



# Conseil économique et social

Distr. générale  
5 juillet 2010

Français

---

## Commission économique pour l'Europe

### Conférence des statisticiens européens

#### Groupe d'experts des recensements de la population et des habitations

##### Treizième réunion

Genève, 7-9 juillet 2010

Point 3 de l'ordre du jour

##### Plans d'évaluation de la qualité des recensements

## Qualité des recensements

Note établie par l'Institut national de la statistique et des études

## **La qualité du recensement**

*Michel CÉZARD, Olivier LEFEBVRE, Insee*

**Décembre 2008**

### **Résumé**

*Le recensement français repose depuis 2004 sur une nouvelle méthode, alliant deux principes : la collecte tournante sur un cycle de cinq ans et le recours au sondage dans les communes de 10 000 habitants ou plus en s'appuyant sur un répertoire exhaustif des immeubles (le RIL). Ces principes figurent dans la loi qui fonde le recensement. Ainsi rénové, le recensement permettra de fournir chaque année des résultats détaillés à tous les niveaux géographiques d'intérêt, du pays entier à la commune et au quartier (§ 1 à 3).*

*Cette réforme du recensement répond au besoin, de plus en plus marqué, de données démographiques récentes, surtout à des niveaux géographiques fins. Elle permet également de lisser les coûts du recensement, qu'ils soient humains ou budgétaires, et de mieux organiser et contrôler les opérations (§ 4). Enfin, elle permet de réduire les omissions, grâce notamment au répertoire d'immeubles.*

*Le recensement tel qu'il se pratique en France répond aux cinq critères essentiels définis par l'ONU (§ 5), ainsi qu'aux critères généraux de qualité des opérations statistiques (§ 6).*

*La qualité est gérée au fil des opérations de préparation, de collecte, de traitement et de diffusion du recensement. Elle repose sur des protocoles de contrôle et de mesure de la qualité et des bilans mobilisant les différents acteurs du recensement : communes, acteurs de la collecte, équipes de l'Insee, sous-traitants (§ 7 et 8). Des investigations en cours ou en projet devraient enrichir une analyse quantitative et qualitative de la couverture du recensement (§ 9).*

## **A/ Présentation générale du recensement**

### **1. Les fondements juridiques du recensement**

Le recensement est fondé par une loi votée en 2002 qui en énonce les finalités et les grands principes : organisation et modalités des enquêtes de recensement, modalités de partenariat entre l'Insee et les communes, principe d'une publication annuelle des chiffres de population de toutes les communes. Cette loi a été complétée par plusieurs décrets et arrêtés qui en précisent les modalités d'application.

Les chiffres de population tirés du recensement servent de référence pour l'application de nombreuses dispositions concernant les collectivités locales (dotations de l'Etat aux communes, modalités d'organisation des conseils municipaux, statut et rémunération des personnels de la fonction publique territoriale) mais aussi pour l'urbanisme, l'implantation des pharmacies, etc. Au total plus de 350 dispositions font référence à un chiffre de population, à fournir avec un haut degré de qualité pour chacune des 36 700 communes, dont près de 10 000 comptent moins de 200 habitants.

Enfin, le recensement français est compatible avec le règlement européen publié le 12/08/2008. Sa méthode figure parmi les méthodes reconnues pour fournir les informations prescrites.

## 2. La méthode

Le premier principe de la méthode est de répartir la collecte de l'information sur un cycle quinquennal glissant, pour produire chaque année une information relative à l'année médiane du cycle. Chaque année A, les données seront donc produites à partir des collectes des cinq années A-4 à A et seront représentatives de l'année A-2.

Le second principe est de recourir au sondage dans les plus grandes communes, celles dont la population est suffisante pour qu'un sondage fournisse une information robuste. Ce recours au sondage répond au souci d'alléger la charge que représente le recensement, pour l'Insee comme pour ces communes et les personnes enquêtées ; il permet également de mieux contrôler la qualité de la collecte et notamment de réduire les omissions, plus nombreuses dans les grandes villes. Le sondage s'appuie sur un répertoire exhaustif des immeubles, le RIL (Répertoire d'Immeubles Localisés).

Compte tenu du très grand nombre et de la très grande hétérogénéité des communes françaises (presque 37 000 communes, dont la un quart ont moins de 200 habitants) un dispositif de collecte particulier a été mis en œuvre.

### Les communes de moins de 10 000 habitants sont recensées exhaustivement, à raison d'un cinquième chaque année.

Les 35 750 communes de moins de 10 000 habitants, qui représentent la moitié de la population de la France, ont été réparties en cinq groupes. Chaque année, l'enquête de recensement concerne toutes les communes de l'un de ces groupes, et est exhaustive sur chacune de ces communes. Au bout de cinq ans, on réinterroge les communes du premier groupe, et ainsi de suite.

Ces cinq groupes sont équilibrés, c'est-à-dire qu'ils sont répartis de la même façon sur une dizaine de critères démographiques (population, répartition par sexe et grands groupes d'âges) ou relatifs au parc de logements (nombre de logements, nombre de résidences principales). L'équilibrage est assuré au niveau national et pour chacune des 26 régions de France. Il permet d'atteindre, à taille égale, une meilleure précision.

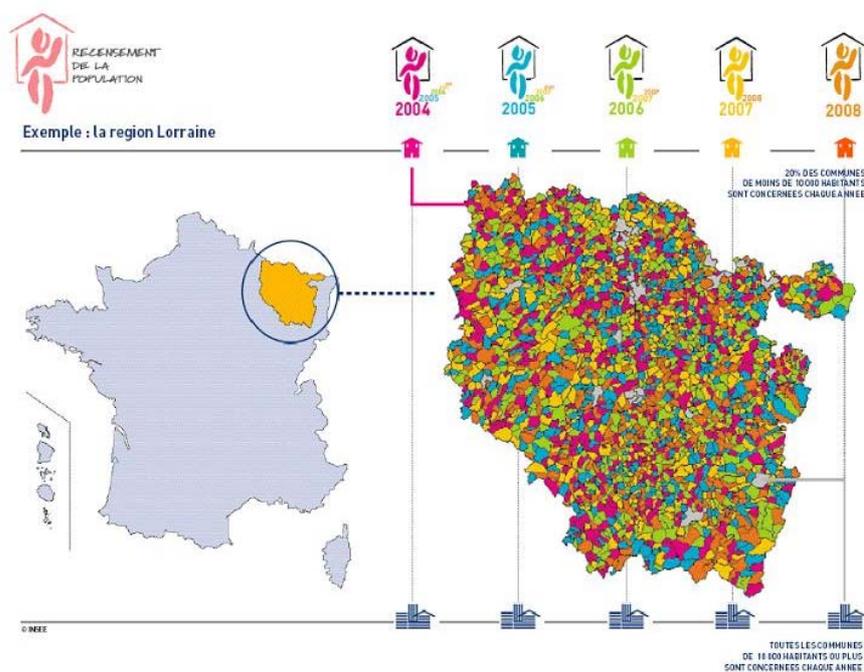


Fig. 1 Les cinq groupes de communes pour la région Lorraine

**Quant aux 900 communes de 10 000 habitants ou plus, qui regroupent l'autre moitié de la population du pays, elles réalisent tous les ans une enquête de recensement, mais seulement auprès d'une partie de leur population.** Chaque année, l'enquête porte sur 8 % des logements, répartis sur l'ensemble du territoire de la commune. Au bout de cinq ans, 40 % de la population de



chacune de ces communes aura donc été enquêtée, un taux suffisant pour garantir la fourniture de données robustes sur la commune et ses quartiers.

La base de sondage, dans chaque grande commune, est constituée par son répertoire d'immeubles localisés (RIL). Ce répertoire est une liste exhaustive des immeubles (qu'ils soient d'habitation, administratifs, industriels ou commerciaux), identifiés et localisés à leur adresse, grâce à un système d'information géographique qui en permet une représentation graphique. Le RIL a été initialisé en partant du recensement général de 1999 et mis à jour depuis au moyen de fichiers administratifs (permis de construire, fichiers de gestion de la fiscalité locale) ou postaux (fichiers d'adresses de La Poste). Il est soumis chaque année à l'expertise des communes – chacune pour ce qui concerne son territoire - et validé *in fine* par l'Insee.

L'expertise par les communes est indispensable car les communes ont une connaissance de leur territoire qui complète les informations dont dispose l'Insee. Chaque commune est sollicitée sur les flux de mise à jour du RIL, puis sur le stock. Sur les flux, l'Insee envoie aux communes des listes d'adresses à confirmer, c'est à dire des adresses qu'il souhaite voir ajoutées au RIL, supprimées ou modifiées et pour lesquelles il sollicite l'avis de la commune. Ces listes sont envoyées deux fois par an aux communes, sauf quand les flux sont faibles : dans ce cas, l'Insee et la commune peuvent souhaiter ne faire qu'un envoi qui regroupe l'ensemble des adresses à confirmer. En mai-juin intervient l'expertise de l'ensemble du RIL, c'est-à-dire du stock d'adresses.

Les enquêtes de mesure de la qualité indiquent clairement que la qualité du RIL est d'autant plus assurée que les communes participent activement à l'expertise.

