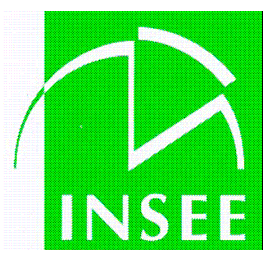
Direction des Statistiques Démographiques et Sociales

N° F1003

Construire une pyramide des âges pertinente pour le calcul des indicateurs démographiques à partir des enquêtes annuelles de recensement

Stéphane JUGNOT, Marie ANGUIS et Catherine BEAUMEL

Document de travail



Institut National de la Statistique et des Études Économiques

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

Série des Documents de Travail
de la
DIRECTION DES STATISTIQUES DÉMOGRAPHIQUES ET SOCIALES
Unité des Études Démographiques et Sociales

N°F1003

Construire une pyramide des âges pertinente pour le calcul des indicateurs démographiques à partir des enquêtes annuelles de recensement

Stéphane JUGNOT, Marie ANGUIS et Catherine BEAUMEL

Juin 2010

Construire une pyramide des âges pertinente pour le calcul des indicateurs démographiques à partir des enquêtes annuelles de recensement

Stéphane Jugnot, Marie Anguis, Catherine Beaumel

Juin 2010

Les effectifs par sexe et âge du recensement servent de base pour construire la pyramide des âges nécessaire au calcul de la plupart des indicateurs démographiques : le taux de fécondité par âge, l'indicateur conjoncturel de fécondité, les quotients de mortalité par sexe et âge, l'espérance de vie, etc. La disponibilité des résultats définitifs du nouveau recensement fin 2008 a conduit à publier pour la première fois une pyramide des âges s'appuyant sur ce dernier, relative au 1er janvier 2006. Le nouveau mode de collecte du recensement modifie le contexte en proposant des résultats issus d'un cumul d'informations collectées sur cinq ans. Lorsqu'il se fait sur une seule année, le recensement propose une photographie de la population à une date donnée ; il est alors indifférent de construire une pyramide des âges à partir de l'âge atteint dans l'année ou selon l'année de naissance. Quand la collecte d'information s'étale dans le temps, il n'en est plus de même. Avoir 20 ans la première année du cycle de cinq ans, c'est en avoir 24, la dernière année du cycle. Cumuler les informations selon l'âge ne revient donc pas au même que les cumuler selon l'année de naissance (ou génération). La disponibilité de résultats définitifs des cinq premières enquêtes annuelles de recensement a donc conduit à développer une méthodologie adaptée au nouveau mode de collecte du recensement. Sans celle-ci, l'analyse des indicateurs démographiques détaillés par âge serait légèrement biaisée. Ce document présente les raisons qui obligent à mettre en place une nouvelle méthodologie, ainsi que cette nouvelle méthodologie.

Sommaire

Introduction

Le recensement sert habituellement de base à la comptabilité démographique. Outre les populations légales des communes, il détermine également la population statistique, « sans double compte » pour utiliser la terminologie employée jusqu'au recensement général de 1999. Rapportée au premier janvier de l'année du recensement quand cela est nécessaire, la population recensée détermine le niveau de population de référence pour les séries d'estimations de population. Les effectifs par sexe et âge, donc la pyramide des âges du recensement, servent également de base pour construire la pyramide des âges nécessaire au calcul de la plupart des indicateurs démographiques : le taux de fécondité par âge, l'indicateur conjoncturel de fécondité, les quotients de mortalité par sexe et âge, l'espérance de vie, etc.

La disponibilité des résultats définitifs du nouveau recensement fin 2008 a conduit à publier pour la première fois une pyramide des âges s'appuyant sur ce dernier. Comme eux, elle a pour date de référence le 1er janvier de l'année médiane du cycle de cinq ans, soit le 1er janvier 2006.

Auparavant, le passage de la pyramide des âges du recensement à une pyramide de référence pour le calcul des indicateurs démographiques consistait essentiellement à rapporter au 1er janvier de l'année du recensement la pyramide du recensement, relative à la date de collecte du recensement. Le nouveau mode de collecte du recensement modifie le contexte en proposant des résultats issus d'un cumul d'informations collectées sur cinq ans. Lorsqu'il se fait sur une seule année, le recensement propose une photographie de la population à une date donnée ; il est alors indifférent de construire une pyramide des âges à partir de l'âge atteint dans l'année ou selon l'année de naissance. Quand la collecte d'information s'étale dans le temps, il n'en est plus de même. Avoir 20 ans la première année du cycle de cinq ans, c'est en avoir 24, la dernière année du cycle. Cumuler les informations selon l'âge ne revient donc pas au même que les cumuler selon l'année de naissance (ou génération). La question du passage de la date du recensement au 1er janvier se pose également même si la date de référence est le 1er janvier, dans la mesure où la collecte sur le terrain se déroule sur quelques semaines à partir de mi janvier ; mais comme nous le verrons, cette question est de second ordre.

