations des résultats statistiques issus du recensement.

Après une présentation générale du recensement de la population (partie 1), le processus de production des estimations de population est détaillé en mettant en avant les contrôles de qualité qui sont effectués à chaque phase (partie 2). Ces contrôles sont de différents types : contrôles de vraisemblance, contrôles de cohérence, etc. Les méthodes de redressement des données y sont également développées. Des indicateurs de qualité sont également établis pour différentes phases du processus (base de sondage, collecte, redressements, etc.). Au-delà de ces éléments internes à la production du recensement, la partie 3 porte sur des comparaisons avec d'autres sources (Échantillon démographique permanent, données d'état civil et données fiscales). Ces comparaisons permettent de mettre en évidence d'éventuelles sur-estimations et des sous-estimations pour certaines populations.

Partie 1 : Présentation du recensement de la population

A) Objectifs et fondements juridiques du recensement

Le recensement de la population en France a évolué au début des années 2000. Le dernier recensement exhaustif a eu lieu en 1999. Le recensement s'effectue depuis 2004 par des enquêtes annuelles. Les objectifs du recensement sont toutefois inchangés. Il vise toujours à dénombrer la population et à décrire les caractéristiques des personnes et des logements, à un niveau territorial très fin. De nombreux utilisateurs ont besoin de ces informations à des fins très diverses : acteurs locaux, décideurs publics, chercheurs, entreprises, particuliers, etc.

Les grands principes du recensement de la population sont énoncés dans la loi n° 2002-276 du 27 février 2002 relative à la démocratie de proximité, à l'article 156 pour ce qui concerne la métropole et les départements d'outre-mer (cf. annexe 3). Cet article précise notamment que le recensement de la population est réalisé sous la responsabilité et le contrôle de l'Insee tandis que les communes ou les établissements de coopération intercommunale (EPCI) sont responsables de la préparation et de la réalisation des enquêtes de recensement. La loi indique également dans les grandes lignes la méthode de recensement, qui diffère selon la taille de la commune (cf. partie 1-B). Elle précise enfin qu'un décret authentifie chaque année les chiffres de population. Ces populations servent de référence à de nombreux textes réglementaires (plus de 350), comme ceux relatifs aux dotations de l'État aux communes.

Les modalités d'application de la loi de 2002 sont définies, pour ce qui concerne le recensement, dans le décret n° 2003-485 du 5 juin 2003 modifié relatif au recensement de la population. Y sont définies notamment les différentes catégories de population (cf. encadré 1-A.1). Ce décret précise également les modalités des enquêtes annuelles de recensement (EAR) et les dispositions financières (dotation forfaitaire de recensement) pour la réalisation de ces enquêtes.

Le règlement général 2016/679 du 27 avril 2016 sur la protection des données (RGPD) ainsi que la loi n°78-17 du 6 janvier 1978 modifiée, relative à l'informatique, aux fichiers et aux libertés, s'appliquent à l'enquête de recensement. Les droits d'accès, de rectification ou de limitation de traitement peuvent être exercés auprès des directions régionales de l'Insee. Le recensement de la population est aussi encadré par la loi sur l'obligation, la coordination et le secret en matière de statistiques du 7 juin 1951, s'agissant d'une enquête statistique et de la confidentialité des données,

Plusieurs autres textes réglementaires (décrets, arrêtés) sont relatifs au recensement, mais sont moins structurants pour ce document. Ils ne sont donc pas présentés dans cette partie.

Encadré 1-A.1 : Les différentes catégories de population

Le décret n° 2003-485 du 5 juin 2003 fixe les catégories de population et leur composition.

La **population municipale** comprend les personnes ayant leur résidence habituelle sur le territoire de la commune, dans un logement ou une communauté, les personnes détenues dans les établissements pénitentiaires de la commune, les personnes sans abri recensées sur le territoire de la commune et les personnes résidant habituellement dans des habitations mobiles, recensées sur le territoire de la commune. L'enquête de recensement sur le territoire de la commune permet de déterminer cette population.

La **population comptée à part** comprend certaines personnes dont la résidence habituelle est dans une autre commune mais qui ont une résidence sur le territoire de la commune :

- personnes mineures dont la résidence familiale est dans une autre commune, mais qui résident du fait de leurs études dans la commune ;

- personnes résidant dans une communauté de catégorie 1, 2 ou 3 (établissements sanitaires et sociaux de moyen ou long séjour, maisons de retraite, foyers ou résidences sociales, communautés religieuses, casernes ou établissements militaires) d'une autre commune mais ayant une résidence familiale sur le territoire de la commune ;

- personnes majeures âgées de moins de 25 ans ayant leur résidence familiale sur le territoire de la commune et qui résident dans une autre commune pour leurs études.

La **population totale** est la somme de la population municipale et de la population comptée à part.

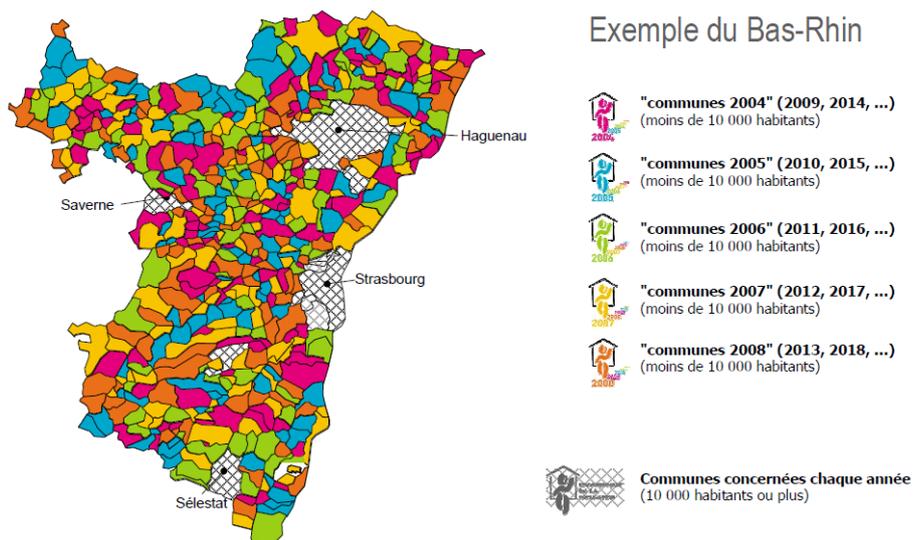
B) Méthode de recensement

Depuis 2004, le recensement repose sur un cycle quinquennal glissant. Ainsi, les informations faisant référence à une année donnée sont élaborées à partir des résultats de 5 enquêtes annuelles de recensement.

Deux catégories de communes (cf. encadré 1-B.1) sont à distinguer.

Les communes dont la population municipale est inférieure à 10 000 habitants (dénommées « petites communes » dans la suite de ce document) sont recensées de manière exhaustive tous les cinq ans. Pour ce faire, ces petites communes ont été réparties en cinq groupes – dits groupes de rotation – en s'assurant que ces groupes soient homogènes sur plusieurs caractéristiques (population par sexe, par âge, logements par catégorie, etc.) [cf. schéma 1-B.1]. Chaque année, les communes d'un groupe de rotation sont recensées. Au bout de cinq ans, l'ensemble des petites communes ont été recensées de manière exhaustive.

Schéma 1-B.1 : Représentation de la répartition des communes de moins de 10 000 habitants pour le département du Bas-Rhin



Les communes dont la population municipale est supérieure à 10 000 habitants (dénommées « grandes communes » dans la suite de ce document) sont recensées chaque année, seulement sur une partie de leur parc de logements. Chaque année, environ 8 % des logements de chaque grande commune sont recensés. L'échantillon de logements à recenser est tiré à partir d'une base de sondage d'adresses issue d'un répertoire exhaustif d'adresses d'habitation, le Répertoire des immeubles localisés (RIL). Ce répertoire, mis à jour en partenariat entre les communes et l'Insee, est un élément clé de la réalisation du recensement dans les grandes communes ; il fournit également le nombre de logements de la commune au

1^{er} juillet de chaque année, soit un des éléments du calcul des populations. Il est présenté de manière détaillée dans la partie 2-A. Les adresses d'habitation sont réparties en cinq groupes de rotation (cf. schéma 1-B.2), et chaque année, environ 40 % des logements d'un groupe de rotation font partie de l'échantillon. Sur un cycle de 5 ans, ce sont ainsi 40 % des logements qui sont recensés. Ce taux est suffisant pour fournir des statistiques robustes sur chaque grande commune (cf. partie 2-F).

Encadré 1-B.1 : la gestion des immeubles et des ensemble immobilier à cheval sur plusieurs communes

Un immeuble est composé au sens de l'Insee, d'un bâtiment unique. Un ensemble immobilier est composé de plusieurs immeubles.

Lorsqu'un même immeuble dispose de plusieurs entrées, l'Insee détermine une entrée principale et une ou des entrées secondaires. Cette distinction est réalisée systématiquement dans les communes de plus de 10 000 habitants dans le répertoire des immeubles localisés. Les critères utilisés pour déterminer le caractère principal ou secondaire d'une entrée sont d'abord de privilégier comme principale l'entrée pour piéton (un accès au garage sera considéré comme secondaire). S'il y a plusieurs entrées pour piéton, l'entrée principale sera celle qui bénéficie des services publics : dépôt du courrier (présence des boîtes aux lettres), ramassage des ordures ménagères...

En cas d'immeuble à cheval sur plusieurs communes, c'est le positionnement de l'entrée principale qui détermine la commune d'appartenance de l'immeuble. Dans les concepts du recensement, un immeuble n'est donc jamais découpé en plusieurs morceaux afin de les affecter à des communes différentes.

Lorsqu'un ensemble immobilier composé de plusieurs immeubles est à cheval sur plusieurs communes, on affecte chaque immeuble à la commune sur laquelle il est situé. Les limites cadastrales sont plus facilement identifiables lorsqu'elles ne coupent pas des immeubles entiers.

Les mêmes règles sont appliquées pour le cas des communautés à cheval sur plusieurs communes.

Schéma 1-B.2 : Représentation de la répartition des adresses au sein des groupes de rotation, pour une commune de plus de 10 000 habitants de France métropolitaine



C) Déroulement de la collecte

L'enquête de recensement de la population est réalisée par les communes. Des agents recenseurs recrutés par chaque commune entrent en contact avec les ménages qui doivent être recensés. En amont de la collecte, les agents recenseurs mettent dans les boîtes aux lettres des logements concernés par le recensement des lettres aux habitants : celles-ci informent les personnes concernées que leur logement sera recensé, que c'est une enquête d'utilité publique obligatoire pour connaître la population et aider les acteurs publics à programmer des infrastructures communales adaptées. L'agent recenseur appose également autant que possible des affiches dans les halls d'immeubles concernés.

Ensuite, lors de la première prise de contact, l'agent recenseur s'assure que le logement est occupé au titre de résidence principale. Si ce n'est pas le cas, seules des informations relatives au logement sont collectées et aucun habitant n'est comptabilisé ; l'agent recenseur précise alors la catégorie du logement : résidence secondaire, logement occasionnel ou logement vacant. S'il s'agit d'une résidence principale, l'agent recenseur propose en premier lieu de répondre par Internet. Pour cela, l'agent recenseur fournit une notice et des identifiants de connexion pour remplir le questionnaire en ligne du recensement. Si le ménage n'est pas en mesure de répondre par Internet, il dépose des questionnaires papier qu'il reviendra chercher, une fois remplis, quelques jours plus tard (méthode du dépôt-retrait). Le questionnaire du recensement est donc auto-administré. Lorsque la situation l'exige (personne n'étant pas en capacité de répondre par elle-même par exemple), l'agent recenseur peut remplir le questionnaire en procédant par interview.

Si l'agent recenseur n'arrive pas à entrer en contact avec le ménage à recenser, il prend contact avec le voisinage afin de déterminer si le logement à recenser est une résidence principale et, le cas échéant, le nombre de personnes qui y résident de manière permanente. En l'absence d'informations, une estimation du nombre d'habitants est réalisée par l'Insee à partir des autres logements recensés (cf. partie 2-D).

Un nouveau protocole a été testé. Il concerne les seules adresses individuelles (*i.e.* les adresses avec un seul logement) et consiste à déposer dans la boîte aux lettres la notice et les identifiants de connexion pour remplir le questionnaire en ligne du recensement. Le ménage peut alors répondre spontanément, sans avoir eu de contact direct avec l'agent recenseur, ce qui allège le travail de collecte. Si le ménage ne répond pas spontanément, l'agent recenseur prend alors contact avec lui pour le relancer et on revient au protocole habituel. Ce protocole sera généralisé à l'ensemble des adresses individuelles pour l'EAR 2021.

Deux paramètres collectés lors des enquêtes annuelles de recensement sont fondamentaux pour le dénombrement de la population. Le premier est la catégorie du logement, afin d'éviter que des ménages soient comptabilisés deux fois, une fois dans leur résidence principale et une fois dans leur résidence secondaire. Le second est le nombre d'habitants permanents du logement. La définition du caractère « permanent » est fixée par le décret n° 2003-485 du 5 juin 2003. Des consignes sur le questionnaire en ligne et sur le questionnaire papier explicitent la manière de traiter chaque situation (colocation, multi-résidence pour raisons professionnelles, études, garde partagée, etc.). Cette définition est conforme aux règlements européens et normes internationales et s'appuie sur la notion de « population de résidents habituels » (cf. encadré 1-C.1 et partie 2-B). Les questionnaires papier sont visualisables sur le site internet le-recensement-et-moi.fr.

Encadré 1-C.1 : la population de résidents habituels, définition de référence aux niveaux européen et international

La définition de la population faisant référence au niveau européen et au niveau international est la « population de résidents habituels ». Elle est définie de la manière suivante (extrait des « recommandations de la conférence des statisticiens européens pour les recensements de la population et des habitations 2020 » de l'Unece [Unece, 2015]) :

« La population servant de base à utiliser pour les comparaisons internationales est la "**population de résidents habituels**". 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 douze mois. L'expression "période continue" signifie que les absences (du pays de résidence habituelle) dont la durée est inférieure à douze mois ne modifient pas le lieu de résidence habituelle. Les mêmes critères s'appliquent à toutes les divisions territoriales du pays.

Le "**lieu de résidence habituelle**" est le lieu géographique où la personne recensée passe habituellement son temps de repos quotidien au cours d'un laps de temps défini, comprenant le temps de référence du recensement. »

Après la phase de collecte, l'Insee réalise des contrôles qui peuvent, dans certains cas, nécessiter de retourner sur le terrain.

Pour les personnes ne vivant pas en logements ordinaires (personnes en habitation mobile, sans-abris, bateliers ou personnes vivant en communauté¹), le protocole du recensement est différent (cf. encadré 2-B.2).

D) Estimations de population

Les résultats des enquêtes annuelles de recensement font l'objet de redressements pour corriger la non-réponse totale et la non-réponse partielle. Une fois les données redressées, elles sont utilisées pour calculer les populations légales annuelles authentifiées par décret. L'Insee diffuse, en fin d'année N+2, les données de population relative au 1^{er} janvier de l'année N. Pour cela, les données du dernier cycle quinquennal sont mobilisées, *i.e.* les données des EAR N-2 à N+2. Les méthodes d'estimations diffèrent selon la taille de la commune (cf. parties 2-E et 2-F pour plus de détail sur les méthodes). Pour les petites communes, le recensement étant exhaustif une année sur cinq, l'estimation de population consiste à estimer le niveau de population de l'année N à partir de la population des deux dernières collectes et de l'évolution observée dans les données fiscales. Pour les grandes communes, l'estimation s'appuie sur le nombre de logements issus du RIL et le nombre moyen de personnes par logement observé lors des cinq dernières enquêtes de recensement.

L'Insee diffuse, à la fin de chaque année N, les populations légales de l'ensemble des circonscriptions administratives, qui sont authentifiées par décret. Ces populations ont comme date de référence le 1^{er} janvier N-2. Par exemple, les populations légales des communes diffusées fin 2019, à utiliser pour l'année 2020, ont une date de référence au 1^{er} janvier 2017. Chaque mois de juin, les populations légales sont complétées par la diffusion de résultats statistiques détaillés. En octobre, des données infra-communales sont également diffusées.

1 La définition des communautés dans le cadre du recensement est fournie dans l'annexe 1.

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.

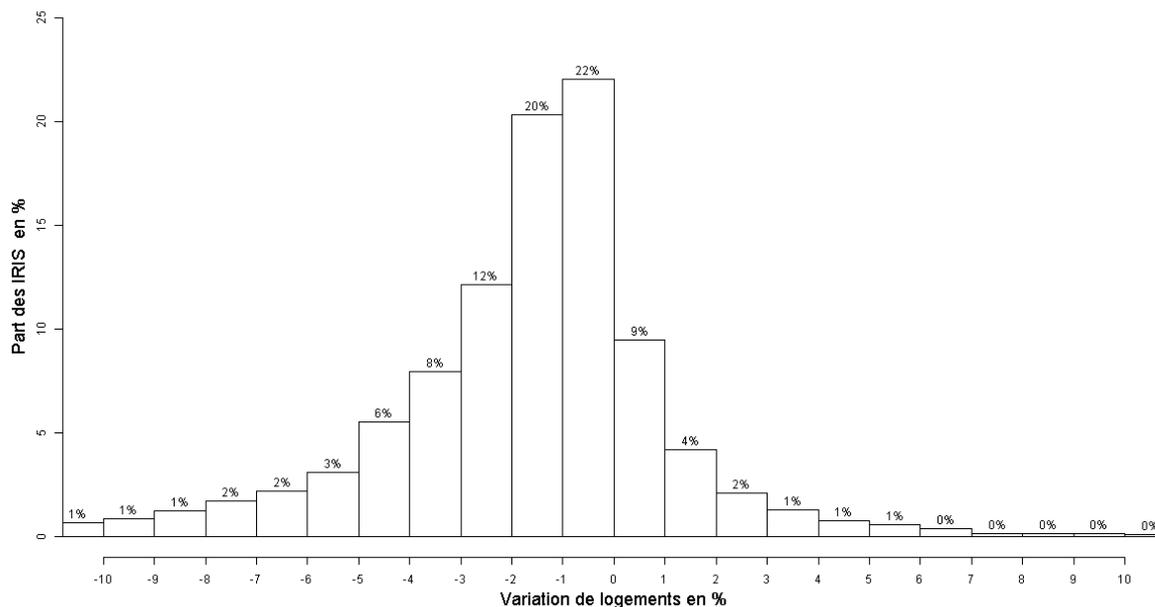
Ce cadre d'assurance qualité du RIL préconise de suivre plusieurs indicateurs qualité.

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

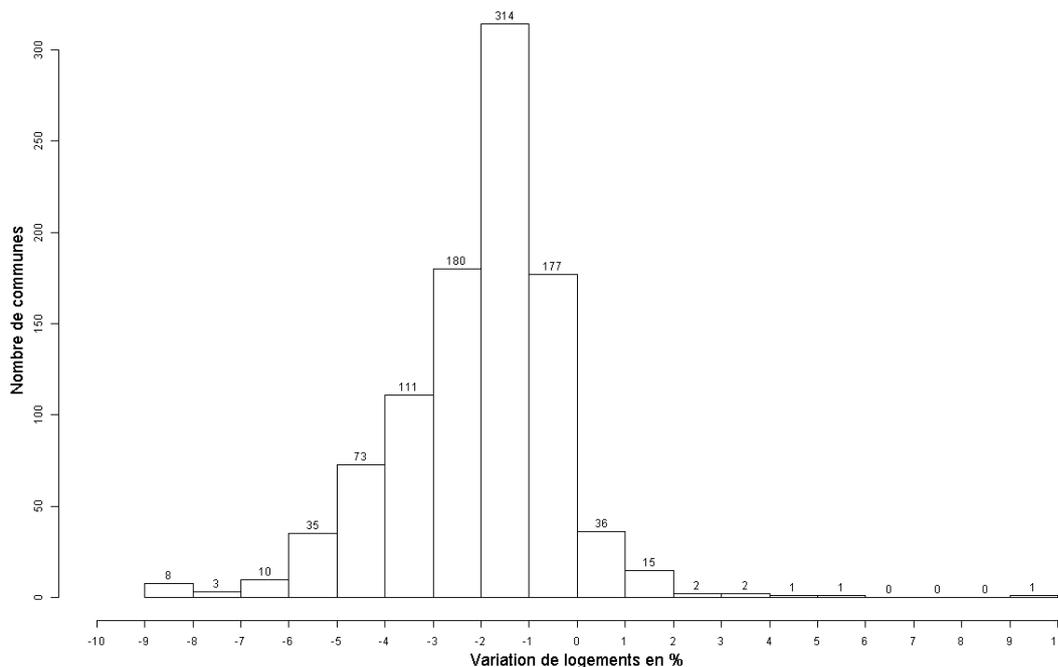


Lecture : pour 22 % des Iris, le nombre de logements de l'échantillon baisse de moins de 1 % à la suite de la collecte.

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

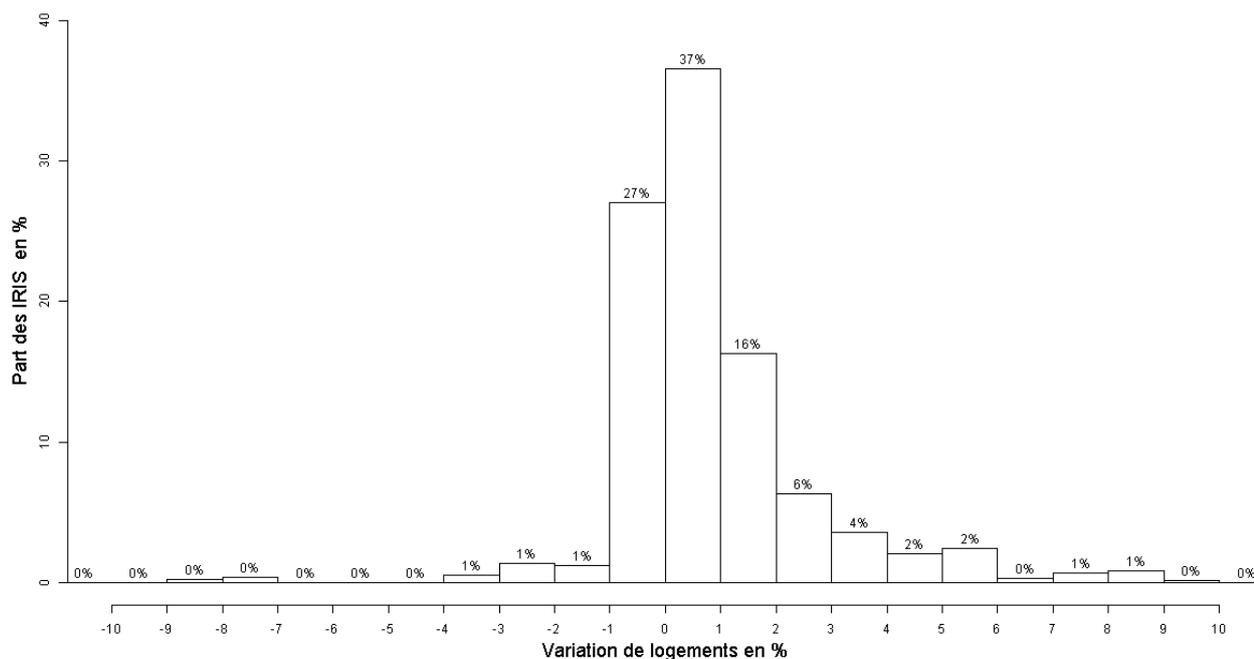
Les EMQR servent à mesurer la qualité du RIL en métropole et à La Réunion. Dans cette partie, nous nous restreignons aux résultats relatifs à la métropole, territoire sur lequel la dernière EMQR a eu lieu en 2017. Il s'agit d'effectuer un ratissage complet d'un échantillon d'Iris. Des enquêteurs de l'Insee 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

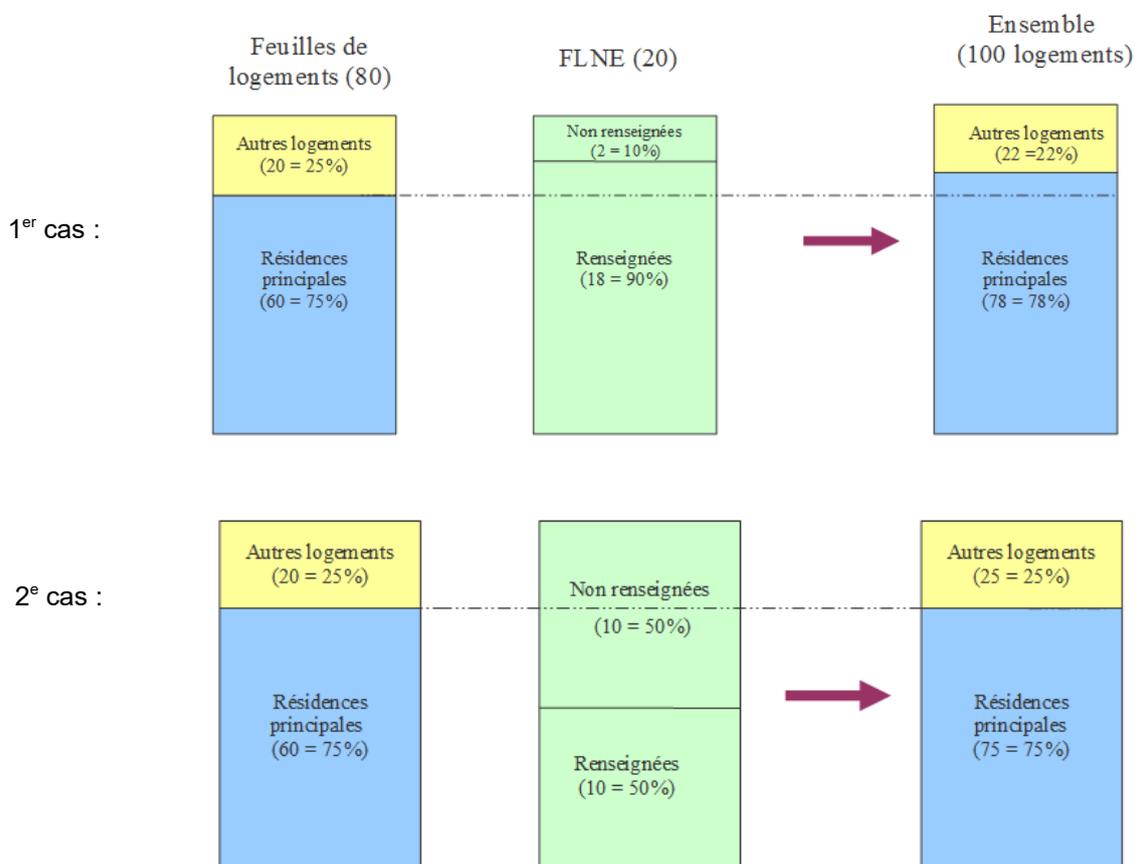
Une FLNE doit normalement être remplie lorsque le logement est une résidence principale non répondante. Néanmoins, dans de nombreux cas, les agents recenseurs ne sont pas en mesure de déterminer si un logement vide est ou non une résidence principale. Une FLNE peut donc dans certains cas signaler qu'un 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.