Grâce à l'expertise de la commune, l'Insee peut utiliser une base de sondage aussi à jour que possible pour tirer l'échantillon des adresses qui vont être enquêtées dans la commune en janvier et février suivants.

Pour chaque commune, la liste des logements soumis à une enquête de recensement résulte d'un tirage au sort parmi les immeubles (adresses) d'habitation, selon la stratégie d'échantillonnage suivante :

- pour éviter des « effets de grappe », (*cf. encadré*) on définit des « grandes adresses », qui seront enquêtées exhaustivement sur le cycle quinquennal et sont donc réparties pour ce faire en cinq groupes annuels. Ce sont les immeubles qui comptent le plus de logements dans chaque commune, avec les contraintes suivantes : ces immeubles doivent comprendre au moins 60 logements ; l'ensemble des grandes adresses d'une commune ne doit pas représenter plus de 10% des logements de la commune) ;
- les adresses neuves sont également recensées exhaustivement, car on ne dispose pas d'informations permettant de les échantillonner ; elles sont également réparties sur cinq groupes ;
- les « autres adresses » sont réparties en cinq groupes équilibrés, comme le sont les groupes de communes de moins de 10 000 habitants, sur des critères démographiques ou relatifs au parc de logement, chacun des groupes étant uniformément réparti sur le territoire de la commune (une rue donnée comprendra ainsi des adresses des différents groupes) ; chaque année, l'échantillon d'adresses à recenser est tiré dans le groupe de rotation de l'année, de façon que le total des adresses à recenser (grandes adresses + adresses neuves + « autres adresses ») représente environ 8 % du total des logements de la commune.

**Le phénomène des effets de grappe.**

Dans le ou les immeubles que l'on trouve à une adresse donnée, il y a une forte probabilité de ressemblance des logements qui les composent et de leurs habitants quant aux variables relatives au confort du logement, à la catégorie sociale des habitants, à leur nationalité, etc. Si on pratiquait un sondage, le fait que cette adresse soit tirée ou pas dans l'échantillon aurait un impact certain sur l'estimation de ces variables. C'est ce qu'on appelle l'effet de grappe. L'impact serait d'autant plus fort que l'adresse correspondrait à de nombreux logements et que son poids dans le domaine géographique où sont produites les estimations serait grand : l'impact serait perceptible au niveau de la commune et encore plus à celui de l'IRIS ou de n'importe quelle zone incluant cette adresse. Le fait qu'en cinq ans toutes les adresses au-delà d'une certaine taille seront recensées élimine, par construction, cet inconvénient. L'effet de grappe subsiste pour les adresses de moindre taille mais, parce qu'elles sont plus homogènes, son impact est moins accusé.

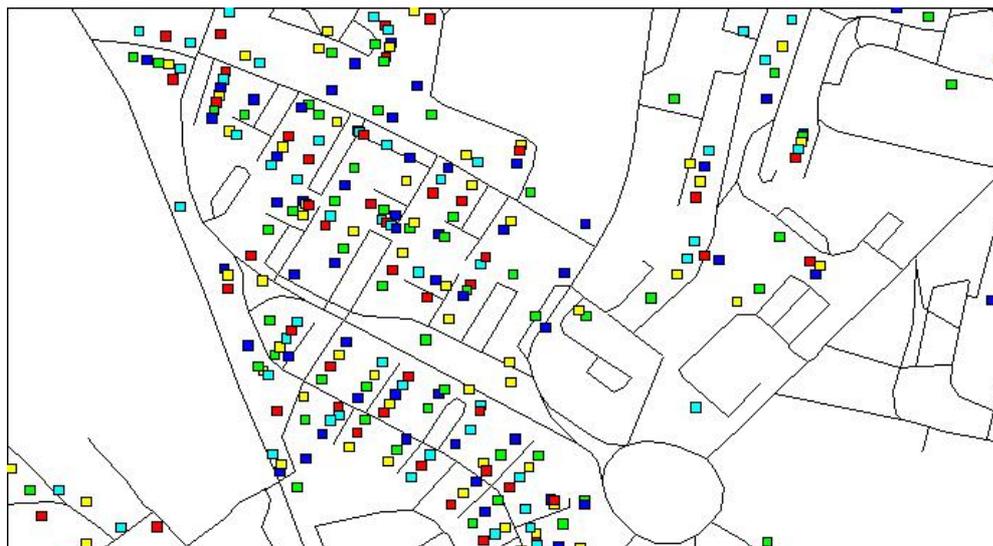


Fig. 2 : les cinq groupes d'adresses

Chaque année, on recense donc exhaustivement un cinquième des communes de moins de 10 000 habitants et 8 % de la population des communes de 10 000 habitants ou plus. Cela représente environ 4,5 millions de logements et 9 millions d'individus. **Au bout de cinq ans, la totalité des petites communes aura été recensée exhaustivement et 40 % de la population de chaque grande commune.**

**Au total, 70 % de la population du pays est recensée au cours d'un cycle quinquennal.**

Encadré : veiller à l'équilibre des groupes de rotation

Dans les communes de moins de 10 000 habitants, la composition des groupes de rotation ne change qu'à l'occasion de fusions ou scissions de communes ou de changement de mode de collecte (de passage de l'exhaustif au sondage ou l'inverse). Compte tenu de la taille importante des groupes, l'équilibrage a très peu de chances d'être altéré au cours du temps. Les comparaisons entre 1990 et 1999 montraient une très grande stabilité des équilibres entre groupes de rotation. L'exercice devra être repris avec les données du nouveau recensement. Il convient de noter qu'un tel déséquilibre n'affecterait pas la population de chaque commune, mais plutôt la qualité d'une enquête annuelle, et la précision des estimations.

Dans les communes de plus de 10 000 habitants, la composition des groupes de rotation varie avec l'apparition ou la disparition d'immeubles. Les adresses nouvelles sont réparties entre les groupes de rotation en veillant à conserver cet équilibre. Là encore, les confrontations entre 1990 et 1999 montrent une bonne stabilité des équilibres. L'exercice devra également être repris avec les données du nouveau recensement.

Les communautés (maisons de retraite, internats, établissements pénitentiaires, communautés religieuses, centres d'hébergement...) sont recensées exhaustivement. Là aussi, la collecte est tournante et étalée sur le cycle de cinq ans. Les communautés situées dans une commune de moins de 10 000 habitants sont enquêtées l'année de l'enquête de recensement de leur commune. Celles qui sont situées dans une commune de 10 000 habitants ou plus sont réparties en cinq groupes et chacune est enquêtée tous les cinq ans avec toutes celles de son groupe. L'Insee gère un répertoire des communautés à des fins d'organisation de la collecte et d'actualisation des populations. Comme le RIL, ce répertoire a été initialisé avec le recensement de 1999 et est actualisé avec des sources administratives (fichier des installations sanitaires, des internats, des établissements pénitentiaires...). Ce répertoire est expertisé par les communes.



Les personnes sans abri, les personnes vivant en habitation mobile terrestre et les marinières sont recensés exhaustivement, une fois tous les cinq ans.

### **La collecte reste organisée par dépôt-retrait des questionnaires.**

Si l'échantillonnage représente une innovation majeure, le mode de collecte reste celui des recensements généraux : un dépôt-retrait par des agents recenseurs. Chaque ménage se voit remettre deux types d'imprimés : une feuille de logement et des bulletins individuels. Ces questionnaires sont très proches de ceux utilisés lors des recensements précédents.

Le premier, intitulé « feuille de logement », permet d'établir la liste des habitants du logement et comprend une quinzaine de questions sur les caractéristiques et le confort du logement, ainsi que l'équipement du ménage en automobiles.

Le second, intitulé « bulletin individuel », comprend 25 questions : âge, sexe, lieu de naissance, nationalité, lieu de résidence cinq ans plus tôt, diplôme, activité professionnelle, lieu d'études ou de travail.

### **Des fiches de logement non enquêtés pour tenir compte des non-réponses**

Les fiches de logement non enquêté (FLNE) sont une innovation importante dans le nouveau recensement afin de tenir compte des non-réponses. Le principe en est le suivant : pour les résidences principales que l'on n'a pas réussi à enquêter (personnes impossibles à joindre, absentes de longue durée ou qui refusent de répondre), l'agent recenseur remplit une Fiche de Logement Non Enquêté (FLNE), en mentionnant le nombre de personnes qui habitent ce logement, ce qui permettra de redresser les chiffres de population. **La population de ces logements est ainsi prise en compte dans les décomptes de population.**

Pour chaque résidence principale de la commune, on recueille donc soit une feuille de logement et un ou plusieurs bulletins individuels, soit une FLNE. Ce point est contrôlé au moment de la collecte et ensuite à l'Insee.

### **3. La méthode de calcul des chiffres officiels de population et des résultats statistiques.**

**L'Insee doit satisfaire deux conditions pour la publication des chiffres de population :**

- publier **tous les ans** la population de **toutes les communes (obligation prescrite par la loi)**
- pour des raisons d'égalité de traitement des communes entre elles et de qualité pour la population de zones supra-communales (cohérence des chiffres à additionner), **cette population doit être relative à la même année pour tout le monde.** On ne peut pas retenir pour une commune le chiffre de 2004 et pour une autre le chiffre de 2008.