La disponibilité de résultats définitifs des cinq premières enquêtes annuelles de recensement a donc conduit à développer une méthodologie adaptée au nouveau mode de collecte du recensement (et à réviser les estimations annuelles réalisées depuis le recensement de 1999[1]). Après une présentation rapide de la nouvelle méthode de collecte du recensement, ses conséquences et les options envisageables sont abordées dans la partie I. La solution retenue pour les estimations de population de l'Insee est présentée dans la partie II. La partie III aborde les usages, pour rappeler que, pour la plupart des usages, les résultats du recensement, désormais disponibles annuellement, suffisent car la méthode présentée est destinée avant tout à disposer d'une pyramide des âges adaptée pour le calcul d'indicateurs démographiques nationaux détaillés selon l'âge.

Ce document de travail présente les travaux réalisés en 2008 pour le groupe de travail « pyramide des âges et indicateurs démographiques » à partir du cadre de réflexion fixé dans une note de Guy Desplanques du 12 juin 2007. Y ont participé : Stéphane Jugnot, Marie Anguis, Catherine Beaumel (INSEE, division Enquêtes et études démographiques), Laurent Toulemon (INED), Anne Thérèse Aerts, Noëlle Serruys (INSEE, département de l'action régionale), Olivier Léon (INSEE, PSAR Emploi et population).

[1] Ce point n'est pas traité dans ce document. Une note méthodologique est disponible sur www.insee.fr (« Révision des estimations de population nationale par sexe, âge et situation matrimoniale du 1er janvier 2000 au 1er janvier 2006 pour tenir compte des résultats du recensement 2006 », *Note méthodologique* du 15/01/2010). Les estimations révisées sont proposées dans « Données détaillées sur les estimations définitives de population et des indicateurs démographiques de 1999 à 2006 », *Insee Résultats*, n°106, Février 2010.

I- Les enquêtes annuelles de recensement : une nouvelle méthode et son incidence

Jusqu'au recensement général de population de 1999, le dénombrement de la population s'appuyait sur une collecte de bulletins aussi exhaustive que possible, concentrée sur un mois, les recensements étant espacés de plusieurs années. On disposait donc d'une photographie quasi instantanée de la population à la date moyenne du recensement, aux aléas de collecte près. Cette photographie de la population à la date du recensement était ensuite rapportée au 1er janvier de l'année de collecte pour disposer d'une pyramide des âges au 1er janvier. Celle-ci servait de référence pour caler la série des pyramides des âges annuelles et calculer les différents indicateurs démographiques. Pour rapporter la pyramide à la date du recensement au 1er janvier de l'année du recensement, il suffisait de réimputer les décès selon le sexe et l'âge survenus entre le 1er janvier et la date du recensement, de supprimer les naissances intervenues sur la même période et de retirer également le solde migratoire estimé. En d'autre termes : ramener à la vie les décédés, ne pas faire naître les nouveau-nés, faire sortir les immigrants et entrer les émigrants sur la période de quelques semaines considérée.

Le nouveau mode de collecte du recensement, mis en œuvre depuis 2004, s'appuie sur des enquêtes annuelles, dont la collecte reste concentrée sur une courte période, auprès d'un échantillon de logements : pour les « petites communes »[2], une sur cinq est recensée exhaustivement chaque année ; pour les autres communes, dites « grandes communes », un échantillon de 8% des logements est recensé chaque année[3]. Les résultats du recensement, notamment les populations légales des communes, sont ensuite calculés à partir de cinq collectes successives. Pour cela, une pondération spécifique, dite de cumul, est utilisée. Celle-ci ne se déduit pas de la pondération de chaque enquête annuelle[4].

La méthode d'estimation utilisée doit en effet fournir une estimation de la population totale relative à l'année médiane, quel que soit le niveau géographique, de façon à assurer une égalité de traitement entre les communes, quelle que soit leur taille et, pour les petites communes, quelle que soit leur année de collecte. Pour les grandes communes, l'estimation de population s'appuie sur un répertoire d'immeubles localisés daté du 1er janvier de l'année médiane, donc sur une base exhaustive des immeubles à cette date. Pour les petites communes, la population est extrapolée pour les communes recensées avant l'année médiane et interpolée pour celles recensées après l'année médiane, en utilisant les statistiques de la taxe d'habitation et une estimation de l'évolution de la taille moyenne des ménages.