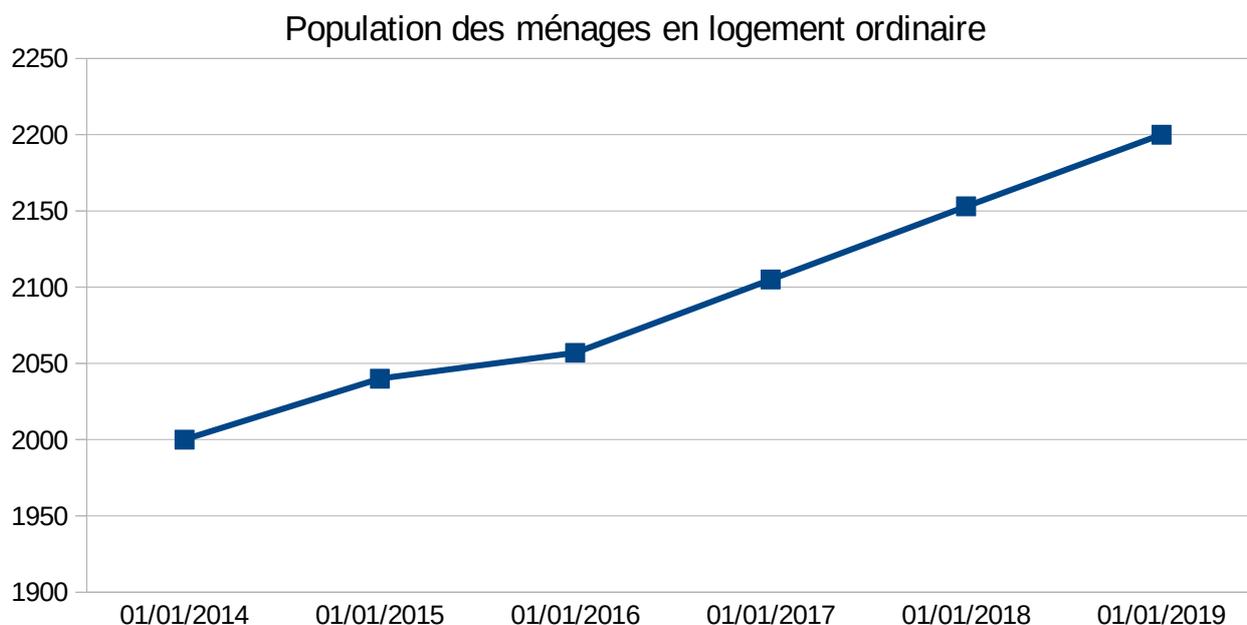
Fin 2020, il s'agit d'actualiser la population au 01/01/2018.

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 Brilhault, 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.).

Partie 3 : Des analyses quantitatives de la qualité des estimations de population

Les contrôles présentés dans la partie 2 sont indispensables pour assurer une bonne qualité des estimations de population. Ils permettent notamment de s'assurer que les traitements statistiques sont réalisés de manière correcte. Ils permettent aussi d'encadrer le processus de collecte des données afin d'éviter de s'éloigner de la réalité du terrain. Toutefois, il n'est pas possible de maîtriser l'ensemble des aléas. Notamment, le recensement reste une enquête auto-administrée, et la réponse des personnes recensées est considérée comme exacte. On fait donc implicitement l'hypothèse que toutes les consignes ont été comprises et respectées.

Cette partie prend plus de recul par rapport au processus de production du recensement pour analyser la qualité des estimations produites. Elle cherche à repérer d'éventuelles sur-estimations ou sous-estimations. Pour ce faire, plusieurs moyens peuvent être imaginés. Le plus intuitif serait de réaliser une enquête de mesure de la qualité du recensement qui viserait à retourner dans un échantillon de logements recensés pour vérifier que les déclarations des personnes correspondent bien à ce que l'on a cherché à collecter : déclaration de la bonne catégorie du logement, déclaration du bon nombre d'habitants permanents du logement, bonne déclaration des personnes multi-résidentes, bonne exhaustivité des logements en petite commune, etc. De telles enquêtes ont été réalisées en France à la suite des recensements de 1962 et de 1990. Coûteuses, elles n'ont à ce jour pas été reproduites¹⁹ depuis le recensement rénové.

Le changement de la feuille de logement en 2018 a mis en évidence que l'ancien questionnaire conduisait à comptabiliser deux fois certaines personnes multi-résidentes. Ce phénomène concernait majoritairement des enfants en résidence partagée qui étaient déclarés en tant qu'habitant permanent du logement par chacun de ses deux parents. En 2018, la manière de « déclarer » les enfants en situation de résidence partagée a été modifiée et les consignes ont été éclaircies. L'Insee estime l'effet du changement de questionnaire en 2018 à une baisse d'environ 500 000 personnes sur la population française, soit 0,75 % (cf. notes techniques 2020, Insee).

Faute de disposer de résultats d'une enquête de mesure de la qualité, nous comparons dans cette partie les estimations de population produites à partir du recensement à d'autres bases de données : l'échantillon démographique permanent (EDP), les données de l'état civil et les bases de données fiscales. Ces comparaisons permettent de mettre en évidence certaines divergences entre les sources, qui tantôt s'expliquent par des différences de concept entre les sources et parfois peuvent laisser soupçonner des défauts du recensement par rapport à ce qu'il cherche à mesurer.

19 À l'exception des enquêtes de mesure de la qualité du RIL présentées dans la partie 2-A, qui fournissent un état des lieux de l'exhaustivité de la base de sondage en grande commune qui est utilisée pour les estimations de la population sur ces territoires. Elles ne constituent toutefois qu'une partie du dispositif d'enquête de mesure de la qualité du recensement qu'il faudrait réaliser.

A) Estimation des doubles comptes statistiques au recensement à partir de l'Échantillon démographique permanent

Cette partie a été rédigée par Laurent Toulemon (Institut national d'études démographiques [Ined]), Sébastien Durier (Insee, Direction des Statistiques Démographiques et Sociales, au moment de la rédaction de cet article) et Benjamin Marteau (Ined).

Ce texte reprend partiellement et met à jour des résultats présentés aux Journées de méthodologie statistique en 2018 (Toulemon, Durier et Marteau, 2018), ainsi que le fruit de discussions avec des collègues de l'Insee. Qu'ils en soient ici remerciés. Comme il est d'usage, les erreurs restent les nôtres. Ce travail a bénéficié du soutien de l'Agence nationale de la recherche (subvention "ANR-16-CE41-0007-01") dans le cadre du projet BigStat (<https://big-stat.site.ined.fr>).

A-1. Les doubles comptes dans le recensement rénové

La qualité statistique des recensements se mesure d'abord par la qualité du comptage des habitants. Les deux erreurs de comptage sont les doubles comptes (personnes recensées deux fois) et les omissions (personnes non recensées). À partir de l'enquête post-censitaire de 1990, la proportion de doubles comptes probables a pu être estimée à 0,7-0,9 %, celle des omissions à 1,5-2,0 %. Ces omissions concernaient principalement des logements occupés par une seule personne, qui sont moins facilement accessibles ou parfois moins visibles. Les personnes rattachées à plusieurs logements ordinaires font partie des personnes à risque d'être soit omises soit en double compte. Les doubles comptes ou les omissions touchaient majoritairement les jeunes adultes âgés de 20 à 24 ans (Coeffic, 1993).

Le but de cet article est d'estimer la proportion d'habitants en situation de double compte dans le recensement rénové. Après avoir défini les doubles comptes statistiques et les méthodes permettant de les identifier, nous présenterons rapidement l'Échantillon démographique permanent, source de données qui permet de réaliser plusieurs estimations, avant de discuter les limites et les prolongements de nos analyses.

Recensement exhaustif et recensement fondé sur des enquêtes annuelles

Dans un recensement exhaustif, les doubles comptes peuvent être identifiés facilement grâce à la présence de deux bulletins correspondant au même habitant. Les doubles comptes peuvent être évités grâce à une règle simple consistant à remplir « un bulletin et un seul par personne ». Avec les enquêtes annuelles de recensement (EAR), le taux de sondage varie entre 8 % dans les grandes communes et 20 % dans les petites, une petite commune sur cinq étant recensée intégralement chaque année (Desplanques, 2008). Les notions d'omissions et de doubles comptes doivent alors être adaptées pour pouvoir être mesurées. Dans le recensement rénové, les doubles comptes ne conduisent au remplissage effectif de deux bulletins la même année qu'environ une fois sur sept (entre 8 % et 100 % des cas, selon que les deux logements sont dans la même commune ou non) : il faut que les deux logements soient inclus dans la même vague d'enquête de recensement. Les doubles comptes sont donc le plus souvent invisibles pour les personnes recensées, seule une minorité donnant lieu au remplissage effectif de deux bulletins la même année, que l'on appelle un bulletin double dans le fichier du recensement (Mambetov, 2014). Nous parlerons de doubles comptes statistiques pour les distinguer des bulletins doubles.

En l'absence d'enquête de qualité du recensement, on n'observe pas les erreurs de mesure du recensement. On ne sait pas si une personne non présente dans le recensement aurait dû l'être et nous n'étudierons pas les omissions dans cette partie. On ne sait pas non plus si une personne présente dans le recensement y est recensée à tort. En revanche, on peut affirmer qu'une personne en bulletin double est recensée à raison dans un logement et à tort dans un autre, et on peut estimer les doubles comptes statistiques à partir des bulletins doubles la même année t , au prix de quelques hypothèses.

Le recensement regroupant les données de cinq vagues annuelles, la même personne peut figurer plusieurs fois dans le fichier si elle a déménagé. Il ne s'agit d'un double compte que si elle a été recensée une fois à tort et l'autre fois à raison. Nous utiliserons les paires d'années de collecte t et $t+1$ pour proposer une estimation améliorée des doubles comptes tenant compte du risque accru de double compte dû au fait que, le plus souvent, une personne ayant deux logements n'est recensée une année donnée que dans un seul de ses logements.

L'échantillon démographique permanent permet d'estimer les doubles comptes

L'identification des individus appartenant à l'Échantillon démographique permanent (EDP – cf. encadré 3-C.1) se fait à partir de chaque vague d'EAR. Avec le recensement rénové, il n'y a de suppression des bulletins doubles qu'au sein d'un même logement. La même personne peut apparaître plusieurs fois la même année (ou des années proches) au recensement dans deux logements différents, sans que l'élimination d'un éventuel bulletin « à tort » ne soit entreprise. On dispose ainsi, pour tous les individus EDP, des informations sur toutes les vagues d'enquête de recensement où ils sont présents. Nous avons calculé des scores d'appariement entre bulletins du recensement attribués au même individu EDP pour nous assurer que les bulletins doubles ne résultaient pas d'erreurs d'appariement. Le détail de ce calcul est fourni dans [Toulemon, Durier et Marteau, 2018] et nous supposons ici qu'il n'y a pas d'erreur de ce type à l'EDP. Pour les analyses fondées sur les données de deux années successives, t et $t+1$, nous appelons bulletins doubles la présence d'un bulletin de recensement une année t et d'un autre bulletin l'année $t+1$, dans lequel, à la question sur la résidence un an avant, « Où habitiez-vous le 1^{er} janvier [t]? », la personne répond « Dans le même logement que maintenant ». Toutes nos analyses sont fondées sur les données brutes du recensement, pour éviter que les imputations ou corrections ne créent des faux bulletins doubles. Bien que l'EDP contienne des variables socio-fiscales très riches (Insee, 2018), nous n'utilisons ici que l'information issue des vagues de recensement.

Méthodes d'estimations, à partir des données d'une année

On se place dans la situation où il n'y a pas d'omission. On suppose également que les habitants sont recensés au plus dans deux logements. En pratique, presque aucun individu n'est recensé trois fois ou plus la même année dans des logements différents. On cherche à estimer la proportion de doubles comptes statistiques une année donnée à partir des bulletins doubles observés cette même année. Dans les calculs, on tient compte du fait que les probabilités d'inclusion sont différentes (20 % pour les logements des petites communes et pour les logements neufs ou les grandes adresses des grandes communes, 8% pour les autres) ainsi que du fait que deux logements ne sont pas indépendants s'ils appartiennent à la même petite commune ou à la même adresse d'une grande commune.

La probabilité qu'une personne ayant deux logements habituels soit interrogée dans ces deux logements une année donnée varie donc de $0,2*1 = 20$ % pour les habitants ayant deux logements dans la même petite commune à $0,08*0,08 = 0,64$ % pour ceux vivant dans deux adresses différentes en grande commune.

Tous les calculs tiennent compte précisément du plan de sondage des EAR dans le calcul de la probabilité de présence simultanée de deux logements dans la même vague de recensement (Toulemon, 2018). Cependant, pour simplifier la présentation de la méthode, on suppose que tous les logements ont la même probabilité d'inclusion et que, pour deux logements différents, les probabilités de sélection sont indépendantes. En appelant p la probabilité de sélection d'un logement, on peut écrire $w = 1/p$ la pondération de l'individu dans son logement.

Quatre catégories d'habitants

On peut définir quatre catégories au sein de la population, dont on va estimer l'importance et le poids dans la population recensée, classées ici par risque croissant de double compte. Le recensement est fondé sur cinq vagues annuelles. On ne s'intéresse qu'aux personnes rentrant dans le champ du recensement (personnes

ayant leur résidence principale en France) : une personne ayant un logement répond au recensement ; une personne ayant plusieurs logements répond au moins une fois ; on suppose également que les habitants ont deux logements au maximum. On omet également la pondération EDP tenant compte du fait que l'échantillon EDP couvre une personne sur 100 jusqu'en 2008 et une personne sur 25 à partir de 2009 (Jugnot, 2014) ; nous calculons des proportions de doubles comptes dans le champ EDP et nous l'appliquons à l'ensemble de la population.

Catégorie A : individus a ayant un seul logement habituel. Pour les individus ayant un seul logement, on définit $Y(a,i)$ comme l'indicatrice de présence de l'individu a dans le fichier du recensement dans le logement i , et $Y(a)$ la somme des $Y(a,i)$ pondérée par le poids $w(i)$ de l'individu a dans chaque logement i . En pratique, le poids est calculé pour le logement i et affecté à tous les habitants du logement.

$$\text{Pour } a \text{ dans } i : Y(a,i) = 1$$

Soit J l'ensemble de tous les logements recensés une année donnée. En écrivant $Pop(A)$ la taille de la population appartenant à la catégorie A , et $Rec(A)$ la taille de la catégorie A au recensement :

$$Rec(A) = \sum_{i \in J} p \sum_{a \text{ dans } i \text{ et } a \in A} wY(a,i) = p w Pop(A) = Pop(A)$$

Logiquement, le recensement estime correctement l'effectif des habitants ayant un seul logement habituel. Les personnes ayant une résidence secondaire dans laquelle elles ne sont pas susceptibles de remplir un bulletin du recensement (le logement étant repéré sans ambiguïté comme une résidence secondaire au recensement) appartiennent à cette catégorie.

Pour les individus ayant deux logements habituels dans lesquels un agent recenseur est susceptible de leur laisser un bulletin individuel, on définit $Y(a,i)$ l'indicatrice de présence au recensement de a dans le logement i en t . On suppose que la personne habite dans deux logements i et j . On suppose que le logement i est leur résidence principale au sens du recensement, la personne est donc recensée à raison dans le logement i et à tort dans le logement j , mais on ne connaît pas la résidence principale dans le fichier. On peut envisager 4 cas pour les individus ayant deux logements i et j :

Cas 1 : Ni i ni j ne sont dans J : a n'est pas interrogé par le recensement.

Cas 2 : i est dans J , et pas j : a est interrogé dans i .

Cas 3 : j est dans J , et pas i : a est interrogé dans j .

Cas 4 : i et j sont dans J : a est interrogé deux fois, dans i et dans j .

Dans un recensement exhaustif, J représente l'ensemble des logements et seul le cas 4 existe ; avec un recensement par enquête, on peut écrire pour une année donnée, si chaque logement a une probabilité p d'inclusion et si les logements sont indépendants les uns des autres :

$$P(\text{cas 1}) = (1-p)^2 ; P(\text{cas 2}) = p(1-p) ; P(\text{cas 3}) = p(1-p) ; P(\text{cas 4}) = p^2$$

La probabilité pour une personne ayant deux logements de participer au recensement dans un seul logement (cas 2 ou 3) est donc de $2p(1-p)$, tandis que la probabilité de se voir proposer deux bulletins est de p^2 . Nous allons distinguer trois catégories parmi les habitants ayant deux logements habituels.

Catégorie B : individus ayant deux logements et capables d'identifier le logement où ils doivent être recensés. Seuls ceux qui doivent être recensés dans i remplissent un bulletin en t :

$$\text{Pour } a \text{ dans } i \text{ et } i \text{ résidence principale de } a : Y(a,i) = 1$$

Les habitants de la catégorie B sont également interrogés dans leur logement j , mais ils ne remplissent pas de bulletin, car j n'est pas leur résidence principale et ils ont lu, compris et appliqué les consignes (le logement j est pour eux une résidence habituelle, mais pas leur « résidence principale » au sens du recensement).

$$\text{Pour } a \text{ dans } j \text{ et } j \text{ non résidence principale de } a : Y(a, j) = 0$$

Pour les habitants de la catégorie B , $Y(a)$ vaut w dans les cas 2 et 4, 0 sinon

À partir de l'échantillon J , on estime au recensement la population de B , notée $Rec(B)$ par :

$$Rec(B) = \sum_{i \in J} p \sum_{a \text{ dans } i \text{ et } a \in B} wY(a, i) = (p(1-p) + p^2) w Pop(B) = Pop(B)$$

Les habitants de la catégorie B ont deux logements, mais une seule résidence principale. Du point de vue du recensement, ils ne produisent pas de double compte.

Catégorie C : individus ayant deux logements, ne lisant pas les consignes, mais sensibles au fait qu'ils ne doivent pas être recensés deux fois. Ces personnes ne lisent pas les consignes et ont deux résidences habituelles dans lesquelles elles sont prêtes à remplir un bulletin. Cependant si les deux logements sont présents la même année, elles ne remplissent qu'un seul bulletin (pour éviter le double compte, avec le raisonnement qu'elles n'ont pas à remplir deux bulletins). On suppose que le logement i est leur résidence principale au sens du recensement. Si i et j font partie de la même vague de recensement, on suppose qu'elles répondent dans le premier logement avec une probabilité 1 et dans le deuxième avec une probabilité 0 : elles évitent de participer deux fois au recensement la même année. Par contre, si un seul logement est inclus dans l'échantillon une année donnée, elles remplissent un bulletin dans le logement. En particulier, si un logement est inclus dans le recensement en t et un autre en $t+1$, elles remplissent deux bulletins, car elles ont oublié avoir déjà répondu au recensement ou ne perçoivent pas le double compte.

Dans le cas où les deux logements sont interrogés l'année donnée, seuls les habitants qui sont d'abord recensés dans i remplissent un bulletin en i ; les autres remplissent un bulletin en j ; finalement, tous ne remplissent qu'un seul bulletin :

$$\text{Pour } a \text{ recensé dans } i \text{ et } i \text{ avant } j : Y(a, i) = 1$$

$$\text{Pour } a \text{ recensé dans } j \text{ et } j \text{ avant } i : Y(a, j) = 1$$

Si un seul logement est interrogé l'année donnée, les habitants de la catégorie C remplissent un bulletin en i ou en j :

$$\text{Pour } a \text{ recensé dans } i : Y(a, i) = 1$$

$$\text{Pour } a \text{ recensé dans } j : Y(a, j) = 1$$

Pour les habitants de la catégorie C , $Y(a)$ vaut w dans les cas 2, 3 et 4, 0 sinon.

À partir de l'échantillon J , on a :

$$Rec(C) = \sum_{i \in J} p \sum_{a \text{ dans } i \text{ et } a \in C} wY(a, i) = (2p(1-p) + p^2) w Pop(C) = (2-p) Pop(C)$$

Les habitants de la catégorie C ne sont jamais en bulletin double la même année ; cependant, ils sont le plus souvent en double compte statistique, et leur poids total (effectif pondéré de la catégorie C) n'est pas $Pop(C)$ mais $(2-p) Pop(C)$. On attribue à chaque bulletin des habitants de la catégorie C un poids w , car on ignore que le même habitant est également susceptible d'être interrogé et de remplir un autre bulletin dans son autre logement.

Catégorie D : individus ayant deux logements, ne lisant pas les consignes et répondant au recensement dans tous leurs logements. Ces personnes ne lisent pas les consignes et ont deux résidences habituelles dans lesquelles elles sont prêtes à remplir un bulletin. Les habitants de la catégorie *D* remplissent un bulletin en *i* et un bulletin en *j*, même si ces deux logements sont recensés la même année :

$$\text{Pour } a \text{ recensé dans } i : Y(a, i) = 1$$

$$\text{Pour } a \text{ recensé dans } j : Y(a, j) = 1$$

Pour les habitants de la catégorie *D*, $Y(a)$ vaut w dans les cas 2 et 3, $2w$ dans le cas 4, 0 sinon.

À partir de l'échantillon *J*, on a :

$$Rec(D) = \sum_{i \in J} p \sum_{a \text{ dans } i \in D} wY(a, i) = (2p(1-p) + 2p^2) wPop(D) = 2Pop(D)$$

Pour un recensement exhaustif, $p=1$, les catégories *B* et *C* sont identiques et ne donnent pas lieu à des doubles comptes ; les doubles comptes correspondent à la seule catégorie *D*. Dans un recensement par enquête annuelle, avec $p < 1$, les habitants de la catégorie *C* s'ajoutent à ceux de la catégorie *D*, et la proportion de doubles comptes augmente. Les habitants de la catégorie *D* sont en double compte statistique (leur poids total dans le fichier du recensement est $2 Pop(D)$ au lieu de $Pop(D)$) ; on peut parler pour les habitants de la catégorie *C* de doubles comptes partiels, leur poids total étant inférieur à $2 Pop(C)$, car quelques habitants de la catégorie *C* « évitent » le double compte : s'ils sont interrogés dans leurs deux logements, ils ne répondent qu'une seule fois.

Estimation finale des doubles comptes. Les habitants de la catégorie *D* sont en double compte dès que l'un de leurs logements est recensé, puisqu'ils répondent et sont inclus dans les fichiers dans chaque logement ; les habitants de la catégorie *C* sont en double compte uniquement si un seul de leurs logements est recensé une année donnée. Quel est le poids relatif de ces catégories dans les fichiers du recensement ? Une personne susceptible d'être recensée dans deux logements et effectivement recensée dans un logement une année donnée a respectivement les probabilités $2p(1-p)$ et p^2 d'être recensée une ou deux fois une année donnée. Dans le recensement fondé sur 5 EAR, les mêmes formules s'appliquent, à ceci près que, parmi les personnes dont les deux logements sont inclus dans l'échantillon du recensement, une personne sur 5 est recensée deux fois la même année et 4 sur 5 sont recensées deux années de suite ou à deux ans d'écart. Enfin, certaines personnes ne sont recensées que dans un logement sur les cinq ans et d'autres ne sont pas recensées du tout. Avec $p^2=1/36$, probabilité observée pour les doubles comptes en $t; t^{20}$, on obtient les ordres de grandeur présentés au tableau 3-A.1. Les probabilités exactes, calculées pour chaque strate de type de commune et de logement, conduisent à la même estimation qu'en appliquant une probabilité moyenne de sélection.

20 Les doubles comptes sont plus fréquents dans les grandes communes et, dans les grandes communes, dans les grandes adresses et les logements neufs, ce qui aboutit à un poids moyen observé pour les personnes en bulletin double plus faible, de l'ordre de 36 (soit un taux de sondage moyen de 1/6) au lieu de 50 (pour un taux de sondage de 14 %, environ 1/7).

Tableau 3-A.1 : Probabilités de sélection de deux logements pour une EAR ou au recensement (cumul de cinq EAR)

	Nombre d'EAR		Avec $p=1/6$	
	Une EAR	Cinq EAR	Une EAR	Cinq EAR
Deux logements recensés...	p^2	$25p^2$	1/36	25/36
- la même année	p^2	$5p^2$	1/36	5/36
- deux années de suite		$10p^2$		10/36
- à deux ans d'intervalle		$10p^2$		10/36
Un seul logement recensé	$2 p (1-p)$	$10 p (1-5p)$	10/36	10/36
Aucun logement recensé	$(1-p)^2$	$(1-5p)^2$	25/36	1/36
Compte : habitant A ou B	1	1	1	1
Compte : habitant C	$2-p$	$2-p$	11/6	11/6
Compte : habitant D	2	2	2	2

Note : ces valeurs sont estimées sous l'hypothèse d'indépendance d'inclusion de chaque logement. Elles fournissent des ordres de grandeur. En pratique, nous avons calculé les probabilités pour les strates de type de commune et de logement (Toulemon, Durier et Marteau, 2018). Les comptes (poids moyen pondéré) valent 1 pour les personnes comptées une fois et 2 pour les doubles comptes. Les comptes pour les 5 EAR sont divisés par 5. On voit que les habitants de la catégorie C sont en double compte partiel ($11/6=1,83$), car une fois sur six ils sont recensés deux fois la même année et ne remplissent qu'un seul bulletin.