L'Insee calcule ces populations de la manière suivante :

#### **La population des ménages**

Pour les communes de 10 000 habitants ou plus, le chiffre repose sur une moyenne mobile établie à partir des échantillons de cinq années successives. Par agrégation des cinq échantillons des années A-4 à A, on calcule une population moyenne par logement, représentative de la situation du milieu de période (l'année A-2). On multiplie ensuite ce ratio par le nombre de logements au début A-2 tiré du répertoire d'immeubles localisés (RIL) pour obtenir la population des ménages de la commune.

Parallèlement, pour les communes de moins de 10 000 habitants, il faut ramener le chiffre de la population à l'année médiane du cycle quinquennal, pour être cohérent avec les communes de 10 000 habitants ou plus.

On s'appuie sur les enquêtes de recensement, selon le schéma suivant.

A-4	A-3	A-2	A-1	A
Recensement				
	Recensement			
		Recensement		
			Recensement	
				Recensement

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

Pour les communes recensées en A-1 et A, on obtient la population en A-2 par interpolation entre les résultats des deux dernières enquêtes de recensement (entre ceux de la dernière enquête de recensement et ceux du recensement de 1999 au démarrage du système).

Pour les communes recensées en A-4 et A-3, on procède par extrapolation entre le résultat de l'enquête de recensement et A-2 ; cette extrapolation s'appuie sur les données de la taxe d'habitation (TH), qui fournissent une indication sur l'évolution du nombre de logements par commune. Elle est affinée pour tenir compte du différentiel entre la croissance du nombre des logements et la croissance du nombre des personnes. Ce différentiel, mesuré entre les derniers recensements, est appliqué à l'évolution mesurée par la TH pour donner l'évolution du nombre de personnes.

### La population hors ménages

La population des communautés est ramenée au 1<sup>er</sup> janvier A-2, soit à la même date que la population des ménages. Si les communautés de la commune ont été recensées en A-2, le chiffre issu de la collecte est retenu. Si elles ont été recensées en A-4 ou A-3, la population des communautés est actualisée en ajoutant la population des communautés nouvelles et en retranchant celle des communautés disparues. Une communauté nouvelle est une communauté apparue depuis la dernière collecte ; on en connaît l'existence grâce à des sources administratives (fichier des établissements sanitaires et sociaux, des internats, des établissements pénitentiaires...). Les évolutions prises en compte sont déterminées à partir du répertoire des communautés. Si les communautés de la commune ont été recensées en A-1 ou A, la population de ces communautés au 1<sup>er</sup> janvier A-2 est calculée par interpolation entre les chiffres des deux derniers recensements des communautés.

Les populations des personnes sans abri, des personnes vivant dans des habitations mobiles ou dont la résidence principale est l'hôtel, ne sont pas actualisées. Elles sont reportées à l'identique durant les quatre années qui suivent l'année de la collecte. La nouvelle collecte remplace la précédente dès qu'elle est disponible.

### 4. Les raisons de la réforme

La réforme du recensement a été motivée par deux facteurs : un besoin de données récentes et plus régulières et le souhait de lisser la charge, financière et humaine, que représentait un recensement général.

Le besoin de données récentes était de plus en plus nettement exprimé par les utilisateurs des données de l'Insee. Les recensements généraux, de plus en plus espacés (neuf ans entre les deux derniers, réalisés en 1990 puis 1999), permettaient de moins en moins de rendre compte des évolutions de la population française : mutations sociales ou familiales, mouvements de péri-urbanisation, aménagements urbains, etc. Une demande forte émanait des collectivités locales, principaux utilisateurs du recensement et dont l'intérêt pour ces données était renforcé par les nouveaux pouvoirs qui leur sont progressivement conférés dans le cadre de la décentralisation : les communes ou leurs groupements ont ainsi vu leurs compétences étendues en matière de transports urbains ; les départements en matière d'aide sociale, d'aide aux personnes âgées, d'investissement et de fonctionnement pour les collèges ; les régions en matière de formation professionnelle, d'aménagement du territoire, d'investissement et de fonctionnement pour les lycées.



L'exercice de ces compétences implique de disposer de données localisées récentes et régulières, afin de prendre des décisions et d'évaluer leur impact. En fournissant des données détaillées chaque année à tous les niveaux géographiques (du pays à la commune et au quartier), le nouveau recensement répond à ces attentes.

Un rapport de 1996 du conseil Economique et Social « Les spécificités démographiques des régions et l'aménagement du territoire », déplore ainsi l'espacement croissant des recensements généraux en France.

Très espacés, les recensements généraux étaient également lourds. Leur coût, humain et financier, était difficile à absorber. Le recensement initialement prévu en 1997 a ainsi dû être décalé à 1999 pour des raisons budgétaires. En lissant la charge, la nouvelle méthode permet d'en étaler le coût et de disposer d'un budget pérenne (reconduit chaque année). Elle facilite la professionnalisation des acteurs du recensement (dans les communes et à l'Insee), puisqu'ils sont sollicités plus souvent. Elle rend possible une meilleure organisation et un meilleur contrôle des campagnes de recensement, du fait de leur ampleur plus réduite. Par ailleurs, le recours à une base de sondage (le RIL) permet un meilleur suivi de la collecte (voir ci-dessous le point sur l'universalité).

## **B/ L'adéquation aux critères internationaux de qualité**

### **5. Adéquation aux caractéristiques essentielles des recensements définies par l'ONU**

Dans un texte adopté concernant la réalisation des recensements de la décennie 2010, l'Organisation des Nations Unies a émis des recommandations sur ce que devait contenir un recensement de la population (« core variables ») et en a défini les cinq caractéristiques de base (« essential features ») : énumération individuelle, universalité, capacité à fournir de l'information sur de petites zones, simultanéité, régularité. Le nouveau recensement français répond à ces critères.

#### **« Énumération individuelle »**

L'information est recueillie séparément, pour chaque individu ou ménage enquêté. Un fichier-détail (comportant un enregistrement par individu ou ménage enquêté) sera produit chaque année, à partir de 2008, en cumulant les fichiers de cinq années d'enquête, permettant ainsi les tabulations croisées, comme pour les recensements généraux. Chacun des 9 millions d'individus et des 4,5 millions de logements enquêtés chaque année figure dans ce fichier, affecté de son poids de sondage. Chaque personne est ainsi prise en compte dans le fichier-détail.

#### **« Universalité »**

Chaque commune de moins de 10 000 habitants est enquêtée tous les cinq ans. Dans ces communes, toute la population est recensée. L'universalité est donc garantie comme lors des recensements généraux.

Pour les communes de 10 000 habitants ou plus, la totalité du territoire de la commune est prise en compte en cinq ans, comme le stipule la loi. La base de sondage est constituée par la totalité des immeubles d'habitation de la commune, ce qui permet bien de satisfaire cette contrainte et répond au principe d'universalité. Elle est actualisée tous les ans, en s'appuyant sur le fichier d'adresses de La Poste ainsi que sur les fichiers de permis de construire ou les fichiers fiscaux. L'Insee et les communes coopèrent pour cette actualisation.

L'information recueillie par cumul de cinq années de collecte permet donc bien de décrire toute la population et tous les territoires de chaque commune de 10 000 habitants ou plus.

Le taux de sondage cumulé sur 5 ans est très élevé (40 % dans les grandes communes), ce qui garantit la précision des résultats obtenus. On rappelle que, dans les recensements généraux, certaines variables n'étaient exploitées que pour un échantillon d'individus ou de logements. Et certains pays ne passent le « questionnaire long » que sur une fraction de leur population, souvent de l'ordre de 15 % ou 25 %.

#### **L'exhaustivité est mieux assurée, grâce au RIL, aux fiches de logement non enquêté (FLNE) et à une meilleure qualité de collecte.**

Avec le sondage et l'annualisation du recensement, la collecte est de meilleure qualité car les efforts des responsables sont plus concentrés, à l'Insee et dans les communes. En outre, dans les communes de 10 000 habitants ou plus, les agents recenseurs disposent d'une liste d'adresses à



recenser, ce qui augmente leur efficacité (cf. page 8). Les populations vivant en ménage collectif sont enquêtées exhaustivement, ainsi que les personnes vivant dans des habitations mobiles terrestres, les marinières et les personnes sans abri. Enfin, l'utilisation des FLNE permet d'estimer le nombre de personnes n'ayant pas répondu. Chaque personne vivant sur le territoire national est ainsi prise en compte.

#### **« Capacité à fournir de l'information sur de petites zones ou des populations d'effectif réduit »**

Le respect de ce critère découle de celui du critère d'universalité et d'énumération individuelle. Compte tenu du taux de sondage pratiqué (100 % dans les communes de moins de 10 000 habitants et 40 % dans les autres communes), et compte tenu de l'organisation des données en fichiers de données individuelles pondérées, il est possible d'en tirer une information détaillée portant sur des niveaux géographiques fins ou des populations de faible effectif. On rappelle que le fichier cumulé sur cinq ans comporte environ 45 millions d'individus et 22 millions de logements, ce qui permet des tabulations fines.

Le recensement produira chaque année des données détaillées sur chaque commune et, pour les communes de plus grande taille, sur leurs « IRIS », qui sont des quartiers d'environ 2 000 habitants.