Estimation des effectifs des différentes catégories

Bulletins doubles en t et catégorie D : doubles comptes systématiques. À partir des bulletins doubles observés l'année t , on estime directement l'effectif de la catégorie D, qui donne lieu à un bulletin double $BD(t)$ dans le cas 4, à partir de la probabilité du cas 4. En notant $\mathbb{1}(E)$ la variable indicatrice de l'événement E (qui vaut 1 si E est vrai, 0 sinon) et EDP l'ensemble des individus EDP, on a :

$$Pop(D) = \sum_{a \in EDP} \frac{1}{p^2} \mathbb{1}(Y(a)=2w) = w^2 BD(t)$$

Les individus de D ayant une probabilité double d'apparition, leur poids total dans les fichiers du recensement est le double de leur effectif :

$$Rec(D) = 2w^2 BD(t)$$

Bulletins doubles en $t;t+1$ et catégorie C. Sous l'hypothèse que la somme des $Y(a)$ en t et $t+1$ vaut $2w$ quand un logement est inclus dans le recensement en t et l'autre en $t+1$ pour les personnes catégories C et D qui n'ont pas déménagé entre t et $t+1$, on peut estimer la proportion de bulletins doubles $BD(t;t+1)$ en t , à partir des bulletins doubles en t et $t+1$: parmi les habitants des catégories C ou D, certains peuvent être recensés en i en t et en j en $t+1$. Ils seront observés comme des doubles comptes en t sous deux conditions supplémentaires :

- les logements i et j appartiennent à deux vagues annuelles successives. Cela restreint le champ aux doubles comptes des habitants dont les deux logements ne sont ni dans la même petite commune ni à la même adresse dans une grande commune. Pour ceux dont les deux logements sont dans la même petite commune ou à la même adresse, seuls les habitants de la catégorie D sont en double compte, puisque leurs deux logements sont nécessairement recensés la même année ;

- à l'enquête de $t+1$, l'habitant recensé en j doit y résider depuis plus d'un an pour avoir été en double compte en t , quand il a été recensé dans le logement i . Il faut éviter de compter à tort comme doubles comptes les personnes résidant en i en t et ayant déménagé pour j en $t+1$, qui ne sont pas en double compte. On ne conserve donc que les personnes recensées en t dans le logement i et qui, à la question sur

la résidence un an avant, posée en $t+1$ dans le logement j , « Où habitez-vous le 1^{er} janvier [t] ? » ont répondu « Dans le même logement que maintenant ». Ils habitaient donc en t dans les deux logements i et j .

La première condition limite le nombre de doubles comptes au recensement ; la seconde limite leur possibilité d'observation : une personne en double compte a deux fois moins de chances d'être observée comme bulletin double en $t;t+1$ si elle a quitté un de ses logements que si elle n'a pas déménagé : il faut qu'entre t et $t+1$, elle ait déménagé de i (où elle a été recensée en t) mais pas de j où elle est recensée en $t+1$ (et répond en $t+1$ qu'elle résidait déjà dans j en t) ; si elle a quitté ses deux logements, elle ne peut pas être observée comme bulletin double en $t;t+1$. On peut estimer la proportion s , parmi les habitants en double compte en t , de ceux qui vivent dans un ou dans les deux logements depuis plus d'un an, et qui sont susceptibles d'être observés à partir des vagues d'enquête t et $t-1$. Sous l'hypothèse de stationnarité, on utilise cette proportion pour les années $t;t+1$ pour estimer finalement la proportion de doubles comptes²¹. On définit alors $Y(a,j,t+1)$ l'indicatrice de résidence de a dans j en t , repérée dans la vague d'enquête en $t+1$. On a :

$$Pop(C+D) = \sum_{a \in EDP} \frac{1}{2} \frac{1}{p^2} \frac{1}{s} 1 \left(\sum_{i \in J(t), j \in J(t+1), i \neq j} Y(a,i) + Y(a,j,t+1) = 2 \right) = \frac{w^2}{2s} BD(t;t+1)$$

Bulletins doubles sans double compte. Par construction, si une personne figure deux fois dans le fichier de la même vague d'enquête en t , elle est en double compte. À l'inverse, certains habitants peuvent figurer deux fois dans le fichier, une fois en t et une fois en $t+1$, sans être en double compte. Tout d'abord, environ 14 % des personnes recensées en t et ayant déménagé entre t et $t+1$ ont emménagé dans un logement présent dans la vague annuelle de $t+1$: il est très important de les distinguer des personnes en double compte qui vivaient déjà en t dans le logement où elles sont recensées en $t+1$. Nous utilisons l'information brute (non imputée) sur la résidence un an auparavant et nous considérons tous les cas douteux comme des déménagements. Les changements de géographie de la collecte du recensement (en lien notamment avec les fusions de communes) peuvent également rendre quelqu'un éligible dans le même logement deux années de suite. Nous avons éliminé ces cas après avoir vérifié qu'ils étaient très peu nombreux. Enfin, dans certains cas de multi-résidence, certaines personnes pourraient avoir conservé leurs logements, mais changé de résidence principale au recensement entre t et $t+1$. Nous avons identifié trois possibilités :

- Enfants en résidence partagée 50-50 entre les logements M et P de leurs mère et père séparés, qui auraient passé la nuit de référence du recensement en M l'année t et en P l'année $t+1$, M et P faisant partie des échantillons des enquêtes annuelles t et $t+1$ respectivement.

- Jeunes adultes recensés en t dans leur logement d'étude et en $t+1$ chez leurs parents dans leur logement familial (où ils revenaient les week-ends en t) ; en $t+1$, ils peuvent déclarer qu'ils habitaient dans le logement familial un an avant, même s'ils ont été recensés ailleurs.

- D'une manière générale, tous les adultes qui ont deux logements et passent plus de la moitié de l'année dans un logement l'année t et plus de la moitié dans l'autre l'année $t+1$ peuvent figurer à raison dans deux vagues annuelles de suite. La question sur la résidence un an auparavant ne permet pas d'identifier les personnes qui vivaient dans le même logement mais dont la résidence principale était ailleurs : on suppose que les personnes recensées en $t+1$ dans un logement et répondant qu'elles y habitaient déjà en t auraient pu être recensées dans ce logement si le recensement y avait eu lieu en t .

21 L'hypothèse de stationnarité suppose que les doubles comptes sont constants, en nombre et en structure. Elle permet d'identifier les transitions entre t et $t-1$, que l'on observe au recensement à partir de la question sur la résidence un an avant, aux transitions entre t et $t+1$ qui servent à estimer la proportion s de doubles comptes observables. Dans l'analyse fondée sur les EAR de t et $t+1$, l'échantillon est multiplié par deux et le nombre de bulletins doubles est également deux fois plus élevé, puisque les combinaisons (i en t et j en $t+1$) et (j en t et i en $t+1$) sont différentes, tandis que la combinaison (i en t et j en t) est unique. Le produit des poids est donc divisé par $2s$, et non par s .

Dans tous ces cas, un bulletin double identifié en $t;t+1$ conduit à « estimer » à tort $w^2/2$ doubles comptes de catégorie C, puisque le bulletin double est observé à partir de deux vagues successives. Seul le premier cas correspond à des informations cohérentes avec les consignes du recensement, mais on peut le considérer comme assez improbable, d'autant que la date de référence du recensement ne figure ni sur les bulletins individuels ni sur la feuille de logement²². On estime à 400 000 au maximum le nombre d'enfants en garde partagée 50-50 en 2016, soit 3 % des enfants mineurs (Algava *et al*, 2019 ; Chardon et Vivas (2004) l'estimaient à moins de 100 000, sur une période plus ancienne, tandis que Toulemon et Denoyelle (2012) estiment la population d'enfants au risque de double compte entre 250 000 et 500 000 pour la décennie 2000). Une année donnée, on compte, d'après l'EDP, 200 000 enfants en doubles comptes. Selon que ces enfants sont des enfants en résidence alternée 50-50 ou non, restent entre 200 000 et 400 000 enfants « au risque d'être en bulletin double » en $t;t+1$ sans être en double compte si les consignes sont lues et respectées. Comme l'alternance est *a priori* indépendante de la date de référence du recensement, ces enfants ont chaque année une chance sur deux de passer la nuit dans le logement recensé, soit une chance sur quatre d'être comptés deux fois à raison. Cela représente donc entre 50 000 et 100 000 enfants, soit 0,4 % à 0,7 % des enfants mineurs, et 0,16 % de la population au maximum, sous l'hypothèse que les parents dont les enfants sont en garde partagée lisent les consignes et les appliquent, connaissent la date de la nuit de référence du recensement. Il est possible que ce soit le cas pour certains d'entre eux, sachant que les enfants en résidence partagée ont relativement plus de chances d'appartenir à la catégorie D que C, car ils ne remplissent pas eux-mêmes leur bulletin : si deux bulletins sont remplis pour eux, ils ne le sont pas par la même personne. Or le rapport des catégories C et D varie peu avec l'âge (Toulemon, Durier et Marteau, 2018). La correction à apporter serait de l'ordre de 0,1 % sur les doubles comptes $t;t+1$. Nous ne l'avons pas effectuée.

Estimation de la variance des estimations. Compte tenu de la grande complexité du plan de sondage du recensement et du fait que les probabilités de bulletin double varient comme le produit des probabilités de sélection de chaque logement, l'écart-type de chaque estimation est difficile à estimer. Par ailleurs, les effets de grappe sont limités par le fait que les doubles comptes sont rares et dispersés sur le territoire (à l'exception des fratries en doubles comptes dans les mêmes deux logements). Nous avons donc adopté la méthode d'estimation suivante : à partir d'un nombre n d'individus dans l'EDP représentant un groupe (ici les individus présents deux fois au recensement) correspondant à une proportion u de la population, on estime l'effectif N dans la population P et $V(N)$ et $V(u)$ la variance de u à partir des formules suivantes, avec w le poids des individus du groupe (produit du poids EDP et du produit des poids de sondage de l'EAR (inverse de la probabilité d'être sélectionné deux fois, voir tableau 3-A.1 et (Toulemon, 2018) pour le détail des probabilités selon la combinaison des deux logements entre petites et grandes communes) et \hat{w} le poids moyen des individus du groupe :

$N = \hat{w} n$ estimation du nombre d'individus en double compte

$u = N / P$ estimation de la proportion d'individus en double compte

$V(n) = n$ estimation poissonnienne de la variance en première approximation

$V(u) = u^2 n z$ estimation de la variance de la proportion d'individus en double compte tenant compte de la variance des poids (poids EAR * poids EDP pour les bulletins doubles).

On estime le facteur z d'augmentation de la variance dû à la variance des poids des membres du groupe comme $[1 + (1/\hat{w}^2) V(w)]$. Cette formule est obtenue en estimant la variance de l'effectif pondéré comme la variance du produit des deux variables X (être en bulletin double) et w (pondération), sous les hypothèses d'absence de covariance entre X et w et de faiblesse de X devant w : $V(uX) = V(X) V(u) + V(X) + V(u) [E(X)]^2 \approx V(X) [1 + V(u)]$ avec u poids ramené à la moyenne 1 ($u = w/\hat{w}$). Sur un groupe d'effectif similaire (les

²² La date de référence a été ajoutée sur la feuille de logement en 2018, mais nos analyses portent sur des années antérieures.

couples de même sexe repérés dans l'enquête Famille et logements) et également réparti dans l'ensemble du territoire, nous avons montré (Toulemon et al., 2014) que cette estimation reproduit très bien l'ordre de grandeur de la variance, par rapport à une estimation tenant compte de l'ensemble du plan de sondage (strates par région, tirage équilibré dans les régions, grappes de logements).

Entre 1,5 et 3 % de doubles comptes au recensement

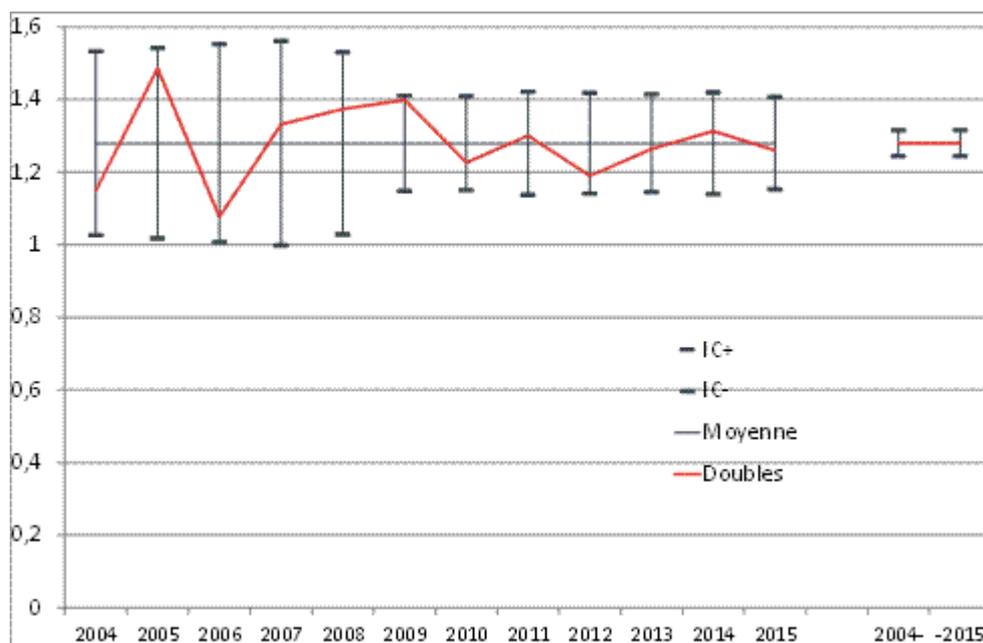
1,3 % de doubles comptes estimés à partir des bulletins doubles la même année

On compte dans l'EDP environ 210 bulletins doubles chaque année jusqu'en 2008, et 910 de 2009 à 2015, après l'augmentation du taux de sondage de l'EDP de 1 % à 4 % (de 4 à 16 jours de naissance par an). Ces effectifs sont tout à fait suffisants pour estimer les doubles comptes de la catégorie *D*, ceux qui peuvent donner lieu à un bulletin double la même année. Le poids moyen dans la vague annuelle de recensement (poids EAR) des personnes recensées deux fois la même année est de 6 en moyenne dans chacun de leurs logements : dans les fichiers du recensement, ils comptent dans les grandes communes pour 1,5 % et dans les petites pour 1,1 %, souvent dans des logements neufs ou des grandes adresses (Toulemon, Durier et Marteau, 2018)²³. Pour les années 2004 à 2015, la proportion de doubles comptes s'établit à 1,3 %. Cette proportion est stable d'une année à l'autre (Graphique 3-A.1).

On peut estimer l'intervalle de confiance en tenant compte des différentes catégories de commune et de la variance des poids : sur l'ensemble des années observées, l'intervalle de confiance est de [1,23 % – 1,33 %], autour d'une valeur centrale de 1,28 %, soit 800 000 personnes en double compte au recensement (entre 770 000 et 830 000), pour une population métropolitaine moyenne de 62,6 millions en 2009. Cet intervalle de confiance est indicatif, car le recensement est une opération complexe, avec un plan de sondage du recensement qui combine découpage en strates, équilibrage des EAR et effets de grappe, et auquel s'ajoute le tirage de l'EDP. Cependant, cette incertitude est très inférieure à la différence avec l'estimation issue des doubles comptes repérés à partir de deux vagues successives (la catégorie *C* s'ajoutant alors à la catégorie *D*).

23 Leur pondération est égale au produit des poids (pondérations de l'EAR) dans chacun des logements où elles habitent, multiplié par leur poids EDP. Leur part est calculée comme le rapport de leur effectif estimé à celui de l'estimation de la population totale d'après la vague EAR de l'année considérée (poids de l'EAR x poids EDP de l'ensemble des personnes recensées).

Graphique 3-A.1 : Proportion d'individus en double compte par année de recensement



Lecture : en 2004, 1,15 % des individus recensés sont en double compte. L'intervalle de confiance est de [1,01 %-1,53 %] autour de la valeur moyenne de 1,28 %, au seuil de 5 % bilatéral.

Champ : bulletins individuels (BI) des individus nés un jour EDP, à l'exclusion des BI avec un poids nul. EAR 2004-2015.

Source : Base étude 2015 de l'EDP.

2,4 % de doubles comptes estimés à partir des bulletins doubles deux années de suite

Pour les années de recensement qui s'étendent de 2011 à 2015, la proportion de doubles comptes statistiques mesurés la même année atteint 1,27 % (graphique 3-A.1). Quelle est la proportion de personnes en double compte en t (catégorie D) susceptibles d'être repérées à partir des EAR t et $t+1$? Il faut pour cela que leurs deux logements ne soient pas recensés la même année et qu'elles n'aient pas emménagé depuis moins d'un an dans le logement où elles sont recensées en $t+1$. Le tableau 3-A.2 présente le détail des estimations. Les personnes ayant leurs deux logements dans la même petite commune ou à la même adresse (0,06 %, tableau 3-A.2, ligne b1) ne peuvent être en bulletin double en t et $t+1$. Les autres doubles comptes en t (1,21 %, tableau 3-A.2, lignes b2 et b3) sont susceptibles d'être en bulletin double en t et $t+1$: recensées en t dans un logement et en $t+1$ dans un autre logement où elles déclarent qu'elles y résidaient déjà un an auparavant (donc en t). En supprimant parmi les doubles comptes en t,t les personnes résidant dans leurs logements depuis moins d'un an et la moitié de celles résidant dans un de leurs logements depuis moins d'un an (et depuis plus d'un an dans l'autre), la proportion descend à 0,76 % (tableau 3-A.2, ligne b2), ce qui est conforme à ce que l'on sait des épisodes de multi-résidences dans les enquêtes longitudinales : d'après l'enquête SRCV, la part des multi-résidences durant plus d'une année est de l'ordre de 50 % à 65 % (Toulemon et Pennec, 2009). Pour le même champ, l'estimation directe des doubles comptes statistiques à partir de deux vagues successives est de 1,60 % (tableau 3-A.2, ligne c2). On suppose ensuite que le rapport $c2/b2$, représentant le rapport entre le nombre de personnes de catégorie $C+D$ et le nombre de celles de catégorie D , s'applique aussi aux personnes ayant déménagé entre t et $t+1$: $c3 = b3 c2/b2 = 0,94$ %. Finalement, on estime à 2,54 % la proportion de doubles comptes en t repérés à partir des bulletins doubles en $t,t+1$ (tableau 3-A.2, ligne c).

Chaque logement des habitants en double compte potentiel ayant une probabilité moyenne observée de 1/6 d'être sélectionné chaque année, au bout de 5 ans, quatre cas sont possibles (tableau 3-A.1) : deux logements interrogés la même année ($p=5/36$) ; deux logements interrogés deux années différentes

($p=20/36$) ; un seul logement interrogé ($p=10/36$) ; aucun logement recensé ($p=1/36$). Quand l'individu est présent au recensement, les deux logements sont interrogés la même année avec une probabilité de $5/35=1/7$, et un seul logement est interrogé, ou les deux logements sont interrogés deux années différentes avec une probabilité de $30/35=6/7$. Finalement, les personnes ayant deux logements habituels et recensées sont donc interrogées deux fois la même année une fois sur sept (dans ce cas, seule la catégorie D est en double compte, $p=1,27\%$) et interrogées une seule fois ou deux fois lors de deux EAR différentes (dans ce cas, les catégories C et D sont en double compte, à condition que les deux logements ne soient pas dans la même petite commune ou à la même adresse, $p=2,54\%$). L'estimation finale de la proportion de doubles comptes statistiques sur l'ensemble d'un recensement s'écrit $e = [d(t;t)+6d(t;t+1)] / 7$, soit $2,36\%$. Ramené à l'ensemble de la population (et non plus l'échantillon EDP), cela conduit à une estimation finale de 1 501 000 personnes en double compte, pour une population métropolitaine de 63 650 000 au 1^{er} janvier 2013. Les variations aléatoires autour de cette estimation sont du même ordre que précédemment (on observe 4 643 bulletins doubles en $t;t+1$ de 2011 à 2015), ce qui conduit à un intervalle de confiance final autour de l'estimation de $2,36\%$ de l'ordre de $[2,27 ; 2,45]$, en tenant compte uniquement des variations liées au sondage du recensement et de l'EDP.

Les doubles comptes sont à peu près deux fois plus probables à partir des bulletins doubles dans deux vagues successives, ce qui confirme notre hypothèse selon laquelle les doubles comptes sont plus souvent visibles et donc évités quand les deux logements sont interrogés la même année (tableau 3-A-2).

On retrouve le même résultat en estimant la proportion d'habitants de la catégorie D à $1,27\%$, puis ceux de la catégorie C à $1,39\%$, dont $1,10\%$ en double compte. La proportion s , parmi les habitants de la catégorie D, de ceux qui n'ont pas déménagé et ont leurs deux logements dans des adresses ou des petites communes différentes (lignes b2/b, $0,76/1,27$) est de 60% . À la ligne c2, on observe $1,60\%$ des habitants, dont $0,76\%$ de la catégorie D et donc $0,84\%$ de la catégorie C. En supposant que la proportion s est identique pour les habitants de la catégorie C, on estime leur effectif à $0,84/60\% = 1,39\%$. Ces habitants sont en double compte partiel (voir tableau 3-A.1) : une chance sur six d'être compté une fois, 5 chances sur 6 d'être compté deux fois ; poids moyen $11/6$). On compte donc $1,39 \times 5/6 = 1,16\%$ doubles comptes théoriques, auxquels on enlève ceux dont les deux logements sont dans la même petite commune ou à la même adresse ; finalement, on compte $1,10\%$ d'individus en double compte de la catégorie C, et $2,36\%$ de doubles comptes au total.

Tableau 3-A.2 : Comparaison en $t;t$ et $t;t+1$ des individus déclarant ne pas avoir déménagé (2011-2015)

	Effectifs pondérés		Proportions (en %)	
	$t;t$	$t;t+1$	$t;t$	$t;t+1$
a. Individus EDP recensés (5 vagues, pondération EAR)	13 804 821	13 804 821	100,00	100,00
b. Individus recensés deux fois en t (catégorie D), dont...	174 896		1,27	
b1. Individus à la même adresse ou dans la même petite commune	7 615		0,06	
b2. Individus déclarant ne pas avoir déménagé (hors même petite commune ou même adresse)	105 462		0,76	
b3. Individus ayant déménagé entre $t-1$ et t (hors même petite commune ou même adresse)	61 819		0,45	
c2. Individus (catégories C+D) déclarant ne pas avoir déménagé (hors même petite commune ou même adresse)		221 033		1,60
c3. Individus (catégories C+D) ayant déménagé entre t et $t+1$ (hors même petite commune ou même adresse)				0,94
c. Estimation d'individus en doubles comptes en $t;t+1$				2,54
d1. Estimation des doubles comptes: catégorie D en $t;t$ et en $t;t+1$			1,27	1,21
d2. Estimation des doubles comptes: catégorie C en $t;t$ et en $t;t+1$			0,00	1,33
d. Estimation des doubles comptes en $t;t$ et en $t;t+1$			1,27	2,54
e. Estimation finale (chaque logement a une probabilité 1/6 d'être inclus en t)			2,36	
e. Estimation finale de doubles comptes au recensement, pour une population totale de 63 650 000 habitants début 2013	1 501 000		2,36	

* Pour $t;t$, sont également inclus les individus déclarant avoir déménagé dans un des deux logements recensés (on divise alors par deux leur poids pour avoir des champs comparables). Pour faciliter la cohérence des totaux, on estime $t;t$ avec t de 2011 à 2015, et $t;t+1$ avec $t+1$ de 2011 à 2015.

Champ : $t;t$: bulletins doubles des individus nés un jour EDP. $t;t+1$: bulletins individuels des individus recensés lors de deux vagues successives. EAR 2011-2015.

Source : base étude 2015 de l'EDP.

Les scores d'appariement montrent que la proximité entre bulletins d'une même personne est plus forte quand les deux bulletins ne sont pas recueillis la même année, mais deux années de suite : dans ce cas, il est plus probable que les deux bulletins ont été remplis par la personne elle-même, tandis que, quand deux bulletins sont remplis la même année, cela peut être par deux personnes différentes, la personne elle-même dans un logement et un parent dans l'autre, par exemple (Toulemon, Durier et Marteau, 2018).

Doubles comptes et estimation de la population

Une estimation finale de 2,4 % de doubles comptes, assez fragile

Ces estimations des doubles comptes sont assez fragiles. Les effectifs sont suffisants pour que les variations aléatoires soient très faibles ; par contre, les sources d'erreur sont nombreuses. Des erreurs d'appariement à l'EDP ou des réponses erronées à la question sur la résidence un an avant créent un bulletin double une fois sur sept (si le nouveau logement en $t+1$ est recensé) et on multiplie ensuite son poids par sept, donc chaque erreur d'appariement ou réponse erronée conduit à un faux double compte. Les

erreurs d'appariement à l'EDP pourraient construire des « faux bulletins doubles », mais nos investigations nous conduisent à penser que les procédures d'appariement sont très rigoureuses et les scores d'appariement sont très élevés pour les différents bulletins du même habitant. De même, nous avons utilisé la variable brute du recensement pour limiter au maximum les erreurs de codage au recensement (Toulemon, Durier et Marteau, 2018).

Les réponses erronées peuvent être plus fréquentes, mais sont difficiles à identifier. D'après les équipes en charge du recensement, dans près d'un cinquième des cas où l'année d'emménagement dans le logement correspond à l'année précédente ou l'année en cours (t), les personnes déclarent être présentes dans le logement depuis plus d'un an à la question « Où habitiez-vous le 1^{er} janvier $[t-1]$? ». Il est possible que, malgré les instructions, certaines personnes répondent à cette question en prenant comme date de référence le 1^{er} janvier de l'année en cours et non le 1^{er} janvier de l'année précédente. C'est probablement la plus grande source de fragilité de l'estimation en $t;t+1$.