En tout état de cause, pour des zones géographiques de petite taille, une information actualisée chaque année, même rendue un peu imprécise par l'aléa de sondage, a été jugée préférable à une information parfois très ancienne. Pour suivre une politique de rénovation urbaine, par exemple, le recensement tournant sera bien plus utile qu'un recensement général décennal. Ce point a fait débat lors de la concertation sur le recensement, certains utilisateurs préférant un recensement exhaustif, permettant des analyses plus fines, à des données annuelles jugées moins précises.

#### **« Simultanéité »**

Chaque année, les enquêtes de recensement ont lieu à la même date pour toutes les communes concernées (elles démarrent le troisième jeudi de janvier).

Les informations collectées sur différentes années sont ramenées à une date unique, celle du milieu du cycle quinquennal glissant.

Dans les grandes communes, on cumule les cinq échantillons, en « calant » ensuite les poids individuels avec le nombre de logements tiré du répertoire d'immeubles localisés de l'année médiane. Dans les petites communes, on procède par interpolation ou extrapolation entre l'enquête de recensement et la date de référence de la population. Les extrapolations sont consolidées par le recours aux données tirées du fichier de la taxe d'habitation, qui donnent une indication sur l'évolution annuelle du parc de logement des communes.

Compte tenu de la durée d'observation prise en compte (cinq ans), la portée de l'interpolation ou de l'extrapolation n'est jamais supérieure à deux ans, ce qui limite les dérives potentielles.

#### **« Régularité »**

Le cinquième critère essentiel est celui de la régularité : les recensements doivent avoir lieu régulièrement, toutes les N années (tous les cinq ou dix ans par exemple). Le recensement français fournira des résultats tous les ans et satisfera donc plus que largement à cette contrainte de régularité.

Il permettra également aux utilisateurs, qui disposeront de données annuelles, de s'affranchir des contraintes de dates posées par le caractère quinquennal ou décennal des recensements traditionnels.

### **6. Adéquation aux critères de qualité retenus par le code de bonnes pratiques d'Eurostat**

Eurostat a dressé une liste de critères de qualité, qui sont repris (avec quelques modifications de terminologie) dans le code de bonnes pratiques de la statistique européenne adopté par le Comité du programme statistique le 24 février 2005. Le nouveau recensement est conforme à ces exigences. Et sa méthode est reconnue dans le règlement sur les recensements européens publié en 2008, au même titre que les recensements généraux ou l'utilisation de registres de population et de logements.

**« Pertinence » :**

La réforme vise à fournir des données répondant mieux aux souhaits des utilisateurs, à savoir des données actualisées régulièrement. Les différentes actions de concertation, menées lors de la phase de préparation et, pour certaines, poursuivies au fil de l'avancement des opérations permettent de prendre en compte les besoins des utilisateurs pour la conception et la mise au point des produits et services de diffusion (cf. parties 1 et 3).

**« Précision » :**

**L'erreur aléatoire liée au sondage** (au maximum 0,05 % pour la France métropolitaine) est inférieure à l'erreur de collecte habituellement relevée dans les recensements généraux (souvent de l'ordre d'un à trois points, dans la plupart des pays pratiquant les recensements généraux) et également inférieure à celle induite par le vieillissement des données (l'évolution annuelle moyenne de la population est d'environ 0,7 % par an).

La précision de l'estimation de population dans les grandes communes

Population	Coefficient de variation (en %)
10 000	1,1
15 000	0,8
20 000	0,7
30 000	0,5
40 000	0,4
50 000	0,4
100 000	0,2
150 000	0,2
200 000	0,1
300 000	0,1
500 000	0,1
1 000 000	0,1

La précision dépend essentiellement de la taille de l'effectif étudié. Ce tableau donne le coefficient de variation du chiffre de population en fonction de sa taille, c'est-à-dire le rapport entre l'écart-type et le chiffre de population. Le coefficient de variation est une mesure de la précision relative du sondage.

**Quant aux erreurs de collecte**, les nouveau recensement permet de mieux les maîtriser (voir le paragraphe sur l'universalité), même si le « zéro défaut » reste hors de portée. Par ailleurs, le redressement rendu possible par les FLNE et l'extrapolation fondée sur le RIL permettent un meilleur contrôle de la sous-évaluation liée aux omissions. En revanche, si la couverture est meilleure, cela s'accompagne d'une légère dégradation du sur-dénombrement. Celui-ci est en effet plus difficile à contrôler dans un recensement tournant et reposant sur un sondage que dans un recensement exhaustif et simultané, où les doublons étaient plus faciles à repérer.

Un point de vigilance quant à la précision des données concerne la **précision des évolutions** qu'il est possible de mesurer : ce n'est pas parce que le recensement est disponible chaque année qu'il constitue une source conjoncturelle. Les variations de court terme doivent être interprétées avec prudence et sous certaines conditions :

- Pour les évolutions de population ou de parc de logements, on peut le plus souvent procéder à des analyses d'une année sur l'autre ;
- Pour analyser des évolutions portant sur les variables statistiques, il faut comparer deux millésimes distants d'au moins cinq ans, pour avoir la certitude de s'appuyer sur des échantillons totalement disjoints. Cela réduit de moitié l'ampleur des intervalles de comparaison, par rapport aux recensements généraux avec, toutefois, une plus grande souplesse dans le choix des années.

**« Actualité et ponctualité » :**

C'est un point fort du nouveau dispositif.

L'actualité des données produites est meilleure qu'avec les recensements généraux (c'est un des objectifs de la réforme) : le nouveau recensement fournit chaque année des données ayant une ancienneté de deux à trois ans contre 18 mois à dix ans dans les recensements généraux.

La ponctualité est inscrite dans la loi (fourniture des premières données avant fin 2008, périodicité annuelle ensuite) ; le respect du planning est une condition majeure de la réussite de l'opération, du fait de son caractère annuel ; enfin, des estimations de population communales ont été fournies en fin d'année en prenant en compte la collecte de début d'année, pour toutes les communes de moins de 10 000 habitants enquêtées en 2004, 2005, 2006 ou 2007, ainsi que pour environ 850 communes de plus de 10 000 habitants, pour lesquelles l'Insee a jugé que l'estimation obtenue après quatre collectes était suffisamment robuste pour être communiquée.

**« Accessibilité et clarté » :**

La diffusion des résultats repose prioritairement sur une consultation en ligne, ce qui permet une accessibilité très satisfaisante. Cette diffusion est précédée d'une communication des résultats les concernant aux communes, avec possibilité pour elles de communiquer à leur tour vers leurs administrés. Enfin, la presse relaie très largement les principaux résultats du recensement.

La documentation des données et l'accompagnement des utilisateurs sont des enjeux extrêmement importants compte tenu de la nouvelle méthode de collecte et d'élaboration des données. L'Insee y consacre des moyens substantiels, pour un effort qui doit se poursuivre sur plusieurs années, compte tenu de la richesse des utilisations du recensement.

**« Comparabilité » :**

La notion de comparabilité recouvre plusieurs aspects :

- Comparabilité temporelle : il y a peu de changements dans les questions posées par rapport aux recensements précédents, et ces changements sont documentés ; par ailleurs, la méthode reposant sur des enquêtes annuelles s'affranchit d'une des difficultés de comparaison dont souffraient les recensements généraux, à savoir l'effet des changements de protocoles de collecte. Désormais, avec des enquêtes régulières reposant sur les mêmes protocoles, la comparabilité est mieux assurée.
- Comparabilité internationale : elle est garantie par le respect des « core variables », de niveau ONU et Eurostat : les questions essentielles préconisées au niveau international figurent dans le recensement français et font référence à des concepts harmonisés.
- Comparabilité entre zones géographiques infra-nationales : elle sera permise par la production annuelle de résultats se rapportant tous aux mêmes dates. Néanmoins, selon les zonages, l'utilisateur devra prendre quelques précautions pour utiliser des variables conjoncturelles comme le chômage. Le rapport du groupe du Conseil national de l'information statistique sur la diffusion donne à ce sujet des exemples et propose des méthodes pour utiliser au mieux les résultats disponibles.

**« Cohérence » :**

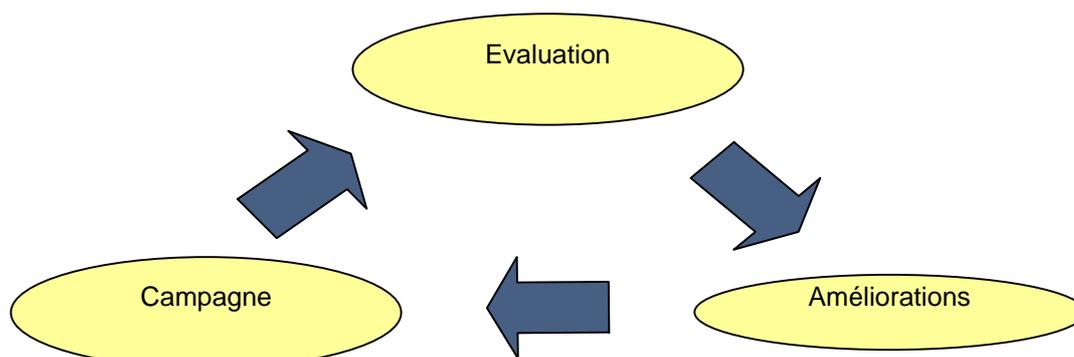
Elle est assurée par les redressements qui visent, entre autres, à rectifier les quelques inévitables incohérences entre variables.