Les hypothèses de comportement sont également assez fortes. Pour une année donnée, une personne ayant deux logements et présente au recensement ne verra ses deux logements recensés la même année qu'une fois sur onze : presque tous les doubles comptes sont invisibles pour les habitants, car ils ne donnent pas lieu à des bulletins doubles ; en ce sens, on peut parler de doubles comptes statistiques : les personnes qui répondent dans leurs deux logements ont deux fois plus de chances que les autres d'être recensées, mais sont très rarement confrontées à deux bulletins dans leurs deux logements. Quand on considère le recensement cumulant cinq EAR, la probabilité que les deux logements soient inclus la même année est un peu plus élevée : elle est de $5/35 = 1/7$; la probabilité d'être recensé deux années de suite est de $10/35=2/7$, et la probabilité qu'il y ait plus d'un an entre les deux inclusions ou que seul un logement soit inclus dans l'échantillon du recensement est de $20/35 = 4/7$ (tableau 3-A.2). Nous avons fait l'hypothèse que les personnes de la catégorie C remplissaient un bulletin en t et un autre en $t+1$ sans se soucier du double compte. Sous l'hypothèse que certains doubles comptes seraient évités en $t;t+1$ mais pas dans les cas où les deux logements sont recensés à un écart temporel plus grand qu'un an ou après un déménagement, notre estimation devrait en être augmentée. En l'absence d'une enquête post-censitaire, tous les doubles comptes sans bulletin double restent invisibles.

Les données de l'EDP permettent donc de proposer une première estimation des doubles comptes au recensement depuis le passage à un recensement par enquête : 2,4 %. Les intervalles de confiance dus aux aléas de sondage sont serrés [2,2 % ; 2,5 %] et difficiles à estimer précisément, les aléas de sondage de l'EDP et du recensement se combinant. On peut cependant les estimer grossièrement et de nombreuses autres sources d'incertitude existent ; l'incertitude totale sur l'estimation est beaucoup plus large. Cependant, la proportion de doubles comptes statistiques est certainement supérieure à 1,5 % et inférieure à 3 %. D'une part, la proportion déduite des bulletins doubles en $t;t$ (1,3 %) est certainement inférieure à la vraie proportion, puisque la grande majorité des doubles comptes ne donnent pas lieu à deux bulletins remplis la même année ; d'autre part, il est probable que peu d'habitants éviteraient un double compte si leurs deux logements sont recensés deux années de suite, mais pas si la durée entre les deux interrogations est supérieure. Nous proposons donc l'estimation de 2,4 % avec un intervalle de confiance de [1,5 % ; 3,0 %], soit entre 950 000 et 1 900 000 personnes.

Autres possibilités offertes par l'EDP

Notre estimation est plus élevée que la proportion de 0,8 % de doubles comptes estimée au recensement de 1990 (Coeffic, 1993), et comparable à ce que l'on obtient dans les recensements étrangers avant correction (Census Bureau, 2012 ; Toulemon, Durier et Marteau, 2018).

L'EDP permet, au-delà du comptage des doubles comptes, un très grand nombre d'analyses que nous avons esquissées ailleurs grâce aux nombreuses informations sur les individus EDP et les autres membres de leur ménage recensés avec eux. Ce sont surtout les jeunes adultes qui sont comptés deux fois, leur profil étant très proche de celui des personnes ayant deux résidences habituelles (Toulemon et Pennec, 2010 ;

Imbert *et al.*, 2014). De plus, on dispose pour ces habitants présents deux fois la même année des informations sur leurs deux logements et leurs deux ménages. La correction des doubles comptes conduirait à diminuer l'effectif total de la population recensée, mais dans une mesure partielle puisque, si l'on supprimait les personnes recensées seules à tort dans un logement, il faudrait également supprimer le logement de la liste des résidences principales (environ 0,8 % de ménages en moins). L'estimation totale de la population étant fondée sur la liste des résidences principales et le nombre d'habitants par logement, elle ne serait modifiée que de 1,5 %. L'écart est plus sensible sur des indicateurs comme la situation familiale des enfants : la proportion d'enfants vivant avec leurs deux parents (Buisson et Lapinte, 2017) passe de 75 % à 78 % si l'on corrige du fait que les enfants comptés deux fois partagent le plus souvent leur temps entre les logements de leurs deux parents séparés. Les définitions des fratries ou des situations familiales méritent d'être analysées pour ces enfants, dans la mesure où les gardes partagées deviennent plus nombreuses.

Conclusion

Compte tenu des nombreuses sources d'incertitude dans les méthodes d'estimation, nous proposons une fourchette large de 1,5 % à 3 % autour de notre estimation centrale de 2,4 % des habitants comptés deux fois au recensement, proportion stable entre 2004 et 2015. Pour certains groupes spécifiques ayant plus que les autres deux résidences habituelles (enfants de parents séparés, jeunes adultes revenant de temps en temps vivre chez leurs parents, personnes recensées seules dans leur logement), cette proportion est plus élevée et la description des situations familiales peut en être affectée (Toulemon, Durier et Marteau, 2018). Les situations de multi-résidence sont très diverses (Toulemon, Pennec, 2010) et il en va de même pour les situations familiales des personnes recensées deux fois (Toulemon, Durier et Marteau, 2018). La modification de la feuille de logement du recensement en 2018 a conduit à une diminution de la population estimée et à un ajustement des mouvements de la population (Papon et Beaumel, 2019 ; Insee, 2020). Il est probable que ce soient surtout des doubles comptes d'enfants qui ont été évités (Toulemon, 2018b), mais cela reste à confirmer. L'amélioration de la feuille de logement a été concentrée sur les enfants, sans question explicite sur l'ensemble des situations de multi-résidence ; l'ajustement prévu, de l'ordre de 500 000 personnes au total, pourrait correspondre à un tiers des doubles comptes.

La comparaison avec les données administratives, qui posent elles-mêmes des difficultés de comptage, est compliquée. L'EDP fournit des informations très riches sur les doubles comptes, qui pourront continuer à permettre des analyses précises ; l'utilisation de sources de données complémentaires au sein de l'EDP, telles que les données socio-fiscales, pourrait faire avancer les réflexions sur les caractéristiques socio-démographiques des individus en double compte ou omis dans le recensement, et est une piste prometteuse à l'avenir.

B) Omissions possibles de jeunes enfants au recensement : que nous apprennent les comparaisons avec l'état civil ?

Les comparaisons entre les effectifs des jeunes enfants et ceux des naissances enregistrées en France à l'état civil suggèrent de possibles omissions dans le recensement, plus particulièrement pour les nouveau-nés, mais également, de manière plus générale, pour les enfants en bas âge.

Ces omissions ne sont pas spécifiques à la méthode actuelle de recensement annuel, fondée sur la collecte d'Enquêtes annuelles de recensement depuis 2004, qui, cumulées sur cinq années successives, donnent les résultats du recensement. Elles s'observent aussi dans les recensements généraux de la population, qui étaient fondés sur une collecte exhaustive environ tous les dix ans.

Le déficit de jeunes enfants au recensement n'est pas spécifique non plus à la France. Il s'observe aussi dans les recensements étrangers, notamment en Grande-Bretagne et aux États-Unis. Ces cas sont connus et documentés, à l'image d'un commentaire rédigé à partir du cas de la France par Laurent Toulemon, chercheur à l'Institut national d'études démographiques (Ined), dans une revue étrangère ayant consacré un numéro au sous-dénombrement des jeunes enfants dans les recensements (Toulemon, 2017).

L'obligation d'attribuer un numéro d'identification au répertoire (NIR) à toutes les personnes nées sur le territoire français et l'obligation de déclarer les nouveau-nés dans des délais brefs après la naissance, ainsi que les contrôles mis en œuvre pour ne pas avoir de doubles enregistrements, font que les statistiques d'état civil sont considérées comme une référence sur le dénombrement des naissances.

On compare donc les effectifs de naissances à l'état civil avec les effectifs des jeunes enfants des générations correspondantes dans le recensement de la population.

Les concepts recueillis dans les deux sources n'étant pas strictement les mêmes, il convient tout d'abord de les préciser et de définir un champ de comparaison le plus proche possible (partie 3-B-1), avant d'exposer une estimation des écarts (partie 3-B-2), quelques caractéristiques des écarts (partie 3-B-3) et l'impact sur la pyramide des âges (partie 3-B-4).

B-1. État civil et recensement : des concepts et champs de couverture différents

Le recensement fournit des statistiques sur les naissances d'enfants dont la mère réside habituellement en France

Le recensement apporte des informations statistiques sur les personnes qui résident habituellement en France. L'analyse est menée ici à partir des enquêtes annuelles de recensement de la population de 2004 à 2018.

Champ géographique

Chaque année, ces enquêtes interrogent un échantillon de personnes en France métropolitaine et dans 4 DOM²⁴ pour fournir des données statistiques sur les personnes résidant habituellement en France. On dispose ainsi d'informations sur les enfants nés au plus tard l'année précédant la collecte et résidant habituellement en France, c'est-à-dire dont la mère est recensée. Par exemple, pour la collecte de l'EAR 2011 réalisée début 2011, les enfants les plus jeunes observés²⁵ sont nés en 2010 et ils ont été recensés, parce que leur mère réside habituellement en France.

24 Le recensement à Mayotte sur cette période est quant à lui exhaustif et quinquennal.

25 Les enfants nés au début de l'année de l'EAR (2011 dans le cas présent) sont également recensés, mais sont écartés pour les exploitations statistiques dans la mesure où la date de référence est le 1^{er} janvier de l'année.

L'état civil fournit des statistiques exhaustives sur les enfants nés en France

Les communes transmettent en continu à l'Insee les informations relatives à l'état civil sur leur territoire. En particulier, elles lui adressent de façon exhaustive les bulletins de naissance d'enfants nés dans la commune. Le bulletin de naissance comporte également des informations relatives à la commune de domicile des parents.

Le champ géographique

Les statistiques concernent l'ensemble du territoire. Mais pour les générations nées avant 1994, l'Insee ne dispose des fichiers statistiques de l'état civil que pour la France métropolitaine.

Naissances enregistrées / naissances domiciliées

L'état civil enregistre toute naissance ayant lieu en France, quel que soit le lieu de résidence de la mère. Il s'agit des **naissances enregistrées**. Le total des naissances en France publié dans les bilans démographiques est le total des naissances enregistrées.

Les parents des nouveau-nés peuvent ne pas habiter sur le territoire. On peut compter les naissances intervenues en France dont la mère réside en France d'après la déclaration de naissance : **on parle alors de naissances domiciliées en France**.

Se ramener à des concepts et champs les plus proches possibles pour comparer les naissances d'après le recensement à celles de l'état civil

On va comparer les effectifs d'enfants à chaque âge, entre 0 et 15 ans, présents dans les enquêtes annuelles de recensement à l'effectif dans l'état civil des naissances de la génération correspondante.

Quelques précautions sont nécessaires, pour raisonner sur des champs et concepts les plus comparables possibles.

Concept retenu pour le recensement : les enfants recensés en France et qui y sont nés

Le recensement porte sur les personnes résidant habituellement en France.

Pour se rapprocher des données de l'état civil, on considère, pour le recensement, uniquement les enfants recensés en France et qui y sont nés. En effet, les enfants nés hors du territoire français ne seront, par définition, pas inclus dans les statistiques de l'état civil. L'analyse est menée en comparant les enfants nés et recensés en France dans les EAR aux naissances enregistrées en France dans l'état civil.

Le champ géographique : France métropolitaine

La première EAR date de 2004, et donne des informations sur les naissances ayant eu lieu au cours de l'année 2003. Les enfants de 15 ans recensés en 2004 sont nés au cours de l'année 1988. En 1988, les fichiers statistiques de l'état civil disponibles couvrent uniquement la France métropolitaine.

Pour raisonner sur le même champ géographique dans les deux sources, l'analyse est donc menée en comparant les naissances d'**enfants nés en France métropolitaine et recensés en France métropolitaine dans les EAR aux naissances enregistrées en France métropolitaine d'après l'état civil**. Les constats dressés restent valables pour la France dans son ensemble (métropole + DOM).

La prise en compte des décès

Bien que ces cas soient rares, des bébés décèdent l'année même de leur naissance ou dans les années qui suivent. Ils ne seront mécaniquement pas comptabilisés dans l'EAR qui suit leur décès. Ainsi, **pour estimer le nombre des naissances de l'année précédant l'EAR à partir de l'EAR, on soustrait aux naissances de l'état civil les décès qui ont affecté la génération considérée**. On procède ainsi à chaque âge. Par

exemple, le nombre attendu d'enfants de moins d'un an de la génération N au 1^{er} janvier de l'EAR N+1 correspond au nombre de naissances de l'année N, moins les décès des enfants de cette génération survenus cette même année. On obtient ainsi, pour chaque âge, un nombre théorique d'enfants comparable à celui de l'EAR.

Ce traitement est rendu possible par les données d'état civil : en effet, les bulletins de décès ayant lieu en France sont transmis exhaustivement à l'Insee par les communes, au même titre que les bulletins de naissance.

Par adéquation au champ géographique retenu, les décès pris en compte sont les décès enregistrés en métropole.

La prise en compte des migrations

Des bébés nés en France, comptabilisés donc dans l'état civil, ont pu quitter ensuite le territoire, par exemple si leurs parents sont partis s'installer à l'étranger : ils ne sont alors pas comptabilisés au recensement après le départ des parents. En toute rigueur, en plus des décès, il faudrait soustraire aux naissances de l'état civil le nombre d'enfants ayant quitté la métropole. **Il n'est toutefois pas tenu compte ici du solde migratoire**²⁶. De manière générale, les enfants, en particulier les plus jeunes, sont peu concernés par les migrations hors de France (Le Penven, 2020). En 2016, 2000 enfants d'1 an nés en France vivaient à l'étranger (source : EDP).

La comparaison de l'effectif d'une génération recensée à un âge donné dans une EAR à l'effectif de cette génération à l'état civil n'est donc pas ajustée pour tenir compte des migrations aux jeunes âges²⁷. À noter que le correctif macro de la pyramide des âges proposé ensuite (cf. annexe 2) tient compte des migrations.

B-2. Moins d'enfants en bas âge au recensement que dans l'état civil

Comparaison du nombre des naissances d'après le recensement au nombre des naissances de l'état civil

Pour des raisons de comparaison avec l'état civil, seuls sont donc considérés dans le recensement les enfants recensés en France et qui y sont nés. Dans l'exploitation qui suit, le champ est restreint à la France métropolitaine.

Le recensement fournit par ailleurs le nombre d'enfants nés hors métropole et vivant en métropole à un âge donné. On en estime les effectifs par âge de manière séparée, pour voir comment l'arrivée des enfants nés hors métropole compense une partie de l'écart observé sur ceux nés en métropole (cf. annexe 2).

26 Pour estimer les effectifs à 0 an, on ne tient pas compte à ce stade du solde migratoire. Le solde migratoire apparent à 0 an, calculé par différence entre les effectifs à 0 an au 1^{er} janvier de l'année N+1 et les naissances de l'année N, corrigées des décès à 0 an de cette même année, est très négatif, du fait de la sous-estimation des bébés au recensement (cf. graphique A2.4 en annexe 2), mais il ne reflète *a priori* pas la réalité, compte tenu des omissions possibles de jeunes enfants.

27 Par ailleurs, le solde migratoire aux jeunes âges, calculé comme différence entre deux pyramides des âges successives, elles-mêmes estimées à partir des EAR, est sujet à caution aux jeunes âges, du fait des omissions probables de jeunes enfants au recensement. Ce point est abordé en annexe 2 sur le correctif macro.

Pour chaque âge entre 0 et 15 ans, on calcule l'écart entre le nombre d'enfants nés et recensés en France métropolitaine et le nombre de naissances enregistrées en France métropolitaine à l'état civil, diminué des décès. On procède ainsi pour chacune des Enquêtes annuelles de recensement (EAR) de 2004 à 2018. Pour chaque âge, on peut ensuite estimer l'écart moyen entre les deux sources, calculé comme la moyenne des écarts constatés dans chaque EAR sur la période 2004-2018 (cf. tableau 3-B.1).

Tableau 3-B.1: Écarts entre les effectifs nés et recensés en métropole et les naissances enregistrées à l'état civil (diminuées des décès)

Enquête annuelle de recensement	Age atteint au 1 ^{er} janvier											
	0 an	1 an	2 ans	3 ans	4 ans	5 ans	6 ans	7 ans	8 ans	9 ans	10 ans	15 ans
2004	-44 035	-42 563	-31 908	-27 987	-23 414	-26 589	-24 211	-23 856	-30 639	-25 332	-21 727	-17 378
2005	-34 354	-39 001	-28 449	-24 733	-14 866	-20 000	-19 093	-19 186	-15 876	-17 877	-17 210	-9 716
2006	-30 857	-35 760	-30 962	-27 708	-15 697	-12 786	-13 797	-10 679	-18 043	-12 883	-18 048	-14 620
2007	-38 987	-40 309	-40 690	-32 608	-22 627	-24 752	-16 379	-16 477	-18 688	-16 285	-17 416	-20 431
2008	-40 283	-43 599	-33 994	-29 599	-23 378	-23 167	-22 212	-13 796	-12 816	-10 664	-17 888	-13 680
2009	-41 788	-45 381	-36 730	-33 145	-26 641	-29 327	-23 302	-21 754	-11 916	-13 130	-18 285	-18 266
2010	-35 320	-50 253	-36 954	-31 579	-27 207	-26 605	-22 865	-17 051	-15 838	-8 199	-10 762	-10 387
2011	-36 760	-40 866	-34 873	-33 276	-29 847	-19 953	-24 631	-18 504	-13 731	-11 964	-3 946	-10 297
2012	-43 345	-43 060	-42 380	-39 773	-35 552	-34 131	-27 160	-27 677	-26 549	-17 618	-18 552	-14 074
2013	-48 041	-49 901	-39 591	-38 264	-34 450	-30 741	-25 473	-30 400	-17 077	-17 069	-16 110	-14 569
2014	-47 016	-53 315	-43 543	-37 314	-29 183	-30 213	-22 793	-21 081	-14 979	-12 656	-14 085	-11 885
2015	-53 288	-57 657	-43 808	-44 637	-39 183	-26 819	-32 718	-23 637	-19 593	-19 338	-17 509	-10 321
2016	-55 208	-68 356	-54 942	-47 482	-43 370	-32 984	-31 702	-23 585	-29 795	-25 737	-17 729	-2 749
2017	-55 303	-64 879	-56 991	-49 627	-45 567	-36 954	-26 031	-31 443	-29 591	-24 682	-21 890	-12 828
2018	-58 043	-63 596	-58 995	-59 237	-54 083	-45 807	-47 321	-35 086	-43 249	-41 593	-37 599	-26 109
Écart moyen	-44 175	-49 233	-40 987	-37 131	-31 004	-28 055	-25 312	-22 281	-21 225	-18 335	-17 917	-13 821

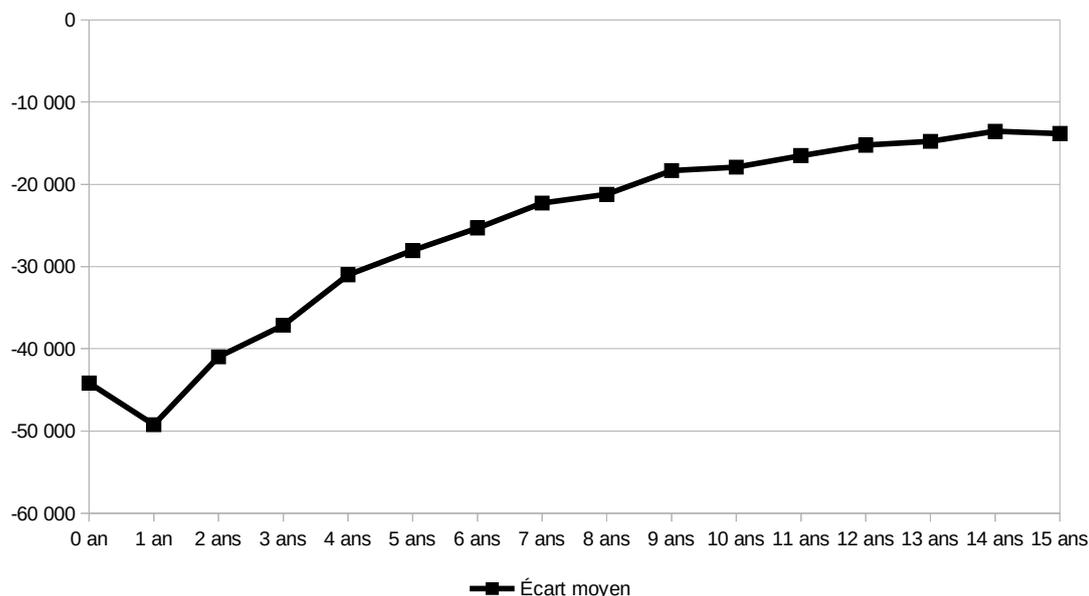
Champ : enfants recensés et nés en France métropolitaine – Naissances enregistrées en métropole, diminuées des décès enregistrés en métropole.

Source : Insee, enquêtes annuelles de recensement de 2004 à 2018 et statistiques d'état civil.

Les écarts entre les deux sources sont les plus importants aux jeunes âges. À 0 an, il y a en moyenne 44 000 enfants de moins dans les EAR que dans l'état civil. À 1 an, le déficit moyen du recensement est de 49 000 enfants. Les écarts sont encore importants aux âges suivants : - 41 000 à 2 ans, - 37 000 à 3 ans, - 31 000 à 4 ans et - 29 000 à 5 ans. Les omissions de jeunes enfants au recensement décroissent avec l'âge. À 15 ans, l'écart moyen entre l'état civil et le recensement s'établit à - 14 000 enfants.

Par ailleurs, les écarts entre les deux sources sont variables selon les années, et ont globalement tendance à s'accroître. Pour les enfants de 0 an par exemple, on trouve entre 31 000 (EAR 2006) et plus de 50 000 enfants (EAR 2015 et après) de moins au recensement qu'à l'état civil.

Graphique 3-B.1 : Les écarts entre le recensement et l'état civil se réduisent avec l'âge



Champ : enfants de 0 an recensés et nés en France métropolitaine (âge atteint au 1^{er} janvier) – naissances enregistrées en métropole, diminuées des décès enregistrés en métropole.

Source : Insee, enquêtes annuelles de recensement de 2004 à 2018 et statistiques d'état civil.

Il est naturel que des écarts existent entre les deux sources, état civil et EAR, du fait de concepts différents. Même en définissant des concepts les plus proches possibles, on ne peut réduire à zéro les écarts dus aux différences de concepts. Pour autant, cela n'invalide pas les constats présentés ici, les effectifs en jeu n'étant pas du même ordre : les écarts que l'on pourrait imputer aux différences de concept ou de temporalité²⁸ entre les deux sources sont vraisemblablement inférieurs aux écarts estimés ici. Par ailleurs, comme mentionné plus haut, une petite partie de ces écarts peut être due à des enfants partis vivre à l'étranger. Enfin, la fiabilité du nombre de naissances de mères résidant en France peut aussi être questionnée : n'y aurait-il pas des mères résidant à l'étranger qui choisiraient d'accoucher en France en déclarant une adresse en France au moment de la naissance de leur enfant ? Cela reste difficile à mesurer. Pour autant, il est probable qu'au moins une partie de ces écarts résulte d'un sous-dénombrement des jeunes enfants au recensement.

B-3. Quels déterminants possibles de ces écarts ?

D'où peuvent provenir ces omissions de jeunes enfants au recensement ? Quels en sont les déterminants ? Il est difficile d'identifier les facteurs d'omission potentiels, de mesurer leur impact et d'établir les facteurs prédominants. Dans la suite, on se concentre sur les enfants de 0 an et on les analyse selon différents critères : région, taille de la commune, nationalité des parents.

Les résultats suivants portent sur l'écart entre :

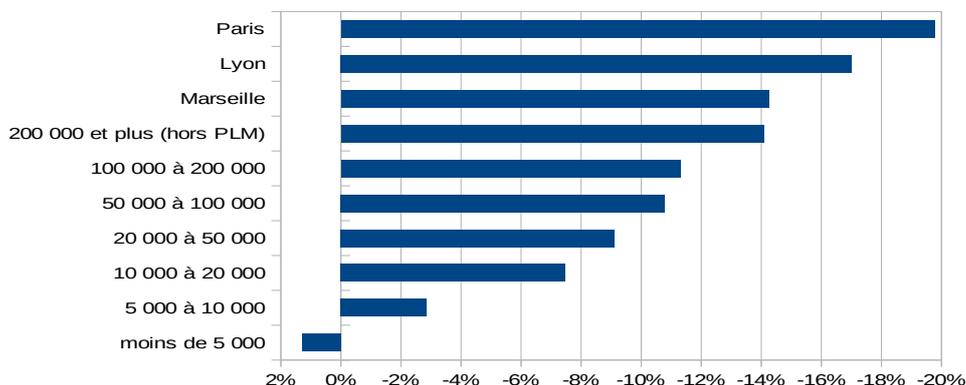
- les naissances enregistrées à l'état civil en métropole, d'enfants de mère résidant en métropole
- et les effectifs d'enfants âgés de 0 an au 1^{er} janvier dans l'EAR, nés et recensés en métropole.

²⁸ Le fait de comparer, par exemple, les effectifs d'enfants de 5 ans dans l'EAR aux naissances survenues 6 ans auparavant.

Des différences accentuées dans les métropoles, en Île-de-France et en Provence-Alpes-Côte d'Azur

Les communes de plus de 10 000 habitants concentrent la quasi-totalité de l'écart (cf. graphique 3-B.2). Plus la taille de la commune est importante, plus l'écart relatif l'est également. L'écart est ainsi particulièrement conséquent pour Paris, Lyon et Marseille.

Graphique 3-B.2 : Écarts de bébés de 0 an entre l'état-civil et l'EAR, selon la taille de la commune

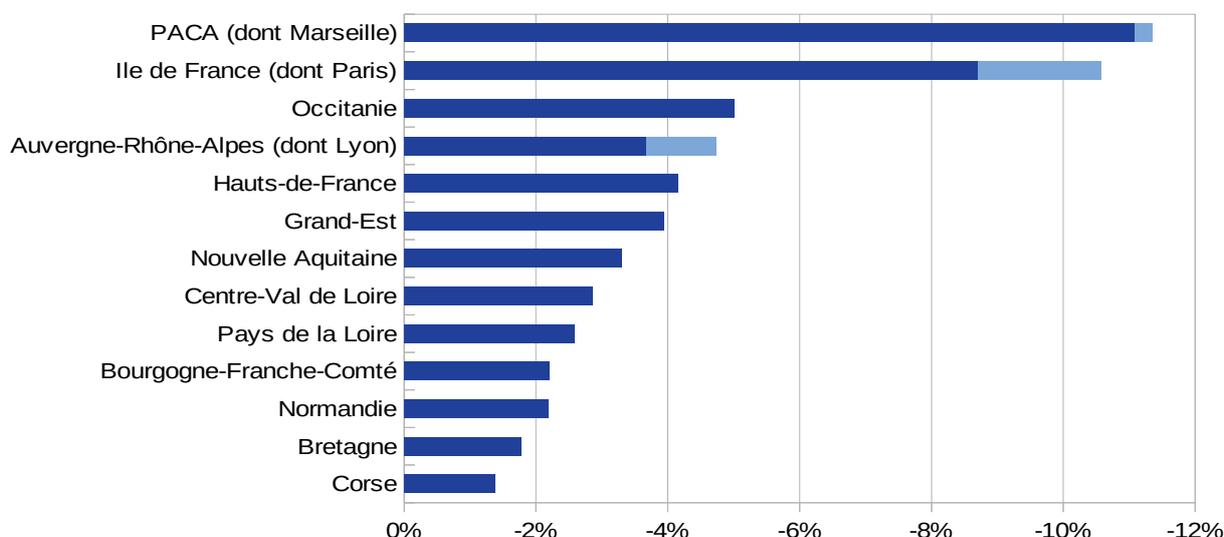


Source : Insee, enquêtes annuelles de recensement de 2004 à 2018 et statistiques d'état civil (comptabilisées au lieu de résidence de la mère).

Ces écarts, plus importants pour les grandes communes, s'expliquent en partie par la mobilité des familles. Car, si les enfants en bas âge sont peu concernés par les migrations hors de France, il n'en est pas de même au niveau local, où la migration des jeunes parents entre en jeu. La mobilité, plus forte aux bas âges, est marquée par un transfert des familles des grandes vers les petites communes : les parents, juste après la naissance, déménageraient souvent des métropoles vers le périurbain pour élever leurs enfants.

Les régions pour lesquelles l'écart entre les naissances au recensement et celles à l'état civil est le plus marqué sont les régions **Provence-Alpes-Côte d'Azur et Île-de-France** (cf. graphique 3-B.3). Les DOM présentent aussi un écart important, hormis la Guyane. Toutes les régions sont néanmoins touchées par le déficit d'enfants en bas âge.

Graphique 3-B.3 : Écarts du nombre de bébés de 0 an entre l'état civil et l'EAR, selon la région



Note : pour les régions Provence-Alpes-Côte d'Azur, Île-de-France et Auvergne-Rhône-Alpes, l'intégralité de la barre représente l'écart du nombre de bébés de 0 an entre l'état civil et l'EAR pour l'ensemble de la région, tandis que la barre foncée représente cet écart sur la région hors la principale commune (Marseille, Paris ou Lyon).

Lecture : pour l'Île-de-France, l'écart est de - 10,6 %, tandis qu'il est de - 8,7 % pour l'Île-de-France hors Paris.

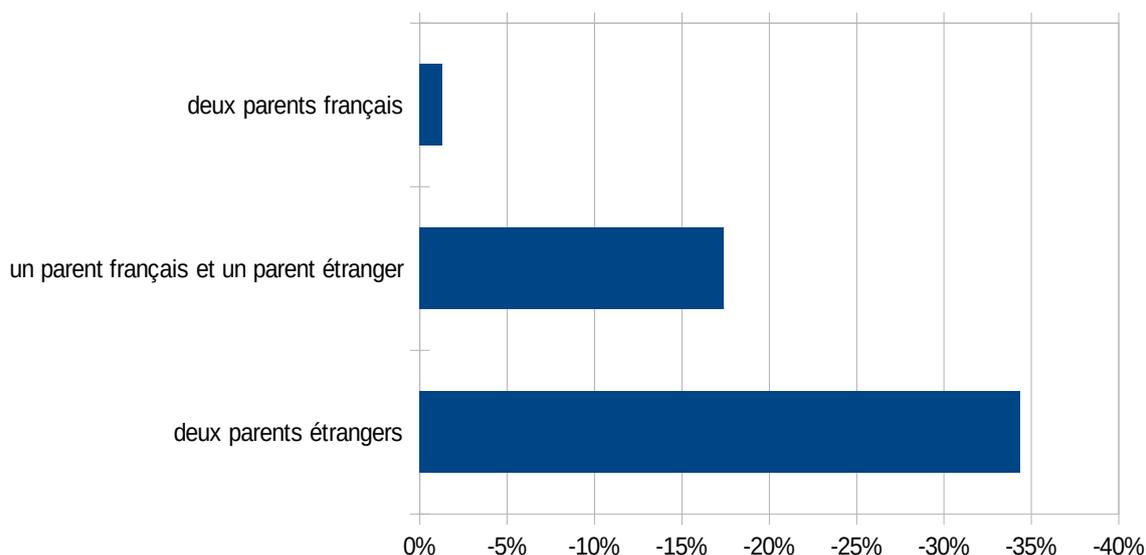
Source : Insee, enquêtes annuelles de recensement de 2004 à 2018 et statistiques d'état civil.

Au niveau régional, il semblerait aussi que, plus les grandes communes pèsent dans l'effectif régional, plus le déficit est grand. En effet, l'Île-de-France et Provence-Alpes-Côte d'Azur sont des régions plus fortement urbanisées que la moyenne des régions françaises.

Des « omissions » d'enfants au recensement plus fréquentes dans les familles étrangères ?

On constate un écart nettement plus conséquent pour les enfants nés de parents étrangers.

Graphique 3-B.4 : Écarts du nombre de bébés de 0 an entre l'état-civil et l'EAR, selon la nationalité des parents



Note de lecture : 1 % des naissances enregistrées dans l'état civil ne sont pas présentes dans le recensement suivant, pour les bébés dont les deux parents sont de nationalité française.

Source : Insee, enquêtes annuelles de recensement de 2004 à 2018 et statistiques d'état civil.

L'écart pour les bébés nés de deux parents français est de près de 1 %, alors qu'il s'élève en moyenne à 34 % pour les bébés nés de deux parents étrangers. Lorsqu'un seul parent est étranger, l'écart est de près de 17 %.

En conclusion, les différents facteurs étudiés permettent de repérer là où les omissions de jeunes enfants semblent les plus nombreuses, mais leur cause n'est pas vraiment identifiée. **Les hypothèses sont donc nombreuses, mais aucune n'a pu être validée ou ne pourrait expliquer à elle seule un sous-dénombrement des jeunes enfants au recensement. Pour les recensements étrangers, les mêmes questions se posent et restent sans réponse.**

Faute de diagnostic plus approfondi, une correction des données du recensement au niveau individuel n'est pas envisageable. Pour effectuer de tels redressements, il faudrait en effet connaître avec précision les déterminants de ces omissions.

Pour affiner ce diagnostic, l'Insee envisage de réaliser une enquête sur la qualité de la mesure de la population consécutive à une enquête annuelle de recensement (cf. conclusion). De telles enquêtes qualité sur le RP ont été menées à deux reprises par le passé : en 1962 et en 1990.

B-4. Impact sur les pyramides des âges

Les pyramides des âges, qui ventilent la population par sexe et âge, sont construites à partir des données des recensements. Le déficit potentiel de jeunes enfants influe donc directement sur ces pyramides. Une méthode de correction des pyramides des âges a été explorée (annexe 2), sur la base des comparaisons entre les données du recensement et celles de l'état civil. Rappelons tout d'abord comment sont construites ces pyramides.

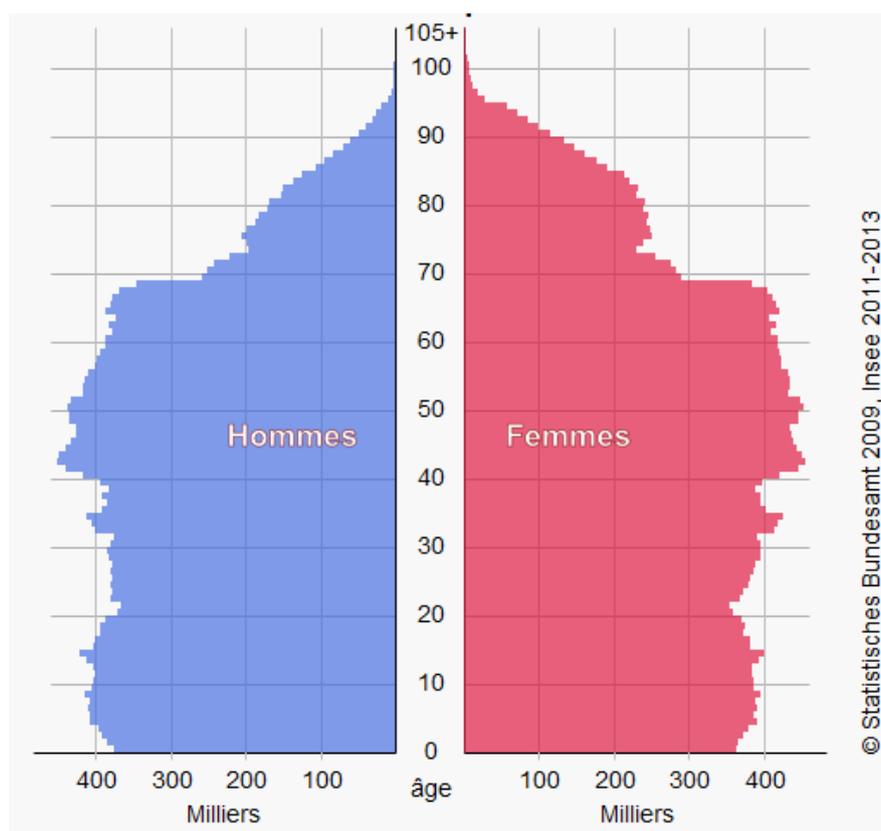
Depuis le recensement en continu, le recensement de la population (RP) cumule les résultats de la collecte de cinq EAR successives ramenées à la situation de l'année médiane : par exemple, le RP 2010 est issu du cumul des EAR 2008 à 2012. Les pyramides des âges définitives sont construites à partir des cinq EAR successives. Les effectifs par âge ne font intervenir que les résultats du recensement de la population. Elles répartissent la population par génération (année de naissance)²⁹, et non par âge atteint au recensement. Il ne s'agit donc pas d'une simple tabulation des données du recensement par âge. Le calcul de la pyramide des âges au 1^{er} janvier de l'année N s'effectue en plusieurs étapes :

- à partir des EAR, on calcule, pour chacune des cinq années de collecte, une répartition de la population par sexe et année de naissance en pondérant les effectifs par la pondération de l'EAR ;
- pour chaque sexe et année de naissance, on calcule les effectifs moyens sur les cinq EAR ;
- des corrections, d'ampleur très faible, sont réalisées avec l'aide des données d'état civil sur les décès, par exemple, pour ne pas tenir compte des personnes recensées en N-3 et N-4, mais décédées avant le 1^{er} janvier N-2 ;
- on applique un coefficient identique à tous les effectifs par sexe et âge pour obtenir le total de la population municipale ; c'est le calage sur la population du cumul RP.

Le graphique 5 présente la pyramide des âges définitive de la France métropolitaine au 1^{er} janvier 2016. Elle représente le nombre d'hommes (à gauche de la figure) et de femmes (à droite) de chaque âge résidant en France. On comprend la difficulté posée par le déficit de jeunes enfants au recensement en analysant la base de la pyramide des âges. Jusqu'à 10 ans environ, la pyramide des âges est rétrécie à sa base, les effectifs allant croissant avec l'âge. La pyramide reflète les naissances passées (les personnes de 10 ans par exemple sont la plupart nées en France, et le nombre de naissances varie d'une année à l'autre, donc le nombre de personnes résidant en France varie d'un âge à l'autre) et les mouvements migratoires (entrées et sorties du territoire). Pourtant, le solde migratoire ne peut expliquer à lui seul le rétrécissement de la pyramide aux jeunes âges. En effet, on constate que les enfants de 0 an au 1^{er} janvier 2015, nés donc en 2014, sont au nombre de 738 452 dans la pyramide des âges. Or, en 2014, l'état civil enregistre 781 167 naissances en métropole et 2 283 décès de nouveau-nés. La pyramide des âges aurait ainsi été « amputée » de 40 000 nouveau-nés, qui refléterait un solde migratoire « potentiel » en 2014 de - 40 000 à cet âge (les sorties du territoire à cet âge excèderaient de 40 000 les entrées). La variation de population entre deux pyramides des âges successives est la somme des naissances diminuée des décès et augmentée du solde migratoire. Ce solde est très élevé par rapport à celui estimé aux âges suivants (5 000 environ en moyenne entre 1 et 6 ans).

29 Document de travail n° F1003, « Construire une pyramide des âges pertinente pour le calcul des indicateurs démographiques à partir des enquêtes annuelles de recensement », S. Jugnot, M. Anguis et C. Beaumel, juin 2010.

Graphique 3-B.5 : Pyramide des âges au 1^{er} janvier 2015



Champ : France métropolitaine.
Source : Insee, estimations de population.

L'Insee a investigué sur les principes d'une méthode de redressement au niveau macro, c'est-à-dire directement au niveau de la pyramide des âges. Toute correction au niveau micro est en effet impossible à ce stade, puisqu'on ne peut estimer précisément ce déficit d'enfants, ni isoler les facteurs qui pourraient l'expliquer (cf. partie 3-B-3). Le principe serait de modifier la répartition par sexe et âge de la population, tout en gardant l'étape de calage sur la population issue du recensement de la population. Il est présenté en annexe 2.

Cette méthode s'appuie fortement sur l'état civil, en supposant donc que le nombre de naissances de mères vivant en France est une donnée robuste. Elle est basée uniquement sur une comparaison macro des effectifs entre l'état civil et les EAR, et ne peut être déclinée à tous les échelons géographiques.

Par ailleurs, si des comparaisons avec d'autres sources administratives tendent à conforter l'hypothèse d'une omission de jeunes enfants au recensement, l'ampleur de celle-ci diffère de celle évaluée à partir des données d'état civil (cf. partie 3-C). Des investigations spécifiques ont été réalisées sur une commune – celles de Grigny (91) – à partir des données de l'EAR 2015 et des données de l'Éducation nationale. Elles ne confortent pas l'hypothèse d'un sous-dénombrement d'enfants par le recensement, et ne permettent pas dès lors d'évaluer son ordre de grandeur ou ses déterminants (cf. encadré 3-C.2).

Au-delà du principe, il paraît alors difficile de savoir de combien il faudrait redresser la pyramide des âges. Dans l'immédiat, l'Insee n'envisage donc pas de mettre en œuvre un tel redressement.

C) Comparaison entre le recensement de la population et les données fiscales : Quelles différences ? Quels enseignements ?

La France n'est pas dépourvue de données administratives sur la population. Depuis 2016, l'Insee a construit et mis à disposition de la statistique publique une base de données statistique sur les logements et sur les individus mobilisant les sources fiscales. Cet assemblage de données administratives, baptisé Fidéli (Fichiers démographiques sur les logements et sur les individus), mobilise les fichiers de la taxe d'habitation, les fichiers des propriétés bâties, les fichiers d'imposition des personnes et les fichiers de déclaration des revenus. Ces données, assemblées au moyen d'identifiants fiscaux sur les foyers et les locaux, constituent le cœur de Fidéli. D'autres données complémentaires contextuelles sont également intégrées : l'appartenance à un IRIS ou un quartier de la politique de la ville, la présence de communautés ou de domiciliation administratives à l'adresse. Enfin, des informations sur le revenu disponible, les grandes masses de revenus déclarés ainsi que sur les montants des prestations sociales sont ajoutées. Fidéli est ainsi conçu pour permettre des analyses statistiques et sert désormais également pour l'échantillonnage des enquêtes ménages de l'Insee.

C.1 - Quelles différences de champ et de localisation entre Fidéli et le recensement ?

Avant de procéder à une comparaison des populations issues du recensement de la population et de Fidéli à différents échelons géographiques, il est important de rappeler comment est construite la source Fidéli et quelles sont les différences de champ ou de définition entre celle-ci et le recensement de la population, différences qui peuvent expliquer une partie des écarts mesurés.

Au niveau des logements

Le répertoire Fidéli couvre tout le bâti provenant des sources fiscales (propriétés bâties et taxe d'habitation), qu'il soit résidentiel ou non. Les locaux d'habitation sont repérés individuellement par un identifiant spécifique.

Par construction, les communautés ne contiennent pas de logements. Elles sont repérées dans Fidéli par un marquage spécifique à l'adresse. Les logements ordinaires non soumis à la taxe d'habitation ne sont pas non plus repérés en tant que tels dans Fidéli. C'est par exemple le cas des foyers logement pour jeunes travailleurs ou personnes âgées. Ces structures ressemblent à des communautés, mais n'en sont pas au sens du recensement, car chaque logement est équipé d'une cuisine et de sanitaires. Les personnes qui y habitent sont néanmoins souvent connues de Fidéli grâce à leur déclaration de revenu.

Le nombre de logements comptabilisés par le recensement est finalement proche au niveau national du nombre de logements de Fidéli : l'écart est de 0,7 % en 2017 en faveur de Fidéli, soit 260 000 logements en plus dans la source administrative. À un niveau plus fin, il y a plus de 3 % de logements de plus dans Fidéli que dans le recensement pour les départements de Seine-Saint-Denis, des Bouches-du-Rhône, du Vaucluse, du Jura et de la Meuse. En revanche, le nombre de logements dans Fidéli est plus faible que dans le recensement de plus de 4 % en Guyane, dans les deux départements de Corse, dans les Hautes-Alpes et en Guadeloupe.

Trop de logements vacants apparaissent cependant dans le Fidéli. Une partie de cet excédent est due à des mises à jour tardives dans les fichiers fiscaux de logements déjà détruits. Une autre partie tient aux règles fiscales : les occupants ont tout intérêt, par exemple en cas de déménagement, à faire en sorte qu'un logement soit considéré comme vacant au 1^{er} janvier afin d'être exonéré de la taxe d'habitation. Les personnes en question ne sont alors connues à l'adresse qu'au travers de l'impôt sur le revenu. Au total,

Fidéli compte environ 0,8 % de résidences principales (au sens de la taxe d'habitation) de moins que le recensement (soit 220 000 résidences principales).

Au niveau des habitants

Fidéli couvre toutes les personnes appartenant à un foyer fiscal sur l'ensemble du territoire français.

Au niveau des habitants, des filtres sont utilisés pour exclure au mieux les personnes qui ne résident pas en France, car cette information n'est pas directement disponible dans les sources fiscales. Par ailleurs, comme plusieurs sources fiscales sont mobilisées, Fidéli effectue des traitements d'appariements entre ces sources sur la base des noms, prénoms, lieu de naissance et identifiant fiscal, encadrés juridiquement. Ces appariements visent à repérer le plus précisément possible les individus distincts. Dans certains cas, ces appariements s'avèrent plus complexes, certains enregistrements étant plus flous (patronyme partiel, lieu de naissance non renseigné, etc.).

Par ailleurs, dans les fichiers fiscaux, il manque des jeunes adultes, majoritairement de 20 à 22 ans. Une des raisons est probablement qu'ils ont quitté le foyer parental et vivent en communauté, ou du moins dans un logement qui n'est pas soumis à la taxe d'habitation, et qu'ils n'apparaissent sur aucune déclaration de revenu. Lorsqu'ils peuvent être retrouvés dans les fichiers fiscaux des années antérieures, ils sont ajoutés au millésime courant. Leur localisation est sans doute médiocre, mais cela permet de corriger un léger creux dans la pyramide des âges au niveau national.

Outre les personnes vivant en communauté ou dans des logements non assujettis à la taxe d'habitation, certains contribuables ne sont connus qu'au travers d'une adresse de domiciliation, donc sans logement identifié (demandeurs d'asile, gens du voyage, sans domicile fixe...). Lorsque ces personnes ne font pas de déclaration de revenus, elles ne sont pas présentes dans les fichiers Fidéli.

En matière de localisation des personnes

Fidéli enregistre toutes les adresses connues pour une personne donnée. Dans les données fiscales de 2016, 7,3 millions d'adultes, soit 13 % d'entre eux, ont ainsi plus d'une résidence déclarée.

Ainsi, l'adresse déclarée comme principale dans les fichiers de la taxe d'habitation (TH) est différente de l'adresse déclarée à l'impôt sur le revenu dans 17 % des cas de multi-résidence avec environ 820 000 personnes localisées (soit 11 % des cas) dans des communes différentes entre ces deux fichiers. La localisation retenue comme prioritaire est alors celle enregistrée comme résidence principale dans les fichiers de la taxe d'habitation.

En matière de localisation géographique, si, dans la plupart des cas, la résidence principale déclarée au recensement coïncide avec la résidence principale au sens de la TH prise comme référence dans Fidéli, ce n'est pas le cas pour certaines populations spécifiques :

- les **étudiants majeurs** vivant dans un logement non soumis à la taxe d'habitation (cité universitaire par exemple) seront localisés chez leurs parents dans Fidéli si ces derniers les font figurer sur leur déclaration de revenu, alors qu'ils seront recensés dans leur logement situé sur leur lieu d'étude selon les règles du recensement³⁰ ;
- les **enfants en résidence partagée** sont comptés dans Fidéli pour moitié dans chacune des communes de leurs parents alors que le recensement les compte là où ils passent le plus de temps (et en cas d'égalité là où ils se trouvent la nuit précédant le début du recensement) ;
- les **personnes occupant un logement occasionnel** (par exemple les célibataires géographiques) pourront déclarer ce logement comme résidence principale à la TH alors que, pour le recensement, ils seront comptés dans leur domicile familial (celui-ci pouvant même éventuellement être à l'étranger).

30 Pour les étudiants mineurs, les règles du recensement prévoient de les compter dans la commune de domiciliation de leurs parents, ce qui est conforme à leur localisation dans Fidéli.

- La source fiscale contient également des personnes connues de l'administration fiscale à leurs **domiciliations administratives** qui ne correspondent pas nécessairement à la commune dans laquelle ils sont recensés (par exemple les gens du voyage, les sans-abris).
- pour des raisons fiscales, certaines personnes choisissent de **localiser fiscalement leur résidence principale dans un logement qui est une résidence secondaire** au sens du recensement.

C.2 - Au niveau national, seulement 0,6 % d'écart sur la population française mesurée par les deux sources en 2017

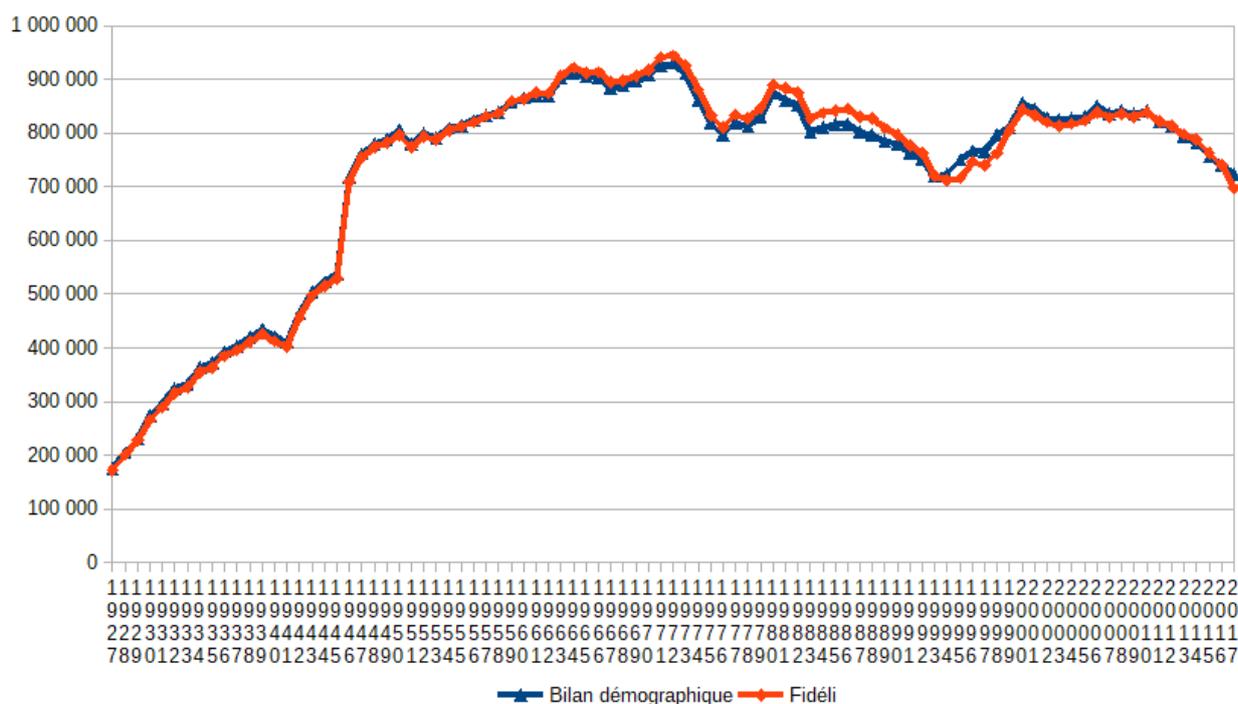
En 2017, d'après le recensement, la population française s'établit à 66 524 000 habitants (métropole et DOM hors Mayotte). Cette population respecte les critères européens et internationaux de population usuelle résidente (Unece, 2015). Selon ces critères, pour être compté comme habitant d'un pays (ou d'une commune), il faut que la personne ait sa résidence usuelle dans ce pays (ou commune), c'est-à-dire qu'elle y passe habituellement ses périodes de repos et qu'elle y réside ou envisage d'y résider au moins un an. En particulier, les étudiants Erasmus ou les travailleurs venus résider quelques mois en France ne sont pas comptabilisés dans la population recensée.