Pour ce qui est de la cohérence avec d'autres sources de données locales, un enjeu important sera de « combiner » le recensement avec d'autres sources locales : le temps où le recensement de la population était la seule source localisée disponible est révolu.

## C/ La gestion de la qualité au recensement

### 7. La démarche d'ensemble

Le caractère annuel des opérations de recensement permet une prise en compte des évaluations selon le schéma suivant :



Les évolutions décidées au vu d'une campagne N ne sont pas nécessairement mises en œuvre dès N+1 ; c'est parfois en N+2 voire N+3. Il faut en effet se donner le temps d'instruire les différentes questions, de préparer les changements et de les mettre en œuvre dans toutes les étapes : protocoles, applications informatiques, information, formation... On constate que souvent derrière une évolution simple en apparence se cachent des coûts importants de conception (pour garantir la cohérence d'ensemble) et de mise en œuvre (prise en compte dans les applications informatiques, les imprimés, les protocoles d'organisation, la formation des acteurs).

L'amélioration des processus est conduite par une démarche d'évaluation systématique de l'ensemble des processus :

- à l'Insee, par des bilans faits auprès de l'ensemble des acteurs, y compris les coordonnateurs communaux. Ces bilans, qualitatifs et quantitatifs, portent sur la pertinence des processus, l'utilisation de telle ou telle fonctionnalité applicative, les organisations mises en place et les charges de travail. Ils sont menés par voie de questionnaires mais sont fréquemment enrichis par des rencontres bilatérales ;
- auprès des sous-traitants chargés de la saisie, qui établissent en fin de campagne un bilan complet des opérations ;
- et via la Commission nationale d'évaluation du recensement associant l'Insee, les communes et les utilisateurs (cf. partie 3) qui se réunit deux à trois fois par an ; cette commission a été créée au sein du Conseil national de l'information statistique (Cnis). Présidée par un sénateur, elle associe l'Insee, les communes, des utilisateurs du recensement (administrations, associations d'élus, chercheurs, etc.). Elle évalue les processus de collecte et de contrôle, propose des adaptations des différents protocoles et devra se prononcer sur les changements souhaitables des textes qui régissent l'organisation des opérations de recensement ;
- enfin, des mesures et des contrôles de qualité sont pratiqués à toutes les phases du dispositif (couverture de la base de sondage, collecte, saisie et traitements statistiques, diffusion).

### **Ce qui change pour le suivi de la qualité par rapport à un recensement général**



Du fait du recours au sondage et de l'annualisation de la collecte, les volumes à traiter (sept fois moins de questionnaires que dans un recensement classique) et le nombre d'acteurs (cinq fois moins de communes, six fois moins d'agents recenseurs) sont réduits. Les efforts de l'Insee, des coordonnateurs communaux (qui sont les responsables du recensement désignés par les maires dans les communes) et des agents recenseurs sont plus concentrés, permettant une collecte de meilleure qualité et des contrôles plus approfondis.

Dans les communes de 10 000 habitants ou plus, chaque agent recenseur dispose de la liste des adresses auxquelles il doit recenser tous les logements. C'est un point très important car l'agent recenseur sait ainsi précisément où il doit aller, ce qui est plus efficace que la technique du « ratissage » des recensements généraux.

## **8. Le contrôle et la validation des données, processus par processus**

La qualité du recensement est contrôlée pour chacun des processus. Aux évaluations évoquées dans le paragraphe ci-dessus s'ajoutent des opérations de contrôle lors des principales phases : contrôle de la qualité de la base de sondage ; contrôle de la qualité de la collecte ; contrôle de la saisie et de la codification ; contrôle de la qualité des données avant diffusion.

Ainsi, l'Insee contrôle ses répertoires par des enquêtes de terrain, et la collecte sur le terrain par la confrontation avec des sources administratives, éventuellement arbitrée par une enquête de terrain. Quant aux opérations de saisie et de codification, elles font l'objet d'une double saisie et d'une double codification sur un échantillon des bulletins, suivies dans les deux cas d'une analyse des divergences.

### **8.1. Les chiffres de population**

Les éléments suivants sont nécessaires pour le calcul de la population des communes :

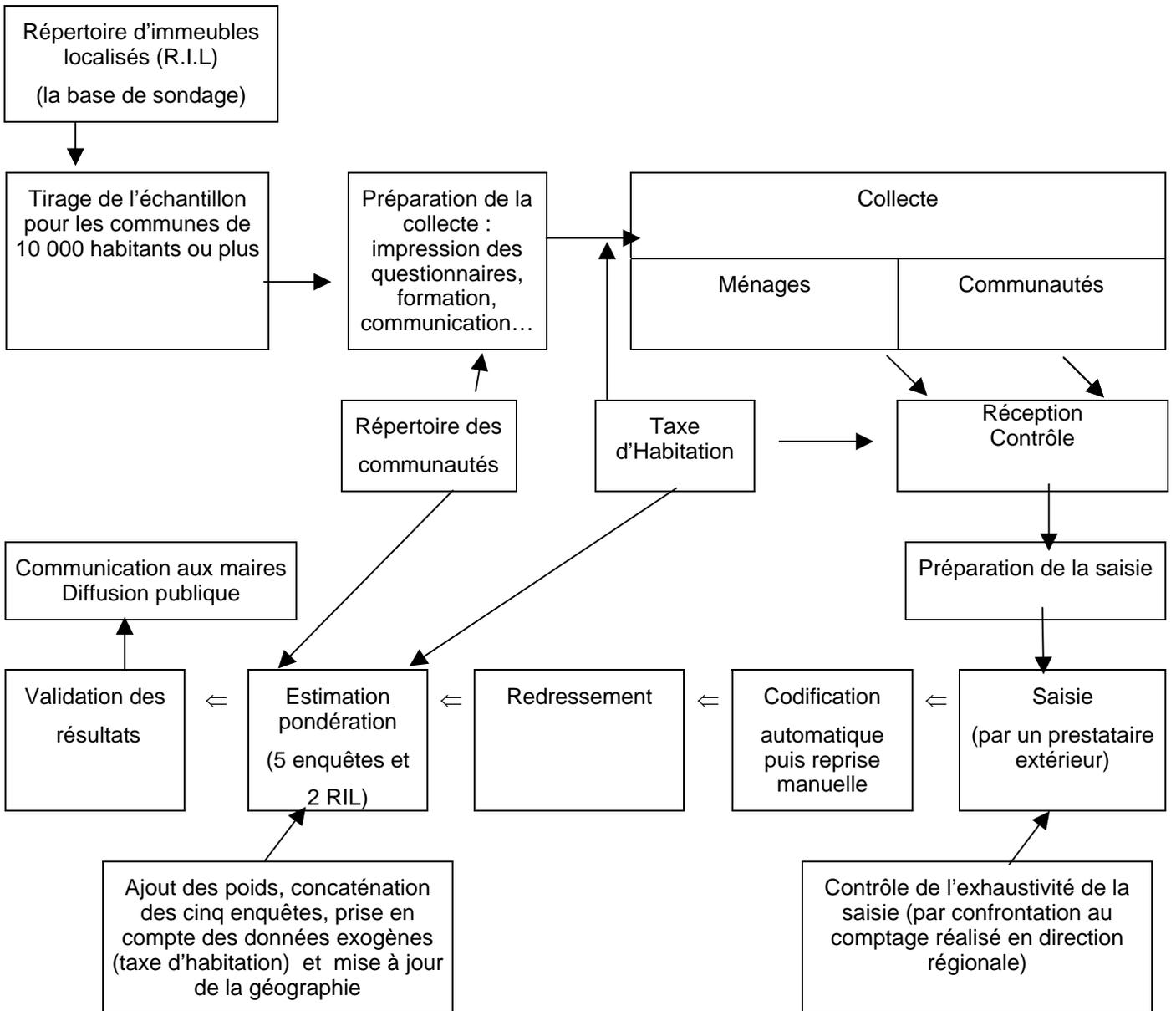
- le stock de logements tiré du répertoire d'immeubles localisés (RIL) ;
- les fichiers issus des différentes collectes ;
- le répertoire des communautés ;
- le fichier de la taxe d'habitation.

La validation porte sur chacun de ces éléments. Elle est menée au fur et à mesure de leur élaboration ou de leur traitement, car les erreurs sont d'autant plus faciles à corriger qu'elles sont détectées précocement.

La validation comprend également l'examen de deux aspects pouvant jouer sur la qualité ou l'interprétation des données : les évolutions dues aux différences de concepts d'un recensement à l'autre et l'effet des redressements. Dans les communes comptant au moins 10 000 habitants, du fait d'une étape de pondération et d'extrapolation des résultats du sondage, la validation porte aussi sur l'effet des « grandes adresses », qui peuvent générer des effets de grappe.

La validation a deux objectifs : d'une part s'assurer de la qualité statistique des résultats que l'on va publier. D'autre part, comprendre les évolutions, estimer leur robustesse et en déduire des précautions d'emploi. Cet aspect est particulièrement important si des différences de concepts affectent particulièrement une commune donnée.