Fidéli comptabilise de son côté 66 948 000 personnes en 2017 sur le même champ, soit un écart de 0,6 %. Il y a donc environ 400 000 personnes de plus dans Fidéli par rapport au recensement.

Pour affiner la comparaison, le graphique 3-C.1 présente l'écart de population entre Fidéli et le bilan démographique par âge fin. La comparaison est faite sur l'année 2018. Il est plus pertinent de s'appuyer sur les données du bilan démographique que sur celles du recensement pour faire une comparaison par âge, dans la mesure où le recensement de la population s'appuie sur 5 enquêtes annuelles de recensement, ce qui tend à lisser la taille des populations entre les différentes générations (Insee, 2018b). Cela est retraité dans le cadre du bilan démographique.

Les profils par âge des populations issues des deux sources sont très proches (cf. graphique 3-C.1).

Graphique 3-C.1 : Population selon la génération dans Fidéli d'une part, et le bilan démographique d'autre part, au 1^{er} janvier 2018



Source : Insee, Fidéli 2018 et bilan démographique.
 Champ : France entière (y compris Mayotte).

Plus précisément (cf. tableau 3-C.1), on constate que Fidéli compte un peu moins d'enfants que le bilan démographique (- 0,2 % pour les moins de 5 ans et - 0,4 % entre 6 et 9 ans) et sensiblement moins de jeunes de 10 à 17 ans (- 1,1 %) et de jeunes adultes de 18 à 24 ans (- 2,4 %). À l'inverse, il y a davantage de personnes âgées de 25 à 54 ans dans Fidéli que dans le bilan démographique.

Plus précisément, Fidéli enregistre un déficit de 27 000 enfants âgés de 0 an par rapport au bilan démographique. En comparaison avec les données de l'état civil, le déficit est alors encore plus important : de l'ordre de 70 000. L'écart est donc plus important qu'entre le recensement et l'état civil (cf. partie 3-B). Ensuite, le nombre d'enfants d'un an est très proche entre les deux sources, et Fidéli compte plus d'enfants que le bilan démographique entre 2 et 5 ans (18 000 d'écart en cumulé). Ce léger surplus pourrait correspondre à un défaut de déclaration des enfants au recensement de la population, défaut évoqué dans la partie 3-B et qui n'est pas propre à la France (Census bureau, 2019 et Toulemon, 2017). Cependant, d'après ces chiffres, la sous-estimation d'enfants au recensement pourrait être moins forte que l'estimation réalisée à la partie 3-B, à moins que le nombre d'enfants soit également sous-estimé à partir des sources fiscales, comme il semble l'être pour les enfants de 0 an.

Par ailleurs, dans certains cas, il est également possible que des enfants soient rattachés fiscalement à des foyers fiscaux sans être effectivement présents (cf. encadré 3-C.2).

Tableau 3-C.1 : Écarts de population par tranche d'âge entre Fidéli et le bilan démographique, au 1^{er} janvier 2018

	Bilan démographique	Fidéli	Écart absolu Fidéli/bilan démographique	Écart relatif Fidéli / bilan démographique (en %)
0 à 5 ans	4 607 468	4 599 414	-8 054	-0,2
6 à 9 ans	3 339 668	3 325 701	-13 967	-0,4
10 à 17 ans	6 694 070	6 618 784	-75 286	-1,1
18 à 24 ans	5 331 462	5 202 930	-128 532	-2,4
25 à 34 ans	7 920 402	8 154 805	234 403	3,0
35 à 44 ans	8 432 449	8 605 587	173 138	2,1
45 à 54 ans	9 051 352	9 154 681	103 329	1,1
55 à 64 ans	8 369 055	8 364 914	-4 141	0,0
65 à 74 ans	6 994 431	6 925 380	-69 051	-1,0

Source : Insee, Fidéli 2018 et bilan démographique.

Champ : France entière (y compris Mayotte).

La population issue de Fidéli apparaît également en déficit pour les jeunes adultes, comme cela était évoqué dans la partie précédente. Ce déficit de 128 000 personnes entre 18 et 24 ans peut s'expliquer notamment par le fait que ces jeunes adultes ne sont plus déclarés dans la déclaration de revenus de leurs parents et ne réalisent pas encore eux-mêmes leur propre déclaration de revenus. S'ils vivent dans un logement non soumis à la TH ou chez des amis, ils peuvent alors n'être présents dans aucun fichier fiscal.

La population Fidéli présente en revanche un excès de 408 000 personnes âgées de 25 à 44 ans. Une des causes de cet excès pourrait être que, malgré les filtres mis en place pour les repérer, les données fiscales contiennent des travailleurs non-résidents pour ces tranches d'âge.

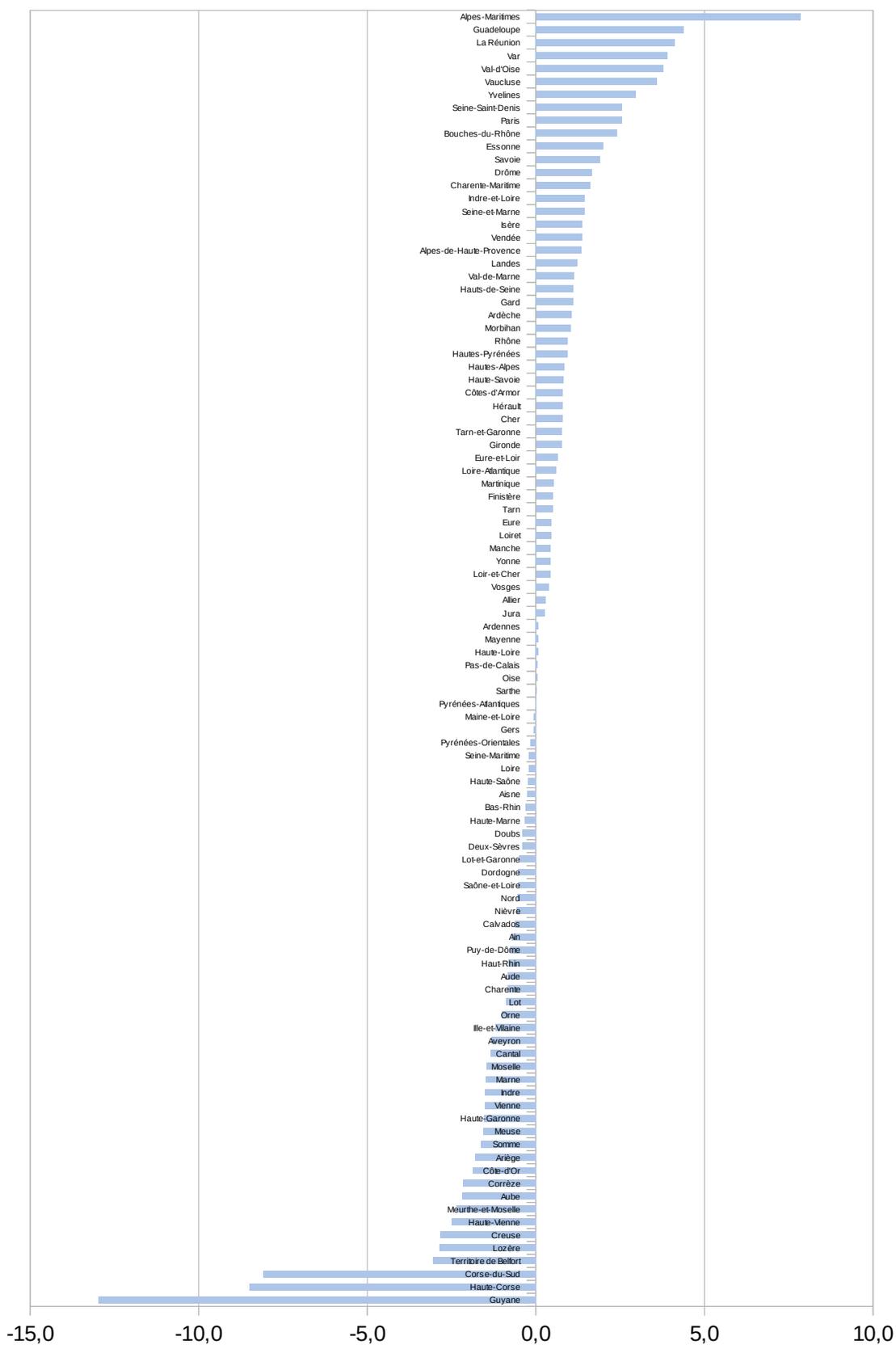
C.3 - Des écarts entre le recensement et Fidéli de signes et d'ampleur différents selon les départements

Les écarts de population selon les départements entre le recensement et Fidéli se situent entre - 13 % (Guyane) et + 7,8 % (Alpes-Maritimes). Hors Dom, Alpes-Maritimes et Corse, la fourchette est plus resserrée, comprise entre - 3 % pour le Territoire de Belfort et + 3,9 % pour le Var.

Ces écarts sont inférieurs en valeur absolue à 1 % dans la moitié des départements, et ils excèdent 3 % dans 10 départements.

En particulier, la population au recensement est supérieure à la celle de Fidéli de plus de 3 % dans quatre départements : la Guyane, la Corse du Sud, la Haute-Corse et le Territoire de Belfort. À l'inverse, les populations fiscales sont supérieures de plus de 3 % aux populations recensées dans six départements : les Alpes-Maritimes, la Guadeloupe, La Réunion, le Var, le Val-d'Oise et le Vaucluse.

Graphique 3-C.2 : écart entre le nombre d'habitants au recensement et dans Fidéli par département en 2017



Source : Insee, Fidéli 2017 et RP 2017.
 Champ : France hors Mayotte.

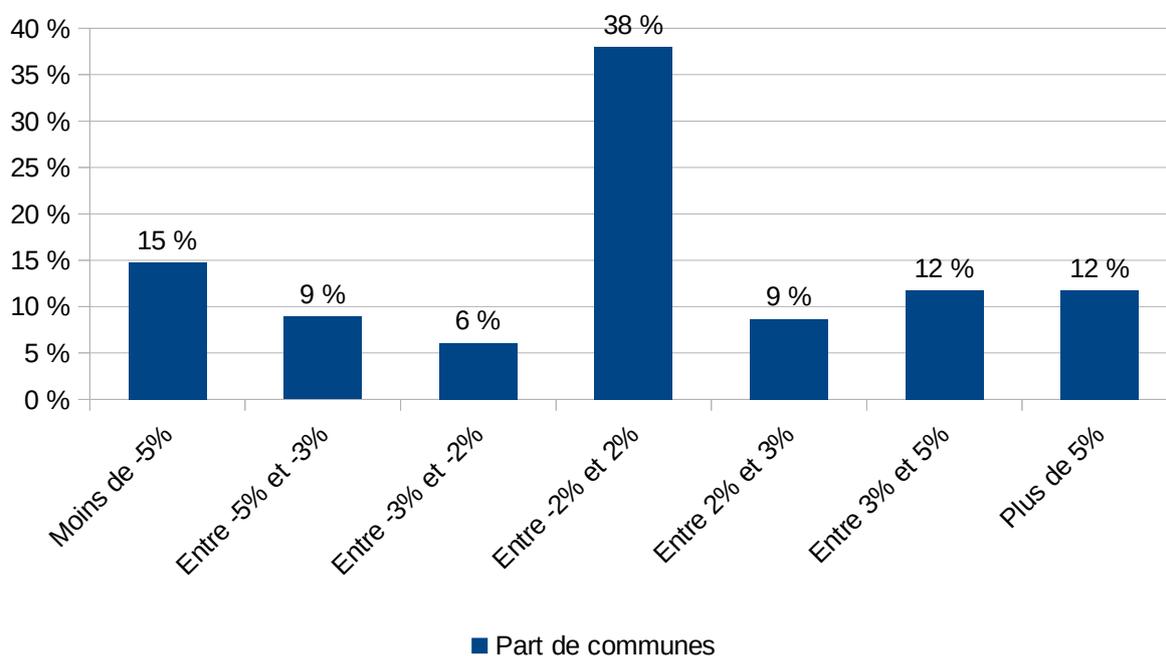
C.4 - Des écarts plus marqués au niveau communal, notamment dans les plus petites communes

Les écarts de population entre recensement et sources fiscales peuvent être plus importants au niveau communal, compte tenu du nombre élevé de communes et de leur hétérogénéité en matière de taille.

Les écarts semblent plus importants pour les plus petites communes, celles de moins de 2 000 habitants. La raison la plus généralement évoquée est que les différences de concepts entre les deux sources et les retards de mise à jour des données fiscales se compensent moins facilement dans les communes de petite taille.

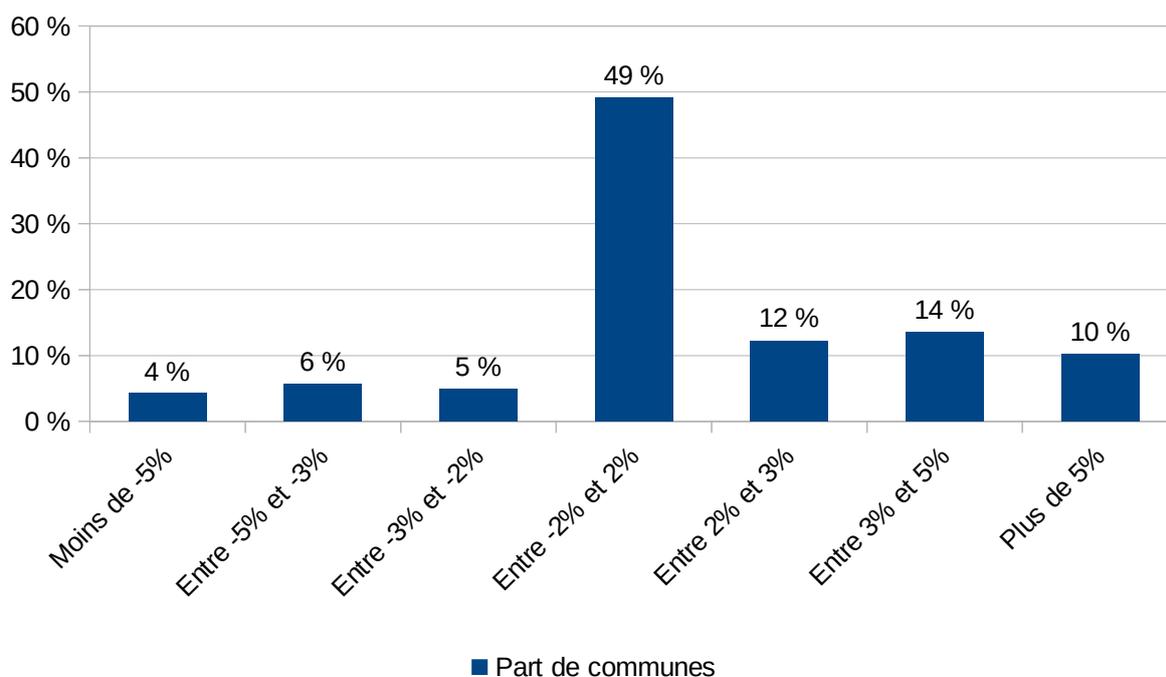
Ainsi, 38 % des 35 000 communes de France hors Mayotte ont un écart entre la population mesurée par Fidéli et le recensement de moins de 2 % en valeur absolue. Mais 62 % des communes présentent un écart de plus de 2 % et 27 % de plus de 5 %. Si on se restreint aux communes de plus de 2 000 habitants au recensement, la distribution des écarts entre les deux sources est plus concentrée autour de 0 %. Sur ce champ, 49 % des communes présentent un écart de moins de 2 % en valeur absolue, et 14 % des communes présentent un écart de plus de 5 %.

Graphique 3-C.3 : Distribution des écarts communaux entre Fidéli et le recensement en 2017 – Ensemble des communes



Source : Insee, Fidéli 2017 et RP 2017.
Champ : France hors Mayotte.

Graphique 3-C.4 : Distribution des écarts communaux entre Fidéli et le recensement en 2017 – Communes de plus de 2 000 habitants



Source : Insee, Fidéli 2017 et RP 2017.
 Champ : France hors Mayotte.

Les travaux d'analyse ont montré par ailleurs qu'une partie des écarts entre sources était due aux personnes vivant en communautés, notamment les jeunes qui, lorsqu'ils ne font pas de déclaration de revenus, ne peuvent être localisés par la source fiscale qu'au domicile de leurs parents (puisque'ils ne paient pas de taxe d'habitation).

Même en ne s'intéressant qu'aux communes de plus de 100 000 habitants, on constate encore des écarts marqués entre sources, parfois positifs, parfois négatifs. En effet, sur les 50 communes concernées, seules 16 présentent des écarts inférieurs à 3 % en valeur absolue. Ce nombre double si on traite au cas par cas les populations vivant en communautés. Reste néanmoins une petite moitié de ces grandes communes pour lesquelles les écarts sont significatifs. Ainsi, par exemple, la population Fidéli est nettement inférieure à la population recensée à Caen (- 7,3 %), Dijon (- 7 %), Rennes (- 6,6 %) ou dans le 16e arrondissement de Paris (- 6,8 %). À l'inverse, la population Fidéli est souvent supérieure à la population recensée en région parisienne, par exemple à Paris 19e (10 %), 12e (8,8 %), à Argenteuil (7,1 %), Montreuil (6,8 %) ou encore à Saint Denis (5 %). C'est le cas aussi à Nice (10,9 %).

La comparaison entre recensement et sources fiscales montre ainsi que, si les résultats en nombre d'habitants sont proches au niveau national, ils peuvent sensiblement s'écarter au niveau local, probablement en raison de différences de localisation de la résidence principale retenue par les deux sources ou de champs pas complètement homogènes (cf partie 3-C.1). Les écarts ne sont ainsi pas systématiquement de même signe. On peut constater qu'en Île-de-France, la population recensée est plutôt inférieure à celle mesurée par Fidéli, tandis que dans les grandes agglomérations de province c'est souvent l'inverse.

Des analyses économétriques ont été menées au niveau des communes et des départements pour tenter d'extraire des facteurs explicatifs des écarts observés. Deux variables ont particulièrement été analysées : la part de résidences secondaires et la part d'étrangers. Ces analyses n'ont pas mis en évidence de corrélation forte entre les écarts mesurés entre sources et ces deux variables.

C.5 - Une autre approche possible pour comparer recensement et données fiscales : retrouver les personnes de l'échantillon EDP recensées en 2018 dans Fidéli 2018

L'objectif de notre analyse a été de partir des bulletins de l'enquête de recensement de 2018 et de voir si l'on retrouvait bien ces personnes dans Fidéli, puis pour les personnes retrouvées de comparer leur lieu de résidence.

L'appariement est conduit sur la totalité du champ des personnes « exploitables », i.e. dont le nom, le prénom, l'année de naissance et le lieu de naissance sont connus. L'appariement est réellement utilisé seulement pour les seules personnes majeures, nées avant 2001. Sur ce champ restreint, qui compte 771 884 individus, 95,3 % des individus disposent d'informations d'état civil exploitables.

Sur le champ des personnes mineures, l'appariement est moins bon, car, dans les fichiers fiscaux, les informations détaillées sur l'état civil ne sont pas toujours disponibles pour ces personnes.

De nombreuses opérations préalables sont nécessaires avant d'apparier les données en raison de problèmes lexicographiques, de l'utilisation du nom d'épouse et d'un seul prénom dans les réponses au recensement de la population.

Au total, 94 % des personnes appartenant à l'échantillon EDP (cf. encadré 3-C.1) et recensées ont pu être retrouvées dans les sources fiscales. Les disparités sont nettes selon l'âge. Sans surprise, les taux d'appariement sont moindres avant 30 ans, âges où des personnes encore en formation peuvent résider en communauté (résidence étudiante...) sans être connues des services fiscaux. Le taux d'appariement est de 86 % pour les moins de 30 ans. Les taux d'appariements sont un peu plus faibles chez les femmes au-delà de 30 ans, en raison des difficultés à repérer leur patronyme sous une forme équivalente dans les deux sources. Certains taux sont en revanche particulièrement élevés, aux alentours de 97 % pour les hommes proches de la cinquantaine. Il est ainsi difficile d'en tirer des conclusions claires sur l'absence de personnes recensées dans Fidéli, la majorité étant probablement due à des défauts de qualité des données servant à l'appariement.

Encadré 3-C.1 : présentation succincte de l'échantillon démographique permanent (EDP)

L'échantillon démographique permanent est un panel socio-démographique de grande taille, mis en place en France, pour étudier la fécondité, la mortalité, les parcours familiaux, les migrations géographiques au sein du territoire national, la mobilité sociale et la mobilité professionnelle, les carrières salariales et les niveaux de vie ainsi que les interactions possibles entre ces différents aspects.

L'échantillon démographique permanent a été créé en 1967.

Il contient des informations issues de cinq sources :

- des bulletins d'état civil de naissance, de mariage, de décès depuis 1968
- des recensements de 1968, 1975, 1982, 1990 et 1999 puis des enquêtes annuelles de recensement à partir de 2004
- du fichier électoral depuis 1967
- du panel " tous salariés " depuis 1967
- des données socio-fiscales depuis 2011 (revenus 2010), notamment venant du dispositif Filosofi.

Pour les personnes appariées se pose la question de la correspondance de la commune de résidence entre les deux sources (recensement et Fidéli). On se limite ici au seul champ des personnes EDP appariées

précédemment et pour lesquelles le code de la commune (ou d'arrondissement) est renseigné dans le recensement, soit environ 282 000 individus. Il en résulte que :

- **pour 94,1 % des personnes, le code commune (ou arrondissement) du recensement correspond au code commune (ou arrondissement) de Fidéli (issu de la taxe d'habitation ou à défaut impôt sur le revenu) ;**
- pour 1,6 % des personnes, le code commune du recensement est retrouvé dans les déclarations fiscales des individus, mais à une autre localisation que celle de Fidéli. Il s'agit dans 86 % des cas de la localisation d'une résidence secondaire, sinon d'une incohérence interne aux fichiers fiscaux (FIP et PLFC) sans doute due à un retard de mise à jour ;
- pour 0,8 % des personnes, le code commune du recensement est celui de la dernière adresse connue, en général une adresse destinée à remplacer l'adresse Fidéli au-delà du 1^{er} janvier (notamment en cas de déménagement lorsque la nouvelle adresse est connue par anticipation, par exemple en cas d'emménagement au-delà du 1^{er} janvier pour être exonéré de la taxe d'habitation) ;
- pour 1 % des personnes, le code commune de Fidéli est différent du code commune actuel du recensement, mais correspond au code commune du recensement l'année précédente ; les deux sources sont cohérentes, les écarts provenant d'un décalage de mise à jour des fichiers fiscaux ;
- pour 2,5 % des individus, la commune diffère entre les deux sources sans qu'il y ait trace de la commune au recensement présente ou passée dans les fichiers fiscaux pour la personne en question.

Comme on pouvait s'y attendre, les taux de correspondance sont meilleurs en ménage ordinaire.

Il en résulte que la localisation des personnes en logement ordinaire est proche dans le recensement et Fidéli, sans toutefois coïncider exactement. Une partie des écarts est sans doute due à des délais de mise à jour de la source fiscale, ou à des choix spécifiques conduisant à ce que la personne déclare au fisc comme résidence secondaire sa résidence principale au recensement. Au-delà de 40 ans, il est très rare que la localisation au recensement soit inconnue dans la source fiscale. Ces écarts tendent toutefois à expliquer la divergence de population observée entre les sources à des niveaux locaux.

Encadré 3-C.2 : Une comparaison recensement / fichiers de l'Éducation nationale sur la commune de Grigny

Une autre expérience de comparaison entre le recensement et un fichier administratif a été menée sur la commune de Grigny (91), suite à des doutes répétés émis par cette commune sur la qualité de la population recensée. Cette étude a permis de comparer au niveau individuel les résultats de l'enquête annuelle de recensement de 2015 et les informations figurant dans le fichier du Rectorat de Versailles pour l'année scolaire 2015-2016.

Dans la mesure où les comparaisons ont été effectuées à partir de données individuelles nominatives, la procédure d'obtention par l'Insee des fichiers de l'Éducation nationale a obéi aux dispositions réglementaires en vigueur, à savoir une demande d'accès aux données au titre de l'article 7bis de la loi du 7 juin 1951 modifiée adressée au Conseil national de l'information statistique (Cnis) et une autorisation de traitement accordée par la Commission nationale informatique et des libertés (Cnil) le 27 novembre 2015.

L'objectif de ce travail était d'identifier si des enfants manquaient au recensement par rapport aux informations présentes dans les bases de l'Éducation nationale en lien avec la problématique développée dans ce document dans la partie 3-B.

La comparaison a reposé sur l'enquête annuelle de recensement de 2015 de la commune de Grigny, dont l'échantillon était composé de 99 adresses et 792 logements. Dans ces logements, se trouvaient 365 bulletins individuels d'enfants nés entre le 1^{er} janvier 2005 et le 1^{er} janvier 2011 (c'est-à-dire âgés de 3 à 10 ans) ou nés en 2004 ou 2012 et scolarisés dans les classes observées (afin de prendre en compte les

enfants scolarisés précocement et ceux ayant redoublé). Ces enfants étaient domiciliés au sein de 228 logements dans 47 adresses.

La méthode de travail a consisté à comparer, pour toutes ces adresses, le nombre d'enfants recensés et le nombre d'enfants scolarisés d'après l'Éducation nationale. Les comparaisons ont aussi reposé sur l'identité des enfants présents dans les deux sources (nom, prénom, date de naissance).

Le premier résultat brut montre que 365 enfants ont été recensés à ces adresses alors que 252 sont enregistrés sur ces mêmes adresses dans la base de l'Éducation nationale de la commune de Grigny.

Plus précisément : 196 enfants ont été retrouvés aux adresses de l'échantillon dans les deux sources, 125 enfants étaient présents dans l'EAR et absents de la base Éducation nationale sur les adresses de l'échantillon et 56 enfants étaient absents de l'EAR et présents dans la base Éducation nationale. Par ailleurs, pour 44 enfants, la date de naissance n'était pas suffisamment complète dans l'EAR pour qu'on puisse les repérer dans le fichier de l'Éducation nationale.