### LES ETAPES DU RECENSEMENT



## a) Le stock de logements tiré du RIL

La qualité du RIL est essentielle pour la qualité du recensement : si la base de sondage comportait trop de déficit, elle ne serait pas représentative. Si elle comportait trop d'excédents, cela alourdirait inutilement la collecte (nombreux déchets). Enfin, un bon équilibre excédent-déficit est nécessaire pour calculer des estimations sans biais au moment du traitement des données.

Au niveau national, voire régional, la mesure de la qualité du répertoire d'immeubles localisés (RIL) est réalisée grâce à une enquête annuelle, menée sur le terrain par des enquêteurs de l'Insee.

Son objectif est de mesurer l'exhaustivité du RIL en termes de déficit et d'excédent d'adresses de la catégorie *habitation*, et de logements associés. Elle vise à s'assurer que toutes les adresses rencontrées sur le terrain sont répertoriées dans le RIL et inversement que toutes les adresses du répertoire existent sur le terrain.

### Le mode opératoire de l'enquête de mesure de la qualité

L'enquête porte sur un échantillon d'îlots. Un îlot est en général un pâté de maisons mais ce peut être une partie d'un grand ensemble immobilier. Pour chacun d'eux, les enquêteurs vont sur le terrain avec la liste de l'ensemble des adresses de l'îlot figurant dans le RIL : les habitations et les autres catégories (établissements, équipements, « autres »). Ils vérifient l'existence de toutes ces adresses, avec des consignes de priorité sur les habitations. Une adresse d'habitation figurant sur la liste et absente sur le terrain constitue un excédent. C'est également le cas d'une adresse qui se révèle sur le terrain n'être qu'un accès secondaire. Une adresse d'habitation existant sur le terrain et absente de la liste constitue un déficit. Les enquêteurs contrôlent également l'adressage, le nombre de logements, les confusions de catégories et les adresses avec plusieurs accès.

Rappelons que l'objectif principal de l'enquête est le contrôle de l'exhaustivité du RIL habitation en termes d'adresses et de logements et en particulier la détection d'un éventuel déficit d'adresses.

L'échantillon de l'enquête de 2007 a porté sur 340 000 adresses et 900 000 logements.

### Les résultats de l'enquête de mesure de la qualité

#### *Déficits et excédents d'adresses d'habitation et de logements du RIL*

		Ril 2003	Ril 2004	Ril 2005	Ril 2006	Ril 2007
<b>Adresses</b>	<i>En déficit</i>	2,3%	1,9%	1,3%	1,3%	1,4%
	<i>En excédent</i>	2,0%	1,9%	1,8%	1,7%	1,4%
	<i>Solde excédent/déficit</i>	-0,3%	0%	0,5%	0,4%	0%
<b>Logements</b>	<i>En déficit</i>	2,3%	1,9%	1,1%	0,9%	1%
	<i>Excédent</i>	2,1%	2,3%	1,8%	1,4%	1%
	<i>Solde excédent/déficit</i>	-0,2%	0,4%	0,7%	0,5%	0%

Source : Insee, enquêtes sur la qualité du RIL de 2003 à 2007

## b) Les fichiers tirés de la collecte

La qualité de ces fichiers tient à deux facteurs : la qualité de la collecte et celle de la saisie. Par ailleurs, les redressements effectués visent à améliorer la qualité, par correction des non-réponses ou des réponses incohérentes. L'étape de codage fait également l'objet d'un contrôle de qualité.

**La collecte** est souvent le maillon le plus délicat du processus statistique. Sa qualité repose sur des procédures précises et rigoureuses, mais également sur l'adhésion des répondants (qui conditionne la sincérité de leurs réponses) et l'engagement de ceux qui recueillent l'information, c'est-à-dire les agents recenseurs. Une préparation rigoureuse (repérages sur le terrain, formation des acteurs) et



des contrôles permettent d'en assurer la qualité. Des contrôles sont réalisés par les communes et l'Insee pendant la collecte. Après la collecte, le contrôle est du ressort de l'Insee ; il permet de vérifier les données collectées et le cas échéant de les corriger.

#### La préparation de la collecte et les contrôles au fil de la collecte

La collecte est réalisée sous la responsabilité de la commune. Le mode opératoire de collecte repose sur un protocole précis, à la charge des agents recenseurs, sous le contrôle du responsable communal et du superviseur de l'Insee. Le manuel de l'agent recenseur explicite les tâches à accomplir. Des documents détaillés sont fournis aux responsables communaux et aux superviseurs. En cours et à l'issue de la collecte, la commune vérifie que l'agent recenseur est bien passé dans chaque logement à recenser. Aucune adresse, aucun logement ne doit avoir été oublié.

Les superviseurs effectuent aussi des contrôles lors de leurs visites dans les communes. Ils vérifient le respect des procédures (confidentialité, respect du calendrier) et le remplissage des questionnaires déjà collectés, ce qui leur permet de détecter les erreurs de collecte les plus fréquentes. En fin de collecte, quand ils transmettent les questionnaires à l'Insee, ils indiquent leur appréciation sur la qualité de la collecte dans la commune et indiquent les éventuelles difficultés rencontrées.

Outre ce dispositif de base, les communes de 10 000 habitants ou plus mettent en place un dispositif de suivi et de contrôle plus précis.

En amont de la collecte, le contrôle des adresses de l'échantillon à recenser est réalisé en deux temps : validation des adresses et du nombre de logements par l'équipe centrale de la commune, puis repérage sur le terrain par l'agent recenseur, qui doit authentifier les adresses et dénombrer les logements (cette phase permet aussi à la commune de tester l'agent recenseur). Un deuxième repérage sur le terrain est effectué en cas de désaccord entre l'agent recenseur et l'équipe centrale (un seuil d'écart peut être accepté, selon les moyens de l'équipe communale). Ainsi quand la collecte débute, il y a accord Insee / coordonnateur communal / agent recenseur sur l'existence de l'adresse, son classement en adresse habitée et le nombre de logements à y recenser. La rigueur de ce contrôle amont garantit une collecte plus « confortable » à l'agent recenseur.

Dans toutes les communes, le suivi de la collecte au fil des opérations et pour chaque agent recenseur permet de vérifier l'avancement de la distribution et du retour des imprimés et de détecter certaines difficultés rencontrées par des agents recenseurs, ou des carences. Pour ce suivi, l'Insee fournit aux communes un outil de contrôle et de suivi de la collecte, mais les communes peuvent également développer ou faire développer leur propre outil, si elles souhaitent des fonctionnalités plus riches. L'examen des premiers documents retournés par l'agent recenseur permet, en analysant la qualité de leur remplissage, d'évaluer la compréhension des consignes par l'agent recenseur et son implication dans la collecte.

Certaines communes procèdent avec chaque agent recenseur à un bilan hebdomadaire d'une durée d'environ 2 heures. Son objectif est de vérifier que sont bien recensés toutes les adresses, tous les logements d'une adresse, toutes les personnes d'un logement. Chaque semaine, l'agent recenseur doit respecter des objectifs chiffrés (imprimés distribués, documents collectés en retour). En fin de gestion, les dernières difficultés de collecte sont analysées. Plusieurs indicateurs sont suivis : le nombre et le pourcentage de FLNE (elles doivent être établies le plus tard possible, en phase de clôture des opérations), le taux de vacance des logements, la qualité du carnet de tournée.

#### Après la collecte : les contrôles menés par l'Insee

Après la réception à l'Insee des documents transmis par les communes, des vérifications sont menées en bureau, puis sur le terrain. Le protocole de contrôle est adapté à la nouvelle configuration des enquêtes de recensement. Par rapport aux recensements généraux, des vérifications plus approfondies sont rendues possibles pour chaque commune par le moindre volume de données collectées chaque année et le plus faible nombre de communes. Le fichier de la taxe d'habitation joue un rôle important dans les contrôles en bureau et permet de vérifier l'exhaustivité de la collecte dans des communes ou des zones de collecte données. Ces contrôles, en bureau ou sur le terrain, sont menés « au plus près de la collecte » et les erreurs qu'ils permettent de détecter sont corrigées.



Ils sont sélectifs car ils résultent de « tamis » successifs :

- un premier contrôle lors de la réception et de l'enregistrement des questionnaires permet d'apprécier la qualité de la collecte pour chacune des communes ;
- cette appréciation débouche, pour environ 15% des communes représentant 30% de la population, sur un contrôle approfondi en bureau ;
- enfin, les cas non résolus en bureau font l'objet de contrôles menés sur le terrain par l'Insee (environ 28 000 contrôles ont été effectués sur le terrain en 2007, concernant 8% des communes).

#### Les contrôles en bureau

La première étape, dite « Réception-enregistrement des questionnaires », porte sur les documents de toutes les communes concernées par l'enquête de recensement de l'année. Les agents de l'Insee en bureau comptent les questionnaires par enregistrement de codes à barres. Ils vérifient les adresses, les bulletins vierges... C'est la phase du « flashage ». Innovation importante du nouveau recensement, ce processus permet d'identifier chaque questionnaire et de le suivre tout au long de la chaîne. Il permet également d'établir des indicateurs par commune et par zone de collecte sur l'évolution du nombre de logements et de personnes, la structure du parc de logements (résidences principales, secondaires, logements occasionnels, logements vacants), la proportion de fiches de logement non enquêté (FLNE). A l'issue de cette étape, ils font une expertise de chaque commune en tenant compte de ces indicateurs quantitatifs, d'indicateurs qualitatifs sur le déroulement de la collecte fournis par les superviseurs et de leur appréciation sur la réception des documents. Ils établissent, pour chaque commune, par une technique de « scoring », une note « qualité », et mettent en contrôle approfondi les communes dont les scores sont les moins bons (environ 15% des communes, représentant 30 % de la population).