Une recherche plus poussée a ensuite été menée sur les 56 enfants non présents dans le recensement. Quatre adresses regroupant 28 enfants ont été particulièrement analysées. Pour 4 enfants, le logement a bien été recensé avec un nom de famille correspondant à l'enfant tel que connu à l'Éducation nationale, mais aucun bulletin individuel n'a été rempli pour cet enfant³¹. Les 24 autres cas correspondent à des noms de famille non recensés à l'adresse. Il est possible que des logements n'aient pas été enquêtés et aient fait l'objet d'une procédure d'estimation sans connaître le nom de la famille (voir redressement des fiches de logement non enquêté - partie 2-D), que des logements aient été inscrits à tort en logement vacant ou aient été oubliés ou que l'adresse connue par l'Éducation nationale pour ces enfants ne soit plus d'actualité au moment de l'EAR en raison d'un déménagement en cours d'année (les adresses de la base éducation nationale étant renseignées en septembre 2014).

Parmi les 125 enfants présents dans le recensement mais pas dans la base Éducation nationale dans les adresses de l'échantillon, 20 ont été retrouvés sur une autre adresse de Grigny (ces enfants sont enregistrés par l'Éducation nationale à une adresse hors de l'échantillon de l'EAR 2015, peut-être en raison d'un déménagement), et 105 enfants sont absents du fichier de l'Éducation nationale de Grigny. En élargissant la recherche aux communes limitrophes, on en retrouve 88 dans une autre commune. Les écarts observés entre les données de l'EAR et celles de la base Éducation nationale ne doivent donc en aucun cas être interprétés en termes d'enfants déscolarisés.

Plus globalement, dans les fichiers de l'Éducation nationale, de la maternelle au CM1, le nombre d'enfants habitant Grigny quel que soit leur lieu de scolarisation est de 3 459. D'après le recensement, en 2015, on en a 4 088 entre 3 et 9 ans ou 4 028 entre 4 et 10 ans. **Le nombre d'enfants à Grigny est donc plus élevé dans le recensement que dans les fichiers de l'Éducation nationale.** Il convient toutefois de noter que certains enfants en âge d'être scolarisés en classe maternelle peuvent ne pas y aller, la scolarisation n'étant pas obligatoire à cet âge en 2015.

Si l'on se restreint au champ de l'école primaire, les résultats sont plus proches : 2 150 enfants nés entre 2005 et 2008 dans les fichiers de l'Éducation nationale de l'année scolaire 2014-2015, 2 280 enfants dans le recensement âgés de 6 à 9 ans et 2 189 âgés de 7 à 10 ans.

Cette analyse pointue menée sur une commune spécifique ne permet donc pas de conforter l'hypothèse d'une omission d'enfants dans le recensement : en effet, les estimations pour cette commune restent cohérentes entre recensement et Éducation nationale. Au-delà, elle montre combien il est difficile de comparer deux sources de données différentes.

31 À noter : jusqu'à l'EAR 2017, une personne qui était déclarée dans la feuille de logement mais sans bulletin individuel associé n'était pas comptabilisée dans le décompte de population, les informations recueillies étant jugées trop pauvres. Depuis l'EAR 2018 et le changement du questionnaire logement, toutes les personnes déclarées dans la feuille de logement avec leurs nom, prénom, sexe et date de naissance sont comptabilisées même si aucun bulletin individuel n'a été rempli.

Conclusion

La grande implication des acteurs (communes et Insee) dans la préparation et la réalisation du recensement induit une très bonne maîtrise du processus. Chaque année, le recensement mobilise environ 450 ETP à l'Insee. Le grand nombre de contrôles réalisés tout au long de ce processus permet de garantir la qualité des estimations produites. Par ailleurs, le recensement bénéficie d'un très faible taux de non-réponse (3,9 % en 2019). Certains éléments échappent cependant à cette maîtrise. Même si de nombreuses consignes sont présentes sur les questionnaires du recensement, les réponses des personnes recensées ne sont pas parfaites, du fait parfois d'une mauvaise compréhension, d'une inadaptation de certaines questions à des situations réelles ou de mauvaises réponses délibérées.

Le recensement de la population évolue régulièrement pour s'adapter aux évolutions de la société, gagner en efficacité et en qualité (Roux, 2020). Ainsi, le questionnaire du recensement a été modifié à deux reprises ces dernières années : en 2015, où trois questions du bulletin individuel du recensement ont été modifiées, et en 2018 où la feuille de logement a été modifiée afin de recueillir plus d'informations sur les situations de multi-résidence. Les protocoles de collecte ont également évolué, avec la généralisation de la collecte par Internet à partir de 2015 et celle du dépôt des identifiants de connexion aux questionnaires Internet dans les boîtes aux lettres des adresses individuelles à partir de l'an prochain.

Les éléments quantitatifs dont l'Insee dispose aujourd'hui pour mesurer la qualité des estimations de population, présentés tout au long de ce document, restent partiels. L'enquête de mesure de la qualité du RIL représente la principale source d'informations sur les excédents et les déficits de logements observés sur ce répertoire : en 2017, le RIL présente un déficit net d'au plus 0,9 % de logements habitables. Des éléments sur la variance liée au sondage des estimations de population dans les communes de plus de 10 000 habitants sont également disponibles. En revanche, l'Insee ne dispose pas d'éléments sur d'autres composantes, comme les erreurs de mesure sur les habitants permanents d'un logement, la catégorie de logement, l'exhaustivité des logements dans les petites communes.

Peu d'éléments existent à ce jour pour évaluer de manière globale les éventuelles sur-estimations ou sous-estimations de la population. Plusieurs analyses, présentées dans la troisième partie de ce document, explorent la voie d'une comparaison des résultats du recensement avec d'autres sources de données. Si elles amènent à fournir des données chiffrées sur d'éventuelles sur-estimations ou sous-estimations, elles reposent soit sur des hypothèses que l'on peut discuter, soit sur des sources dont la qualité peut également être questionnée. Ces résultats ne placent pas le recensement français dans une position singulière par rapport aux recensements des autres pays quand on les compare avec la qualité des recensements au niveau international.

Tous ces résultats mériteraient toutefois d'être consolidés. Pour cela, une mesure directe peut être envisagée par la réalisation d'une enquête sur la qualité du recensement. Une telle enquête a été réalisée à deux reprises en France, après les recensements de 1962 et de 1990. Elles ont permis de mettre en avant les excédents et les déficits de la mesure de la population *via* les recensements passés. L'enquête de 1990 a estimé à près de 1 % le nombre de doubles-comptes et de 1,5 % à 2 % le nombre d'omissions (Coeffic, 1993), ce qui confirme les estimations de l'enquête de 1962. L'enquête de mesure de la qualité du recensement consiste à retourner dans certains logements recensés quelques semaines après le recensement et à interroger en face-à-face le ménage afin d'observer les écarts de déclaration avec les questionnaires du recensement. Cette méthode présente également quelques défauts (biais de déclaration toujours présent, biais lié au mode de collecte, reformulation des questions par l'enquêteur, etc.), mais est *a priori* celle qui permet d'obtenir la meilleure mesure de la vraie valeur que l'on cherche à obtenir, ici le nombre d'habitants permanents dans les résidences principales.

Une telle enquête de mesure de la qualité du recensement n'a jamais été menée depuis que le nouveau recensement a été mis en place en 2004. Le recensement tournant peut compliquer sa mise en œuvre. Par

exemple, la notion de doubles-comptes est plus difficile à appréhender dans la mesure où l'on n'a pas une photographie de tout le monde au même instant. Elle permettrait cependant d'estimer les omissions et le nombre de personnes comptées à tort, *i.e.* qui sont comptabilisées comme habitants permanents d'un logement alors qu'elles ne le devraient pas (résidence non principale, multi-résident devant être comptabilisé dans un autre logement, etc.). Il sera en revanche difficile d'en déduire un nombre de « doubles-comptes », car une personne comptée à tort peut ne pas conduire à une double comptabilisation dans le cas où elle est en omission dans le logement où elle devrait être comptabilisée. Ces informations seraient toutefois intéressantes pour quantifier la qualité des estimations de population par le recensement.

La réalisation d'une telle enquête de mesure de la qualité du recensement est en cours d'instruction au sein de l'Insee et pourrait voir le jour dans les années à venir, si les ressources de l'Institut le permettent. La validation d'autres constatations réalisées dans ce document (par exemple, l'omission probable de jeunes enfants) ne serait pas toujours possible du fait de la taille de l'échantillon d'une enquête de mesure de la qualité. En effet, les écarts mis en évidence portent sur des effectifs faibles, si bien qu'un échantillon de très grande taille serait nécessaire pour produire des résultats suffisamment précis. Il faudrait alors adapter l'échantillon en surreprésentant certaines sous-populations ou sur certains territoires.

Les travaux de comparaison avec d'autres sources – sociales ou fiscales – sont à poursuivre pour mieux documenter les écarts observés et pour les comprendre.

Bibliographie

Algava É., Penant S., Yankan L., « En 2016, 400 000 enfants alternent entre les deux domiciles de leurs parents séparés », *Insee Première* n° 1728, 2019.

Brilhault G., Caron N., « Le passage à une collecte par sondage : quel impact sur la précision du recensement ? », *Économie et statistique* n° 483-484-485, 2016.

Buisson G., Lapinte A., « Vivre dans plusieurs configurations familiales », *Insee première* n° 1647, 2017.

Cases C., Clanché F., Frécon J.-C., Oger P., « Évolution du questionnaire du recensement de la population », *Rapport du Cnis* n° 130, août 2012.

Census Bureau, « Investigating the 2010 Undercount of Young Children – Summary of Recent Research », *Report*, 2019. <https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/final-analysis/2020-report-2010-undercount-children-summary-recent-research.html>

Census Bureau, « Summary of Estimates of Coverage for Persons in the United States », *Census Coverage Measurement Estimation Report*, Washington U.S. Census Bureau, 2012. <https://www.census.gov/programs-surveys/decennial-census/about/coverage-measurement.html>

Cézard M., Lefebvre O., « La qualité du recensement », Insee, décembre 2008.

Chardon O., Vivas É., « Les familles recomposées : entre familles traditionnelles et familles monoparentales », *Documents de travail de la Direction des Statistiques Démographiques et Sociales* n° F0904, Insee, 2004.

Coeffic N., « L'enquête post-censitaire de 1990. Une mesure de l'exhaustivité du recensement », *Population*, vol 48, n° 6, pp 1655-1681, 1993.

Desplanques G., « Avantages et incertitudes des enquêtes annuelles de recensement en France », *Population*, vol. 63, n° 3, pp 477-501, 2008.

Deville J.-C., Tillé Y., « Échantillonnage équilibré par la méthode du cube, variance et estimation de variance », *Actes des journées de méthodologie statistique*, 2000.

Eurostat, « Code de bonnes pratiques de la statistique européenne », novembre 2017.

Godinot A., « Pour comprendre le recensement de la population », *Insee Méthodes hors série*, mai 2005.

Imbert C., Deschamps G., Lelièvre E., Bonvalet C., « Vivre dans deux logements : surtout avant et après la vie active », *Population et Sociétés* n° 507, 4 p, 2014. <https://www.ined.fr/fr/publications/editions/population-et-societes/vivre-deux-logements-vie-active/>

Insee, « Rénovation du questionnaire du recensement de la population. Impact sur les évolutions annuelles de population » et « Rénovation du questionnaire du recensement de la population. Méthodes d'estimation de l'effet questionnaire (ajustement) à partir des enquêtes annuelles de recensement », *Notes techniques*, 2020. https://insee.fr/fr/statistiques/fichier/2383177/Fiche_complexe_V2020.pdf et https://insee.fr/fr/statistiques/fichier/2383177/Fiche_methodo_effet%20questionnaire_V2020.pdf

Insee, « Base étude 2016 de l'échantillon démographique permanent », *Manuel de l'utilisateur*, 476 p, 2018. <https://utiledp.site.ined.fr/fr/variables/variables-de-l-edp/>

Insee, « L'ajustement intercensitaire 1999-2006 », *Fiche conseil*, mai 2018. <https://www.insee.fr/fr/statistiques/fichier/2383177/fiche-ajustement.pdf>

Insee, « Âge, générations et pyramides des âges », *Fiche conseil*, mai 2018b. <https://insee.fr/fr/statistiques/fichier/2383177/fiche-age.pdf>

- Insee, « La précision des résultats du recensement », *Fiche conseil*, mars 2017. <https://insee.fr/fr/statistiques/fichier/2383177/fiche-precision.pdf>
- Jugnot, S., « La constitution de l'échantillon démographique permanent de 1968 à 2012 », *Documents de travail de la Direction des Statistiques Démographiques et Sociales* n° F1404, Insee, 2014. <https://www.insee.fr/fr/statistiques/1381113>
- Le Penven E., « Le nouveau recensement français : un déficit de jeunes enfants ! », *Population & Avenir*, janvier 2020.
- Mambetov D., « Étude des BI multiples associés à une même personne dans les enquêtes annuelles de recensement, à partir de l'EDP (base étude 2011) », pp 9, Insee, 2014.
- Papon S., Beaumel C., « Bilan démographique 2018. La fécondité baisse depuis quatre ans », *Insee Première* n° 1730, 2019. <https://www.insee.fr/fr/statistiques/3692693>
- Pla A., « Bilan démographique 2008: plus d'enfants, de plus en plus tard ». *Insee Première* n° 1220, 2009. <https://www.insee.fr/fr/statistiques/1280757>.
- Roux V., « The French Rolling Census: A Census That Allows a Progressive Modernization' », *Statistical Journal of the IAOS*, vol 36, pp 125-134, 2020.
- Toulemon L., « Combien de personnes ont plusieurs résidences habituelles en France ? », Ch. 1 in Imbert C., Lelièvre E., Lessault D., (eds.), *La famille à distance. Mobilités, territoires et liens familiaux*. Paris : Ined, pp 27-49, 2018a.
- Toulemon L., « Faut-il ajuster les estimations de population de 2016 ? », *Note pour Chantal Cases*, directrice des statistiques démographiques et sociales de l'Insee, 2 novembre, 2018b.
- Toulemon L., « Undercount of young children and young adults in the new French census », *Statistical Journal of the IAOS*, vol 33, pp. 311-316, 2017.
- Toulemon L., Pennec S., « Two-home adults and children: long-lasting or transitory family situation? The case of France », XXVI IUSSP International Population Conference, Marrakech, Séance 501, Unions, families and households, 2009. <http://iussp2009.princeton.edu/abstracts/93441>
- Toulemon, L., Pennec, S., « Multi-residence in France and Australia: Why count them? What is at stake? Double counting and actual family situations ». *Demographic Research*, vol 23, n°1, pp 1-40, 2010. <https://www.demographic-research.org/volumes/vol23/1/>
- Toulemon L., Durier S., Marteau B., « Au recensement, 2,5 % de doubles comptes, d'après l'échantillon démographique permanent », *Actes des Journées de méthodologie statistique*, 2018. http://jms-insee.fr/jms2018s26_2/
- Toulemon L., Morand E., Razafindratsima N., Damian R., « La précision de l'estimation du nombre de couples de même sexe en France à partir d'une étude incluse dans le recensement ». 8e Colloque francophone sur les sondages, Société française de la statistique (SFDS), Dijon, 18-20 novembre, 2014.
- Unece, « Generic Statistical Data Editing Model (GSDEM) – version 2.0 », juin 2019.
- Unece, « Recommendations for the 2020 Censuses of Population and Housing », 2015.

Annexe 1 : définition d'une communauté et gestion du répertoire

Définition d'une communauté dans le cadre du recensement de la population

Une communauté au sens du décret n° 2009-485 du 5 juin 2003 est un ensemble de locaux d'habitation relevant d'une même autorité gestionnaire et dont les habitants partagent à titre habituel un mode de vie commun. Il s'agit, par exemple, des internats, casernes, Ehpad...

La population de la communauté comprend les personnes qui résident dans la communauté, à l'exception de celles résidant dans les logements situés au sein de la communauté et appelés « logements de fonction » (parfois abusivement, car il peut s'agir de logements de résidents et pas uniquement de logements du personnel). Ces dernières sont comptées dans la population des ménages.

Le mode de vie commun mentionné par le décret se traduit concrètement par le fait que la population des communautés vit dans des chambres ou des dortoirs. Les éléments de confort, tels que la cuisine ou les sanitaires, sont partagés avec les autres membres de la communauté. La population des logements de fonction, quant à elle, vit dans des logements équipés de ces éléments de confort (studios, appartements, pavillons individuels...).

On compte environ 30 000 communautés France entière (cf. tableau A1 .1). Elles sont classées en 7 catégories listées dans le décret, elles-mêmes découpées en sous-catégories pour des besoins de gestion (temps de collecte, redressements...). Elles hébergent 1,6 million d'individus (y compris les logements de fonction), soit 2,3 % de la population française. Une communauté bascule dans le champ des logements ordinaires dès lors qu'elle n'est plus constituée que de logements, à l'exception de quelques structures spécifiques qui restent en communauté : gendarmeries, résidences Adoma, Ehpad, etc.

Repérage des unités à enquêter - mises à jour de la base de sondage

Pour repérer les unités à enquêter, l'Insee met à jour en continu un répertoire des communautés à partir de différentes sources d'informations administratives (répertoire Finess de la Drees, fichier des gendarmeries...) ou directement auprès des têtes de réseau (Adoma, ministère de la Justice...).

Par ailleurs, en amont de chaque enquête annuelle de recensement, les communes sont sollicitées pour expertiser la liste des communautés situées sur leur territoire et signaler les ouvertures et fermetures.

Enfin, trois mois avant le début de la collecte, l'échantillon est confirmé par l'envoi de coupons-réponse aux communautés elles-mêmes, ce qui permet de vérifier leur existence ainsi que les informations contenues dans le répertoire et de disposer d'un correspondant au sein de la communauté.

Tableau A1.1 : Nombre de communautés par sous-catégorie (2018)

Catégories		Sous-catégories		Nombre de structures fin 2018
1	Service de moyen ou de long séjour d'un établissement public ou privé de santé, établissement social de moyen ou long séjour, maison de retraite, foyer ou résidence sociale ou assimilé	11	Maison de retraite, EHPAD	7 823
		12	Structure gérée par Adoma	619
		13	Autre foyer de travailleurs	749
		15	Structure pour personnes nécessitant des soins médicaux (enfants, adultes)	1 954
		16	Structure pour les enfants handicapés	301
		17	Structure pour les adultes handicapés	3 318
		18	Structure d'aide sociale à l'enfance et de protection judiciaire pour enfants et jeunes majeurs	1 780
		19	Structure pour adultes et familles nécessitant un accompagnement social et psychologique	1 110
2	Communauté religieuse	21	Communauté religieuse	2 524
3	Casernes, quartier, base ou camp militaire ou assimilé	31	Gendarmerie	3 311
		32	Autre établissement militaire	403
4	Établissement hébergeant des élèves ou des étudiants, y compris établissement militaire d'enseignement	41	Cité universitaire, foyer d'étudiants	391
		42	Autre internat	5 148
5	Établissement pénitentiaire	51	Établissement pénitentiaire	190
6	Établissement social de court séjour	61	Établissement social de court séjour (hébergement du 115)	387
7	Autre catégorie de communauté	71	Autre catégorie de communauté (secte, centre de rétention administrative...)	46
TOTAL				30 054

Source : Insee, répertoire des communautés.

Annexe 2 : une méthode possible pour un correctif macro de la pyramide des âges

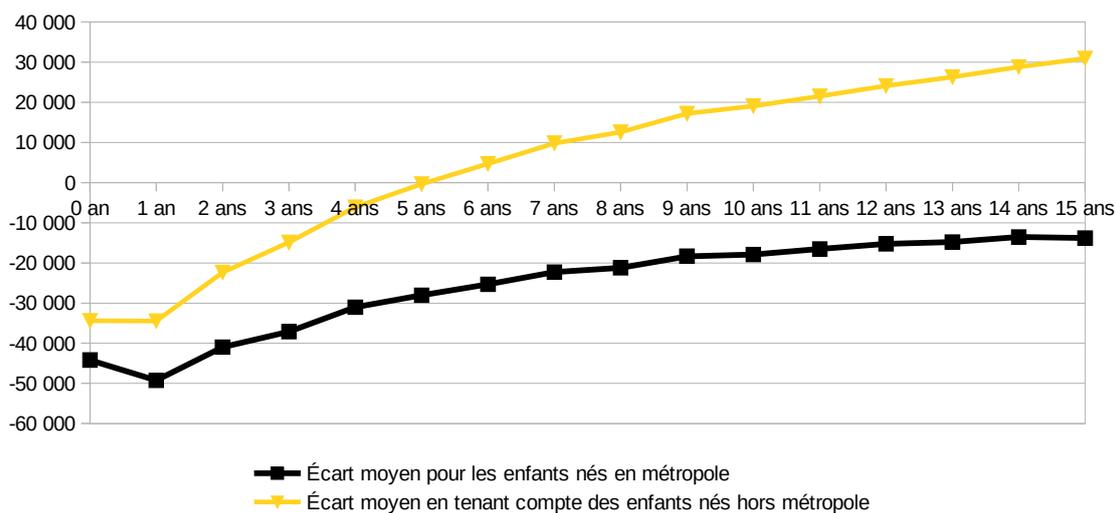
Les chiffres présentés dans le tableau 3-B.1 de la partie 3-B « Omissions possibles de jeunes enfants au recensement : que nous apprennent les comparaisons avec l'état civil ? » concernent les enfants nés en métropole, pour disposer des données nécessaires sur le même champ géographique pour toutes les années pour lesquelles la comparaison entre l'EAR et l'état civil a été réalisée. Dans le même temps, des enfants nés hors métropole arrivent chaque année en France. S'ils ne figurent pas dans les fichiers de naissances de l'état civil³², ils sont comptabilisés au recensement. Celui-ci fournit le nombre d'enfants nés hors métropole et vivant en métropole à un âge donné. Pour chaque âge entre 0 et 15 ans, on calcule dans chaque EAR le nombre d'enfants recensés nés hors France métropolitaine. Puis, pour chaque âge, on en calcule la moyenne sur l'ensemble des EAR. Ainsi, en moyenne sur les EAR de 2004 à 2018, 10 000 enfants nés hors métropole sont recensés en métropole à 0 an. Leur nombre augmente avec l'âge : ils sont 45 000 en moyenne à 15 ans.

On peut alors considérer que ces enfants nés hors métropole viennent « compenser » au recensement, du point de vue du « nombre de personnes d'un âge donné présent en France », une partie des omissions de jeunes enfants nés en métropole, comme en atteste le graphique A2.1.

En tenant compte des enfants nés hors métropole, l'écart entre le recensement et l'état civil devient nul à partir de 5 ans. Au-delà de cet âge, le recensement comptabilise davantage d'enfants que l'état civil. Ainsi, l'augmentation dans les EAR du nombre d'enfants nés à l'étranger ou dans les DOM et les COM et vivant en métropole permet de compenser vers 5 ans le déficit observé sur les enfants nés en métropole. Il n'en reste pas moins que, sur les seuls nés en métropole, lorsque l'on compare les chiffres des EAR à ceux de l'état civil, il reste un déficit plus important et plus persistant, mais dont une partie au moins provient des départs d'enfants nés en France et vivant à l'étranger.

32 C'est en particulier le cas des enfants nés à l'étranger.

Graphique A2.1 : Avant 10 ans, les écarts entre le recensement et l'état civil se réduisent avec l'âge et avec la prise en compte des enfants nés hors métropole



Champ :

pour la courbe sur les enfants nés en métropole : enfants recensés et nés en France métropolitaine – Naissances enregistrées en métropole, diminuées des décès enregistrés en métropole ;

pour la courbe tenant compte des enfants nés hors métropole : enfants recensés en France métropolitaine quel que soit leur lieu de naissance – Naissances enregistrées en métropole, diminuées des décès enregistrés en métropole.

Source : Insee, *Enquêtes annuelles de recensement de 2004 à 2018 et statistiques d'état civil*.

Dans le cadre de ses travaux sur la comparaison entre les effectifs de naissances de l'état civil et les effectifs des jeunes enfants au recensement, l'Insee a élaboré les grands principes que pourrait suivre un correctif macro de la pyramide des âges si l'on souhaitait corriger le sous-dénombrement des jeunes enfants au recensement. Il s'agirait, à un niveau macro, de modifier le nombre d'individus âgés de 0 à 5 ans, en supposant que les effectifs à 6 ans ne sont plus sous-estimés au recensement (cf. graphique A2.1). Les effectifs à 0 an pourraient être estimés, à partir des données d'état civil, par les naissances diminuées des décès. Les effectifs de 1 à 5 ans pourraient être obtenus par interpolation linéaire entre ces effectifs estimés à 0 an et les effectifs, estimés ou observés dans les EAR, à 6 ans : la méthode repose ainsi sur l'existence d'une fonction affine reliant les effectifs d'une même génération entre 0 et 6 ans et sur le suivi d'une génération sur plusieurs enquêtes de recensement successives, ce qui a du sens d'un point de vue démographique.

Après avoir estimé ces nouveaux effectifs pour chaque EAR, les pyramides des âges pourraient être construites selon la méthode habituelle : pyramide des âges calculée par génération, calée sur le nombre total d'habitants du recensement (RP cumulé), comme cela est fait habituellement.

Il s'agirait donc bien d'un correctif de niveau macro qui ne modifierait en rien les données individuelles des EAR, y compris notamment le chiffre connu et définitif de la population française donné par le dernier recensement.

Estimation des effectifs à 0 an

L'estimation des effectifs à 0 an repose sur les données d'état civil. Les effectifs à 0 an dans l'EAR de l'année N sont estimés par les naissances enregistrées de l'année N-1, auxquelles on soustrait les décès enregistrés à 0 an en N-1, et ce pour chacun des deux sexes. Pour l'EAR 2009 par exemple, les effectifs (notés *Eff* dans la formule ci-dessous) à 0 an au 1^{er} janvier seront estimés comme suit :

$$\text{Eff 0 an EAR 2009}^* = \text{0 an génération 2008 au 1}^{\text{er}} \text{ janv 2009} = \text{Naissances enregistrées en 2008} - \text{Décès enregistrés à 0 an en 2008}$$

* : données estimées

Eff : Effectifs

Les naissances et décès enregistrés (naissances en France) ont été préférés aux naissances et décès domiciliés (personnes vivant en France) puisque le redressement porte sur les effectifs globaux des EAR et non uniquement sur les enfants nés en France. Par ailleurs, le fait de ne pas exclure les enfants nés en France de mère domiciliée à l'étranger permet de « compenser » une petite part des enfants nés à l'étranger mais recensés sur le territoire français. Ce choix est en adéquation avec l'objectif fixé de raisonner en déficit « apparent » : on cherche à évaluer l'écart entre les effectifs à 0 an dans les EAR et les naissances (diminuées des décès), indépendamment du lieu de naissance.

Pour estimer les effectifs à 0 an, on ne tient pas compte à ce stade du solde migratoire. Le solde migratoire apparent à 0 an, calculé par différence entre deux pyramides des âges au 1^{er} janvier de deux années successives diminuées des décès à cet âge, est très négatif, du fait de la sous-estimation des bébés au recensement. Le seul moyen de réconcilier les données des naissances de l'état civil et le bas de la pyramide des âges issue du recensement (ou EAR) est de faire partir de nombreux bébés nés en France, d'où un solde très négatif à cet âge. On ne tient donc pas compte de ce solde pour estimer le point de départ des 0 an, ce qui revient, à ce stade du calcul, à faire l'hypothèse que le solde migratoire est nul. On fait cette hypothèse de travail parce qu'aucune donnée ne permet de l'estimer. Pour autant, le solde final ne sera pas nul lorsqu'il sera calculé par différence des pyramides corrigées (cf. graphique A 2.4 en annexe).

Interpolation linéaire des 1-5 ans dans les EAR

Le principe de l'interpolation linéaire est de calculer une droite affine entre deux points déterminés, ici 0 et 6 ans. Les effectifs à 0 an sont estimés comme mentionné ci-dessus. Les effectifs à 6 ans sont ceux de l'EAR pour la génération considérée.

Pour illustrer la méthode de manière plus concrète, on se place ici dans le cas où l'on souhaite établir la **pyramide des âges au 1^{er} janvier 2013**. Cette pyramide est construite à partir de cinq EAR, de 2011 à 2015. Les effectifs des 1-5 ans sont interpolés, par sexe, dans chacune de ces cinq EAR.

Pour déterminer, par exemple, les effectifs âgés de 2 ans corrigés dans l'EAR 2011, le calcul par interpolation linéaire repose sur les effectifs de cette génération à 0 an (que l'on a estimés) et à 6 ans (que l'on constate dans l'EAR 2015). Le calcul s'effectue de cette manière :

$$\text{Eff 2 ans EAR 2011}^* = \text{0 an génération 2008 au 1}^{\text{er}} \text{ janv 2009}^* + 2 \left(\text{0 an génération 2008 au 1}^{\text{er}} \text{ janv 2009}^* - \text{Eff 6 ans EAR 2015} \right) / 6$$

* : données estimées

Eff : Effectifs

De la même manière, les effectifs âgés de 5 ans dans l'EAR 2013 s'obtiennent *via* la formule :

$$\text{Eff 5 ans EAR 2013}^* = \text{0 an génération 2007}^* \text{ au 1}^{\text{er}} \text{ janv 2008} + 5 \left(\text{0 an génération 2007}^* \text{ au 1}^{\text{er}} \text{ janv 2008} - \text{Eff 6 ans EAR 2014} \right) / 6$$

* : données estimées

Eff : Effectifs

On voit ici que les effectifs des enfants de 6 ans non encore connus dans les EAR (car la génération n'a pas encore 6 ans dans l'EAR la plus récente, elle atteindra cet âge dans des EAR à venir dans le futur) posent un problème qui sera traité au paragraphe suivant. Par ailleurs, on notera qu'il est théoriquement possible que les effectifs des EAR pour un âge donné soient supérieurs aux effectifs obtenus par interpolation. Dans ce cas, on choisit de retenir les effectifs interpolés.

Une fois le redressement des EAR ainsi effectué, on procède au calcul de la pyramide des âges comme fait habituellement.

Estimation des effectifs à 6 ans pour les générations qui n'ont pas encore été recensées à cet âge

L'interpolation linéaire telle que présentée ci-dessus ne peut être appliquée directement pour les plus jeunes générations, car leurs effectifs à 6 ans ne sont pas encore connus. Par exemple, les enfants de 1 an dans la dernière EAR n'ont nécessairement pas encore été recensés à 6 ans et ne le seront que dans cinq ans. Il va donc falloir les estimer.

Pour estimer la population à 6 ans qui n'est pas encore connue, on va utiliser le rapport moyen entre la population recensée à 6 ans et celle estimée à 0 an (c'est-à-dire les naissances diminuée des décès) observé sur des générations pour lesquelles on dispose de toute l'information nécessaire.

On se place à nouveau dans le cadre du calcul de la pyramide des âges au 01/01/2013. Le tableau A2.1 montre comment les effectifs à 6 ans des générations 2009 à 2012 sont estimés à partir du rapport moyen entre les enfants de 6 ans et les 0 an observé sur les trois générations précédentes, les générations 2006 à 2008. La moyenne du rapport entre population recensée à 6 ans et population estimée à 0 an sur ces trois générations est de 1,0062 pour les garçons et de 1,0003 pour les filles. Appliquée aux effectifs à 0 an, elle permet d'estimer les effectifs à 6 ans des générations 2009 à 2012.

Tableau A2.1. Estimation des effectifs d'hommes et de femmes âgés de 6 ans des générations 2009 à 2012 à partir de l'évolution moyenne entre 0 an et 6 ans des générations 2006 à 2008

Année de naissance	Hommes			Femmes		
	0 estimé = N-D	6 ans (recensés)	6 / (N-D)	0 estimé = N-D	6 ans (recensés)	6 / (N-D)
2006	406 394	407 040	1,0016	387 979	388 161	1,0005
2007	400 859	404 264	1,0085	382 689	387 461	1,0125
2008	405 399	408 906	1,0086	388 143	383 518	0,9881
Moyenne de 6 / (N-D)			1,0062			1,0003
Année de naissance	0 estimé = N-D	6 ans estimés = (N-D) * 1,0062		0 estimé = N-D	6 ans estimés = (N-D) * 1,0003	
2009	404 453	406 979		386 403	386 535	
2010	408 781	411 334		390 986	391 120	
2011	403 971	406 494		386 772	386 904	
2012	403 463	405 983		384 506	384 637	

Champ : France métropolitaine

Quand il s'agira de calculer la pyramide des âges au 01/01/2014, la moyenne du rapport 0-6 ans sera calculée sur les générations 2007 à 2009 et permettra d'estimer les effectifs à 6 ans des générations 2010 à 2013.

Il faut noter qu'à l'arrivée d'une nouvelle EAR, on recalcule les effectifs corrigés des EAR précédentes, car de nouveaux effectifs à 6 ans sont désormais connus.

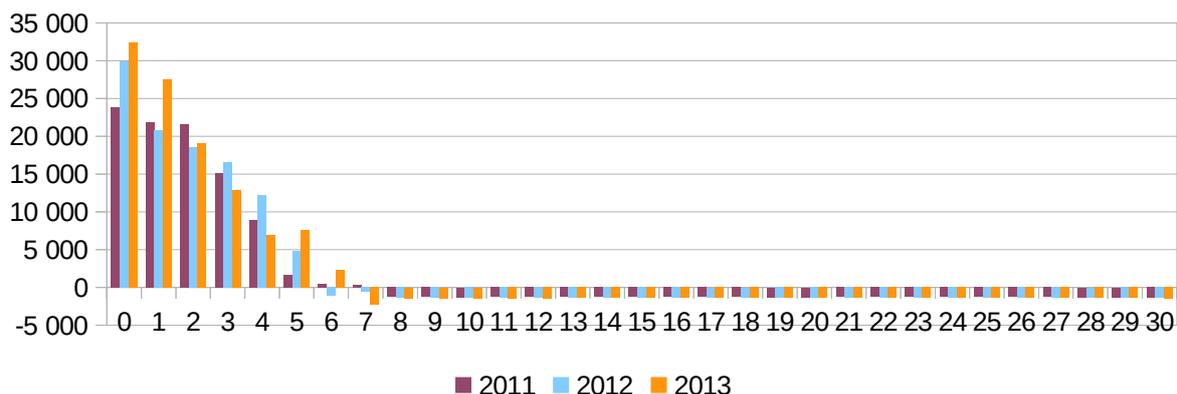
Résultats

On présente ici l'impact de la méthode de redressement sur 3 pyramides des âges pour la France métropolitaine, aux 1^{ers} janvier 2011, 2012 et 2013.

Par rapport aux pyramides actuellement diffusées, les modifications de population concernent principalement, et logiquement, les individus âgés de 0 à 5 ans (cf. graphique A2.2). L'écart le plus important s'observe toujours à l'âge 0 : par exemple, le correctif ajoute 32 400 enfants de 0 an à la pyramide au 01/01/2013. Cela représente 4,3 % de l'effectif initial des 0 an de cette année. De 0 à 5 ans, les écarts sont globalement décroissants avec l'âge.

En dehors de ces âges, la population des « pyramides interpolées » est, en moyenne sur les trois pyramides, légèrement inférieure d'environ 1 400 individus à chaque âge entre 8 ans et 65 ans (de moins en moins au-delà). En effet, la population municipale cible étant inchangée (c'est le total de la population du recensement), l'ajout de jeunes enfants dans les « pyramides interpolées » se traduit par une réduction du nombre d'individus aux âges plus élevés. Les écarts entre la pyramide de départ et la pyramide corrigée pour les âges intermédiaires, 6 et 7 ans, varient d'une année à l'autre, mais restent limités.

Graphique A2.2 : Écarts de population entre les pyramides des âges aux premiers janvier 2011, 2012 et 2013 corrigées par interpolation et les pyramides des âges définitives des Bilans démographiques.



Le tableau A2.2 reprend ces éléments en proposant une comparaison chiffrée des écarts entre les pyramides interpolées et les pyramides des âges non corrigées (i.e. celles diffusées).

Tableau A2.2 : Comparaison selon différents critères des pyramides interpolées et des pyramides actuelles

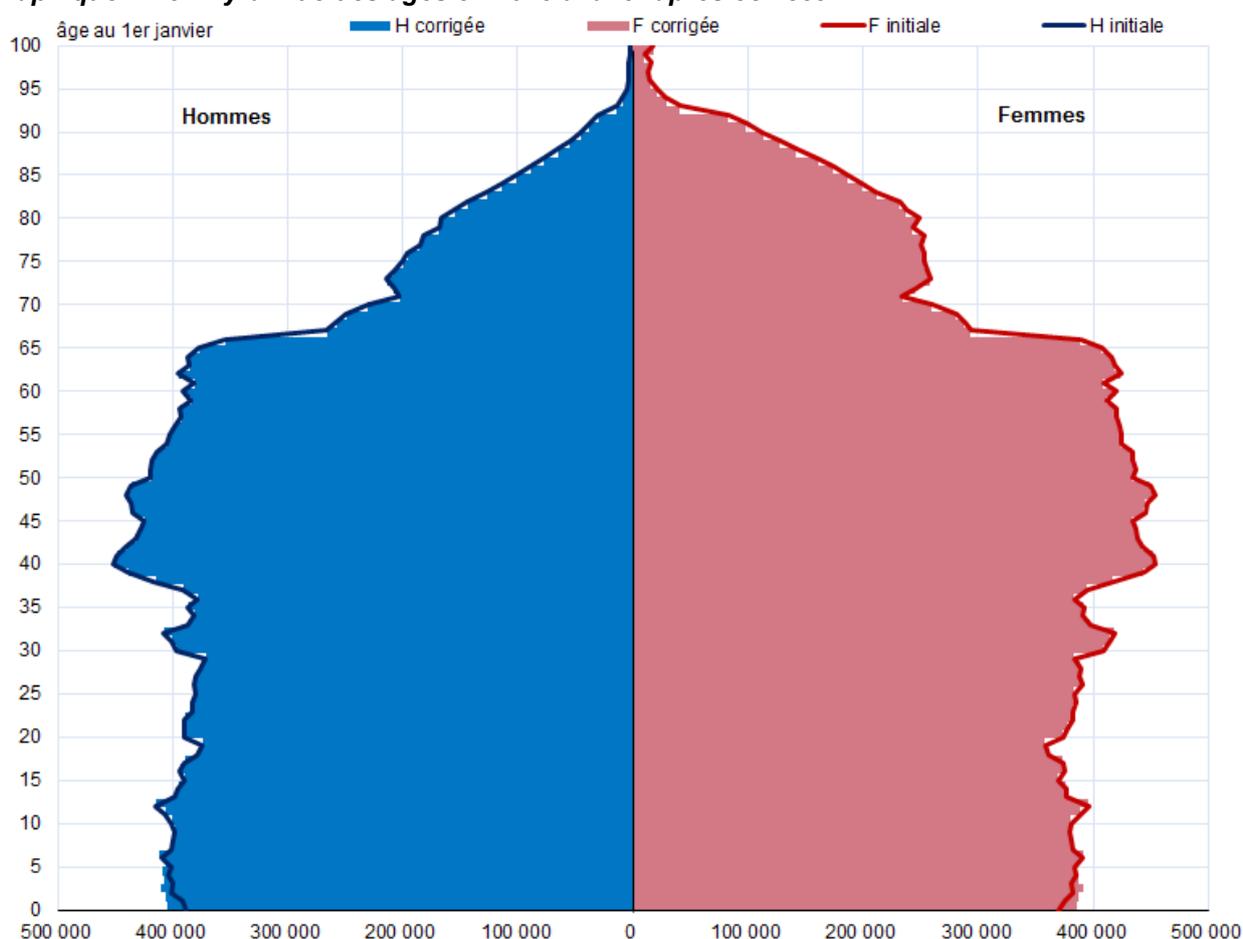
	Pyramide au		
	01/01/2011	01/01/2012	01/01/2013
0 an			
Écart absolu	+ 23 800	+ 29 800	+ 32 400
Écart relatif*	3,1 %	3,9 %	4,3 %
0-5 ans			
Écart absolu total	+ 92 900	+ 102 700	+ 106 400
Écart relatif**	2,0 %	2,2 %	2,3 %
8-65 ans			
Écarts moyens à chaque âge	- 1 300	- 1 400	- 1 500

* Écart relatif calculé par rapport aux effectifs à 0 an de la pyramide actuelle.

** Écart relatif calculé par rapport aux effectifs des 0-5 ans de la pyramide actuelle.

Comme attendu, l'interpolation linéaire corrige significativement le bas de la pyramide des âges. Il en découle qu'elle résout aussi au niveau national le problème majeur du solde migratoire fortement négatif à 0 an³³ et diminue le solde migratoire aux jeunes âges. Le solde migratoire à 0 an est estimé entre - 35 000 et - 40 000 (les sorties excèdent les entrées), surestimé avec les pyramides des âges, est également nettement réduit avec les pyramides obtenues par interpolation : il varie autour de - 6 000.

Graphique A2.3 : Pyramide des âges en 2013 avant / après correctif



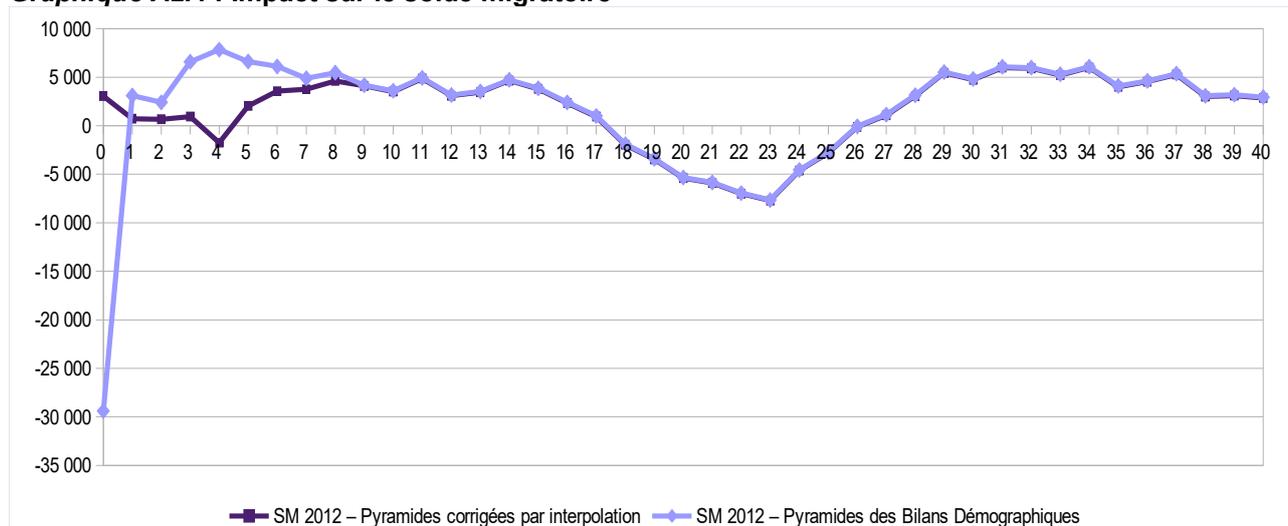
Champ : France métropolitaine.

Source : Insee.

33 La sous-estimation des enfants de 0 an au recensement et le mode de calcul du solde migratoire par différence entre deux pyramides des âges successives impliquent nécessairement un solde migratoire à 0 an fortement négatif quand il est déduit des pyramides des âges actuelles.

Le solde migratoire est déterminé à partir de la différence de deux pyramides des âges successives³⁴, après prise en compte de la natalité et de la mortalité. Ci-après figure le solde migratoire avant et après correction des pyramides des âges.

Graphique A2.4 : Impact sur le solde migratoire



Champ : France métropolitaine.

Source : Insee.

Sous les hypothèses retenues, l'interpolation linéaire est une méthode relativement simple et rapide à mettre en œuvre au niveau macro au niveau national. Elle peut rendre service aux utilisateurs des pyramides des âges qui seraient mis en difficulté par l'impact, sur ces pyramides, du déficit des jeunes enfants tel que mesuré au recensement. Elle nécessite des investigations plus importantes pour une déclinaison par région ou département. Elle n'est pas envisageable à un niveau géographique infra-départemental, du fait de la méthode d'échantillonnage du recensement de la population.

34 Le solde migratoire apparent à 0 an est quant à lui obtenu par la différence entre les effectifs à 0 an et les naissances de l'année précédente corrigée des décès à 0 an de cette même année.

Annexe 3 : principaux textes juridiques encadrant le recensement de la population

Cette annexe présente les fondements juridiques du recensement de la population et reprend les principaux textes encadrant cette opération : la loi n° 2002-276 du 27 février 2002 relative à la démocratie de proximité ; le décret n° 2003-485 du 5 juin 2003 modifié relatif au recensement de la population ; le décret n° 2003-561 du 23 juin 2003 portant répartition des communes pour les besoins du recensement de la population. Au-delà de ces trois textes, de nombreux arrêtés ont été publiés : arrêtés d'application du décret de 2003, arrêtés relatifs aux informations individuelles, notamment celui sur la diffusion des résultats, arrêtés de création du répertoire des immeubles localisés, etc. Pour une vision complète, le lecteur pourra se référer à (Godinot, 2015).

Le recensement de la population est également encadré par le Règlement général sur la protection des données (RGPD), la loi Informatique et libertés du 6 janvier 1978 et la loi sur l'obligation, la coordination et le secret en matière de statistiques du 7 juin 1951,

La loi n° 2002-276 du 27 février 2002 relative à la démocratie de proximité (extrait)

Titre V : Des opérations de recensement

Article 156

Modifié par la loi n° 2017-256 du 28 février 2017 – art. 147

I.-Le recensement de la population est effectué sous la responsabilité et le contrôle de l'État.

II.-Le recensement a pour objet :

- 1° le dénombrement de la population de la France ;
- 2° la description des caractéristiques démographiques et sociales de la population ;
- 3° le dénombrement et la description des caractéristiques des logements.

Les données recueillies sont régies par les dispositions de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques et de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

III.-La collecte des informations est organisée et contrôlée par l'Institut national de la statistique et des études économiques.

Les enquêtes de recensement sont préparées et réalisées par les communes ou les établissements publics de coopération intercommunale, qui reçoivent à ce titre une dotation forfaitaire de l'État.

IV.-Paragraphe modifiant l'article L. 2122-21 du CGCT.

V.-Lorsque l'établissement public de coopération intercommunale a reçu des communes qui le constituent compétence pour préparer et réaliser les enquêtes de recensement, l'organe délibérant de l'établissement peut, par délibération, charger le président de l'établissement de procéder à ces enquêtes.

Dans le cas où une commune ou un établissement public de coopération intercommunale refuserait ou négligerait d'accomplir cette mission, le représentant de l'État dans le département peut, après l'en avoir requis, y pourvoir d'office.

Les enquêtes de recensement sont effectuées par des agents recenseurs, agents de la commune ou de l'établissement public de coopération intercommunale affectés à cette tâche ou recrutés par eux à cette fin. Lorsque l'activité exercée par un agent recenseur présente un caractère accessoire, elle est exclue de l'interdiction prévue par l'article 25 de la loi n° 83-634 du 13 juillet 1983 portant droits et obligations des fonctionnaires. L'inéligibilité prévue au douzième alinéa de l'article L. 231 du code électoral s'applique à tous les agents recenseurs, quel que soit le nombre d'habitants de la commune.

VI.-Les dates des enquêtes de recensement peuvent être différents selon les communes.

Pour les communes dont la population est inférieure à 10 000 habitants, les enquêtes sont exhaustives et ont lieu chaque année par roulement au cours d'une période de cinq ans. Pour les autres communes, une enquête par sondage est effectuée chaque année ; la totalité du territoire de ces communes est prise en compte au terme de la même période de cinq ans.

Chaque année, un décret établit la liste des communes concernées par les enquêtes de recensement au titre de l'année suivante.

VII.-Pour établir les chiffres de population, l'Institut national de la statistique et des études économiques utilise les informations collectées dans chaque commune au moyen d'enquêtes de recensement exhaustives ou par sondage, les données démographiques non nominatives issues des fichiers administratifs, notamment sociaux et fiscaux, que l'institut est habilité à collecter à des fins exclusivement statistiques, ainsi que les résultats de toutes autres enquêtes statistiques réalisées en application de l'article 2 de la loi n° 51-711 du 7 juin 1951 précitée.

À cette fin, les autorités gestionnaires des fichiers des organismes servant les prestations de base des régimes obligatoires d'assurance maladie transmettent à l'Institut national de la statistique et des études économiques les informations non nominatives qu'il appartient à l'institut d'agrégier cinq ans après leur réception, à un niveau géographique de nature à éviter toute identification de personnes.

VIII.-Un décret authentifie chaque année les chiffres des populations de métropole, des départements d'outre-mer, de Saint-Barthélemy, de Saint-Martin et de Saint-Pierre-et-Miquelon, des circonscriptions administratives et des collectivités territoriales.

IX.-Les informations relatives à la localisation des immeubles, nécessaires à la préparation et à la réalisation des enquêtes de recensement, sont librement échangées entre l'Institut national de la statistique et des études économiques, les communes et les établissements publics de coopération intercommunale intéressés.

X.-Le premier décret authentifiant les chiffres de population en application du VIII sera publié à la fin de la première période de cinq ans mentionnée au VI.

NOTA :

Loi n° 2007-148 du 2 février 2007 art. 45 IV : les modifications induites par la présente loi entrent en vigueur à la publication du décret mentionné au I de l'article 25 de la loi n° 83-634 du 13 juillet 1983, dans sa rédaction issue de la présente loi, et au plus tard le 1^{er} juillet 2007.

Article 157

Modifié par la loi n° 2017-256 du 28 février 2017 - art. 147

I. - Jusqu'à la publication du décret mentionné au X de l'article 156, la population des collectivités territoriales et des circonscriptions administratives est celle qui a été authentifiée par décret à l'issue du dernier recensement général de la population effectué en métropole, dans les départements d'outre-mer et à Saint-Pierre-et-Miquelon, modifiée, le cas échéant, par des recensements complémentaires.

À compter de la publication du même décret, les références au recensement général de la population et au recensement complémentaire sont remplacées par des références au recensement de la population dans toutes les dispositions législatives alors en vigueur.

II. - Par dérogation aux dispositions de l'article 156 et du I du présent article, il est procédé, tous les cinq ans, à des recensements généraux de la population en Nouvelle-Calédonie, en Polynésie française et dans les îles Wallis et Futuna. Les opérations de recensement y sont, le cas échéant, organisées avec l'institut de statistiques compétent. Après chacun de ces recensements généraux, un décret authentifie les chiffres des populations de ces territoires, de leurs circonscriptions administratives et de leurs collectivités territoriales.

Ces dispositions s'appliquent en Nouvelle-Calédonie et en Polynésie française dans le respect des compétences définies par les lois organiques fixant leur statut.

Dans les îles Wallis et Futuna, les enquêtes de recensement sont préparées et réalisées par les services de l'administrateur supérieur, qui perçoivent à ce titre une dotation forfaitaire de l'État.

En Nouvelle-Calédonie, en Polynésie française, dans les îles Wallis et Futuna et à Mayotte, lorsque l'activité exercée par un agent recenseur présente un caractère accessoire, les interdictions relatives au cumul d'emplois public et privé prévues par la réglementation du travail en vigueur ne sont pas applicables.

Les dispositions de la dernière phrase du dernier alinéa du V de l'article 156 s'appliquent en Nouvelle-Calédonie, en Polynésie française et à Mayotte.

Article 158

Modifié par l'ordonnance n° 2004-637 du 1^{er} juillet 2004 – art. 12 JORF 2 juillet 2004

Un décret en Conseil d'État définit les modalités d'application du présent titre.

Le décret n° 2003-485 du 5 juin 2003 relatif au recensement de la population

Le décret n° 2003-561 du 23 juin 2003 portant répartition des communes pour les besoins du recensement de la population

Ces deux décrets sont consultables sur le site internet Legifrance.gouv.fr