Le contrôle approfondi en bureau constitue la deuxième étape.

Il peut porter soit sur la commune entière soit sur une portion de commune. Il s'appuie majoritairement sur les fichiers de la taxe d'habitation et est orienté, en fonction du problème mis en évidence par les indicateurs :

- sur l'exhaustivité du nombre de logements :
  - présence ou non de l'adresse ;
  - comparaison du nombre de logements ;
  - comparaison des noms des occupants (communes de 10 000 habitants ou plus) ;
- sur le pourcentage de fiches de logement non enquêté (FLNE) :
  - vérification des FLNE en s'appuyant sur l'information « catégorie de logement » de la TH, nombre supposé de personnes résidant dans le logement ;
- sur la structure par catégorie de logement (résidence principale, résidence secondaire ou occasionnelle, logement vacant) :
  - confrontation avec le fichier de la TH.

Il peut également être procédé à des enquêtes de contrôle par téléphone auprès des habitants.

Tous ces contrôles permettent soit de confirmer la collecte soit de mettre en évidence des erreurs (par exemple des résidences secondaires recensées à tort comme résidences principales). Celles-ci sont alors corrigées.

#### Troisième étape : les contrôles sur le terrain

En cas de divergences nombreuses ne pouvant être surmontées en bureau, les communes font l'objet d'une décision de contrôle sur le terrain. La commune en est alors informée mais elle ne dispose pas de la liste des adresses ou des logements concernés. Ces contrôles portent sur des communes pour lesquelles la qualité est jugée insuffisante (soit pour l'ensemble de la commune, ce qui est très rare, soit pour une partie de la commune) au moment du contrôle en bureau par l'Insee. Dans la très grande majorité des cas, les contrôles confirment le résultat de la collecte, mais ils permettent parfois de détecter et de corriger certaines erreurs. Les contrôles donnent lieu, le cas échéant, à des



corrections de questionnaires, et les communes en sont informées. Ils permettent aussi d'adapter les redressements et de modifier pour l'année suivante les protocoles de collecte ou de vérification.

Les contrôles sur le terrain ont lieu entre avril et juin. Ils sont menés par des enquêteurs de l'Insee qui reçoivent à cet effet une formation appropriée.

Au total plus de 28 000 contrôles ont été menés sur le terrain en 2007 auprès de 691 communes, soit 8,5% des communes concernées par l'enquête de recensement. Près d'un tiers (31,7%) des communes de 10 000 habitants ou plus ont fait l'objet d'un contrôle sur le terrain. Comme ces contrôles portent en priorité sur les cas les plus difficiles ou les incohérences avérées, leur exploitation statistique au niveau de la collecte tout entière n'aurait pas de sens.

Les contrôles de terrain portent soit sur des adresses, soit sur des logements.

Pour les adresses, on contrôle des adresses qui ne semblent pas correspondre à celles demandées dans l'échantillon (communes de 10 000 habitants ou plus), ou celles qui risquent d'avoir été oubliées ou, a contrario, celles que l'on ne retrouve dans aucune autre source (communes de moins de 10 000 habitants). On vérifie aussi à cette occasion le nombre de logements.

6 400 contrôles d'adresse ont été menés sur le terrain en 2007. Cela représente 1,3 adresse pour 1 000 pour les communes de moins de 10 000 habitants et 7,3 adresses sur 1000 dans les communes de 10 000 habitants ou plus.

Les contrôles portent aussi sur des logements, quand le logement a fait l'objet d'une FLNE ou qu'il y a un doute sur sa catégorie (principale, secondaire, vacant, occasionnel). On vérifie alors aussi le nombre de personnes.

Environ 22 000 contrôles sur les logements ont été menés en 2007, soit un contrôle pour 200 logements.

Une grande majorité de ces contrôles (79%) porte sur les logements ayant fait l'objet d'une fiche de logement non enquêté (FLNE). Les contrôles permettent de prendre en compte, dans les résultats statistiques (chiffres de population et résultats détaillés), la population qui n'a pu être contactée directement par les agents recenseurs. Les FLNE établies le sont effectivement pour des résidences principales (4 cas sur 5) et l'estimation du nombre de personnes habitant les logements concernés est robuste (dans 4 cas sur 5 le nombre de personnes est correctement estimé et la part de cas où le nombre de personnes est supérieur à celui inscrit sur la FLNE est équivalente à la part de cas où ce nombre est inférieur).

Les autres contrôles de logements visent à vérifier la catégorie de logement. Celle-ci est confirmée dans 75 à 80% des cas, selon le type de logement. Rappelons ici que les contrôles sont concentrés sur les cas où des anomalies ont été détectées ou sont suspectées.

En conclusion de cette revue des contrôles effectués sur la collecte, il faut répéter que c'est l'un des grands avantages de la nouvelle méthode : du fait de la moindre taille de chaque enquête de recensement, et aussi grâce au système des tamis successifs, les contrôles sont beaucoup plus efficaces que dans un recensement classique.

#### La validation de la saisie

**La saisie** est effectuée par un prestataire externe. Il s'agit d'une saisie optique : les questionnaires sont scannés puis soumis à un logiciel de reconnaissance des formes (chiffres, lettres, croix, codes à barres). Une saisie manuelle des libellés manuscrits non reconnus par la machine complète le processus. Un cahier des charges précis définit ce qu'on attend du prestataire sur la qualité. Il concerne à la fois la sécurité des données, pendant les transports et chez le prestataire (sécurité physique et informatique, engagement de chaque personne à ne pas divulguer les informations traitées) et la qualité des données (nombre de documents saisis par zone de collecte, taux d'erreur



maximal par variable). Enfin il y a comptage des documents, et contrôle des écarts avec le premier comptage réalisé à l'Insee lors du flashage (cf. ci-dessus).

**Au cours du processus, les documents (bulletins individuels et feuilles de logement) sont donc comptés trois fois** : une fois par la commune lors du bouclage de la collecte, une fois par l'Insee lors de la phase de réception et de contrôle, une fois par le prestataire de saisie. La confrontation de ces trois comptages permet de garantir un niveau très élevé de qualité sur le nombre de bulletins.

Enfin, la qualité de la saisie est mesurée par une double saisie (cf. encadré)

#### *Une double saisie pour vérifier la qualité des variables saisies*

*Les erreurs de saisie sont évaluées grâce à une deuxième saisie effectuée sur un échantillon de bulletins par un second prestataire, qui travaille à partir des images scannées des questionnaires de cet échantillon en utilisant les mêmes règles de saisie. Les divergences entre les deux saisies sont analysées à l'Insee, de façon à établir des taux d'erreur imputables au prestataire principal sur le codage de chacune des variables. La double saisie est réalisée au fil de la campagne de saisie, ce qui permet, éventuellement, de corriger les protocoles de saisie pour les lots suivants.*

*Des contacts et des réunions régulières sont organisés avec le prestataire, de manière à anticiper les problèmes ou à les résoudre au plus vite. Des ajustements ou des améliorations lui sont demandés, en cours de campagne ou en période inter-campagnes s'il s'agit d'améliorations plus lourdes, suite aux enseignements de la double saisie. Par exemple, l'optimisation des outils de reconnaissance de forme a permis de diminuer significativement le taux d'erreur de saisie des dates de naissance. Il faut noter à ce sujet que, même sur des données très simples comme les dates de naissance, il est illusoire d'espérer un taux d'erreur nul : l'information étant manuscrite au départ, des erreurs d'interprétation sont inévitables.*

**Les redressements** sont destinés à améliorer la qualité des fichiers tirés de la collecte. Il s'agit d'analyser les non réponses, totales ou partielles, de manière à les remplacer par une information valide, mais aussi de corriger les bulletins qui présentent des incohérences (par exemple une personne de 95 ans qui déclare être active). La procédure utilisée est celle du « hot-deck » (déjà mise en œuvre lors des recensements précédents), qui consiste à compléter l'information manquante en s'aidant de la réponse d'un « donneur » qui présente certaines caractéristiques communes. Là encore, les redressements s'affinent d'année en année à partir des résultats précédents, ce qui constitue un système efficace d'amélioration de la qualité.

En particulier, le redressement des FLNE permet d'estimer, pour le calcul de la population des communes, la population des résidences principales pour lesquelles le contact avec l'agent recenseur n'a pas eu lieu (absences de longue durée, personnes impossibles à joindre, refus). On rappelle que pour ces logements, l'agent recenseur établit une FLNE en indiquant le nombre d'habitants du logement (cf. page 4, le paragraphe sur les FLNE.). Pour les résidences principales dont le nombre d'occupants est renseigné (plus de 8 sur 10), on « impute » pour le logement autant d'individus que la FLNE en indique. Pour les autres FLNE, on procède également à une imputation, selon une méthode visant à reproduire la structure, par taille de ménage, observée en moyenne nationale sur les FLNE renseignées.

#### **c) Le répertoire des communautés.**

Comme le RIL, le répertoire a été constitué à partir des données collectées lors du recensement de 1999 et il est mis à jour avec des sources administratives (fichier des établissements sanitaires et sociaux, fichier des internats, fichier des établissements pénitentiaires ou relevant de la Protection judiciaire de la jeunesse,...). Comme le RIL, il est transmis aux communes pour expertise juste avant l'enquête de recensement. Les communes reçoivent la liste de leurs communautés et il leur est demandé d'indiquer les communautés qui auraient disparu, celles dont l'Insee n'aurait pas connaissance et celles dont l'adresse aurait changé. L'Insee valide ensuite les propositions d'ajout ou de retrait du répertoire formulées par les communes.

Au démarrage de la collecte, une ultime validation de la liste des communautés est demandée à l'agent recenseur des communautés.



Enfin, l'Insee confronte chaque année le RIL et le répertoire des communautés. Il s'agit d'une part de bien géolocaliser chacune des communautés du répertoire ; le RIL est alors utilisé comme liste de référence pour les adresses et leurs caractéristiques géographiques (coordonnées X,Y, appartenance à un quartier, à un canton, à une zone infra-communale). Il s'agit d'autre part de vérifier l'absence de doublons ou d'omissions entre ces deux répertoires (notamment quand une communauté se transforme en ensemble de logements ordinaires ou inversement).

#### **d) Le fichier de la taxe d'habitation**

L'Insee reçoit chaque année de la direction générale des finances publiques (DGFIP) un fichier récapitulatif de l'ensemble des locaux taxés au titre de la taxe d'habitation. Un premier traitement consiste à écarter les locaux qui ne sont pas des habitations (garages, parkings notamment).

Le fichier est ensuite utilisé à deux fins :

- suivi et contrôle de la collecte, pour les communes concernées ;
- établissement de séries du nombre total de logements, du nombre de résidences principales et du nombre de résidences secondaires. Ces séries sont utilisées pour le calcul des populations des communes de moins de 10 000 habitants.

Les communes reçoivent par ailleurs de la DGFIP l'information sur le nombre de rôles émis et sur l'assiette de l'impôt. Leurs réactions, ainsi que celles des contribuables, concourent à l'amélioration de la source. En effet, toute omission se traduisant par de moindres recettes fiscales, la commune a intérêt à signaler les omissions à la DGFIP. Par ailleurs, tout excédent (local figurant à tort dans la liste) se traduisant pour le contribuable concerné par une imposition supplémentaire, les excédents sont également signalés à la DGFIP.

L'Insee analyse systématiquement le profil des séries ainsi constituées et, en cas d'évolution atypique (par exemple une forte baisse suivie d'une forte hausse) interroge la DGFIP pour détecter d'éventuels artefacts de gestion. Les artefacts confirmés par la DGFIP sont corrigés.

L'évaluation de la qualité de cette source indique que, même si le fichier n'utilise pas les mêmes concepts que ceux du recensement, il fournit des indications fiables sur l'évolution du parc de logements. Et c'est bien cette évolution, et non pas le niveau, qui intervient pour le calcul de la population.

#### **e) Les écarts de concepts entre le recensement général de 1999 et le nouveau recensement**

Ces écarts peuvent jouer sur l'analyse des évolutions de la population d'une commune. Par exemple, le fait de compter, au nouveau recensement, les élèves internes majeurs dans la commune de leur internat et non plus dans la commune de résidence de leurs parents fait évoluer fortement la population municipale de la commune de l'internat.

Il convient donc d'établir une liste des communes a priori fortement impactées par ces écarts pour éclairer la validation de ces communes et les analyses statistiques reposant sur ces communes (ou des zones englobantes). Cet effet doit être signalé aux utilisateurs des données.

#### **f) La validation finale**

Chacun des ingrédients étant validé au fur et à mesure de sa mise en place, il n'y a pas à proprement parler d'étape finale de validation. C'est surtout l'occasion de rassembler tous les éléments qualitatifs et quantitatifs qui permettent d'expliquer le chiffre de population. C'est essentiel pour répondre aux éventuelles questions que les communes, mais aussi d'autres utilisateurs des chiffres de population, peuvent être amenés à poser.

## **8.2. La validation des résultats détaillés**

Au-delà du seul chiffre de population, la validation concerne l'ensemble des résultats détaillés. Certes, la validation de la qualité du chiffre total garantit que la plupart des problèmes de qualité (couverture, effets liés à l'échantillonnage....) ont été détectés et qu'il en a été tenu compte. En complément, il est procédé à deux validations complémentaires.

### **La validation de la codification des questionnaires**

**Le codage** comprend d'abord une phase automatique de codification (selon des nomenclatures d'activité et de profession notamment), puis une phase de reprise par des opérateurs en direction régionale des cas non codés. L'Insee met en œuvre actuellement un dispositif de contrôle de qualité sur ce processus. Le principe est de procéder à une seconde codification sur un échantillon de bulletins représentatifs des différents modes de codification (automatique et reprise), puis d'analyser les divergences.

L'objectif est de mesurer la qualité du codage automatique et celui de la reprise manuelle (obtenir d'abord un pourcentage global de cas bien codés, puis à terme des pourcentages par grande modalité des variables, par exemple pourcentage de cadres supérieurs bien codés).

Ce dispositif de mesure et de contrôle permet aussi d'enrichir les fichiers d'apprentissage qui servent à la codification automatique, de manière à en améliorer la performance, de mieux cibler la formation des opérateurs qui assurent la reprise des échecs de la codification automatique, et d'assurer finalement un bon management de la qualité. L'apport de cette étape, mise en œuvre pour la première fois en 2006, a été sensible dès la campagne 2007.

### **Un contrôle de vraisemblance sur les comptages détaillés.**

Les résultats détaillés sont analysés à partir de « tris à plat » (comptages des occurrences des différentes variables : pyramide des âges, état matrimonial, structures familiales, emploi, parc de logements, etc.). Il s'agit d'une part de confronter ces résultats avec ceux provenant d'autres sources, d'autre part d'apprécier leur robustesse en analysant leurs évolutions d'une enquête à l'autre. L'analyse des niveaux peut permettre d'évaluer la qualité des procédures de redressement. Couplée avec des estimations de la précision, l'analyse de la robustesse doit permettre de définir le niveau de finesse souhaitable de l'information diffusée (par exemple diffusera-t-on des âges quinquennaux ou des âges décennaux ?). La validation est aussi l'occasion de cerner et de quantifier si possible l'effet des changements de questionnaires. C'est une information essentielle pour les utilisateurs qui veulent analyser des évolutions par rapport au recensement précédent.

Les analyses menées sur les premières années ont permis de déceler certains effets indésirables de la procédure de redressement (que l'on a limités pour les enquêtes suivantes) ou de protocoles de saisie trop imprécis (erreur rectifiée à partir de 2005). Elles ont également donné lieu à des analyses des effets du changement de certaines questions (notamment sur l'emploi et sur les structures familiales).

Actuellement, un « crible » analyse les évolutions des principales variables statistiques et invite les directions régionales à analyser les cas où ces évolutions posent problème.

## **9. Les perspectives pour l'analyse de la couverture du recensement.**

On l'a vu, un travail important de contrôle *in itinere* est mené au moment de la collecte du recensement et dans les trois mois qui suivent la réception des questionnaires. Ces contrôles donnent lieu à correction immédiate de la feuille de logement et du ou des bulletins individuels concernés.

Au niveau global, il faut aussi s'assurer que le recensement n'a pas compté plusieurs fois les mêmes personnes et n'en a pas oublié d'autres. L'analyse de la couverture du recensement prendra plusieurs formes :



1. Une analyse du sur-dénombrement, fondée sur la détection de personnes recensées en double. On s'appuiera pour cela sur l'échantillon démographique permanent (EDP), échantillon au 1/100e de la population française, alimenté par les données d'état civil et les formulaires de recensement. La procédure d'alimentation de cet échantillon permet de détecter si plusieurs bulletins sont établis pour une même personne. Compte tenu de la méthode de recensement, il convient de disposer des questionnaires de cinq campagnes successives pour en tirer des enseignements fiables. En particulier, on disposera d'une estimation du volume global de sur-dénombrement et on pourra caractériser les populations concernées, notamment en termes de résidence (grande commune ou petite commune, ménage ordinaire ou communauté) ou de sexe et d'âge. Du fait des modalités du recensement et surtout du taux de couverture de l'EDP, les résultats de cette analyse n'auront qu'une portée nationale.
2. Une analyse des omissions, qui pourra s'appuyer sur des fichiers administratifs de couverture universelle tels que les fichiers de type « Assurance-Maladie » ou des fichiers fiscaux.

Ces pistes sont plus prometteuses que celles offertes par une enquête post-censitaire du type de celles qui ont été menées en France en 1962 et 1990. Il paraît en effet plus pertinent de contrôler la qualité d'une enquête en la confrontant à une source administrative plutôt qu'à une autre enquête, qui présenterait les mêmes imperfections (en particulier les biais dus aux difficultés à contacter les ménages).

Ces investigations, centrées sur la collecte, compléteront celles, décrites au 8.1, sur la qualité du RIL.