Outre la disponibilité de données définitives désormais chaque année à partir de fin 2008**, la collecte étalée dans le temps constitue le changement le plus important par rapport à l'ancienne méthode pour le calcul de la pyramide des âges de référence.** Ce point est abordé dans la partie I-A. La partie I-B abordera le passage des pondérations annuelles à la pondération de cumul et les légères distorsions de structures qu'il induit. Quelques autres changements de méthodes et de concepts peuvent affecter le calcul de la pyramide des âges. Ils sont abordés dans les parties suivantes.

I-A) Une collecte étalée dans le temps

Les résultats définitifs du recensement sont obtenus en synthétisant les informations collectées sur cinq enquêtes annuelles consécutives de façon à disposer d'un effectif total daté du 1er janvier de l'année médiane. En revanche, les informations relatives aux caractéristiques des personnes sont celles disponibles à la date du recensement. En d'autres termes, pour une petite commune collectée en 2004, sa population est extrapolée pour correspondre à une population au 1er janvier 2006 ; elle diffère donc du niveau de population collecté. En revanche, les proportions de femmes, de moins de

[2] Dans la terminologie du recensement : communes de moins de 10 000 habitants.
[3] Pour une présentation détaillée du mode de collecte du nouveau recensement, voir « Pour comprendre le recensement de la population », *Insee Méthodes*, hors série, mai 2005.
[4] En particulier, le cumul n'est pas une moyenne des cinq années de collecte : la pondération de cumul n'est pas égale à un cinquième des pondérations annuelles.

18 ans, de plus de 60 ans sont celles observées en 2004. Il en est de même de la taille moyenne des ménages, de la proportion de personnes en emploi, du taux de chômage… Dans les grandes communes, enquêtées chaque année, les caractéristiques reflètent une situation moyenne sur la période.

Dans ce contexte, construire une pyramide des âges nationale à partir de l'« âge » atteint au recensement revient à comptabiliser ensemble des personnes appartenant à cinq générations successives. Elle diffère de ce fait, d'une pyramide des âges qui serait construite à partir de l'année de naissance. Dans le premier cas, le poids d'une classe d'âges correspond *grosso modo*[5] à la proportion moyenne de cette classe d'âge sur cinq ans, ce qui revient à lisser les accidents de la pyramide des âges (passage à une classe creuse, baby boom, etc.). Dans le second cas, le poids d'une classe d'âges correspond *grosso modo* à la proportion moyenne de la génération concernée sur cinq ans.

Dans la suite, l'accent sera mis sur l'âge et les exemples seront généralement présentés, par souci de simplicité, par genre, mais les calculs présentés doivent être réalisés en réalité sur le croisement sexe-génération ou sexe-âge.

Simulation

Pour illustrer les différences, un exercice fictif simplifié a été réalisé en s'appuyant sur les pyramides des âges déjà publiées dans la *Situation démographique 2006*[6], disponible à la date de réalisation des travaux présentés, avant la disponibilité des résultats définitifs du recensement de 2006. Depuis, les pyramides des âges datées du 1er janvier 2000 et des années suivantes ont été révisées, de même que le calcul des indicateurs démographiques des années 1999 et suivantes. Toutefois, la simulation réalisée et présentée par la suite reste utile et pertinente pour simuler de façon plausible ce qu'auraient pu être les résultats de cinq enquêtes annuelles successives. Pour cela, sont prises en compte les cinq pyramides des âges au 1er janvier entourant 1999, année du dernier recensement. Il est alors possible de construire une pyramide des âges moyenne en cumulant les informations des cinq années successives, l'une par âge, l'autre par génération, puis de les comparer à la pyramide « vraie », celle effectivement publiée pour le 1er janvier 1999. Les pyramides obtenues par cumul le sont en calculant, pour un âge ou une génération donnée, la moyenne simple des effectifs de l'âge ou de la génération considérée sur les cinq années étudiées, puis en recalant les effectifs ainsi obtenus par sexe sur le total publié pour le 1er janvier 1999 (les deux coefficients de recalage, sur le total des hommes et sur le total des femmes, se situent entre 0,9980 et 0,9987).

Par rapport aux estimations faites avec les enquêtes annuelles, l'exercice est simplifié, d'une part parce que l'on fait ici l'hypothèse que les collectes annuelles portent sur la situation au 1er janvier (ce point est discuté dans la partie I-C). D'autre part, parce que l'on fait comme si la méthode d'estimation utilisée par le recensement consistait en une moyenne simple, alors qu'elle est plus complexe (voir I-B). Enfin, parce que le raisonnement est effectué séparément pour les deux genres, afin de mettre l'accent sur la principale difficulté : l'âge.

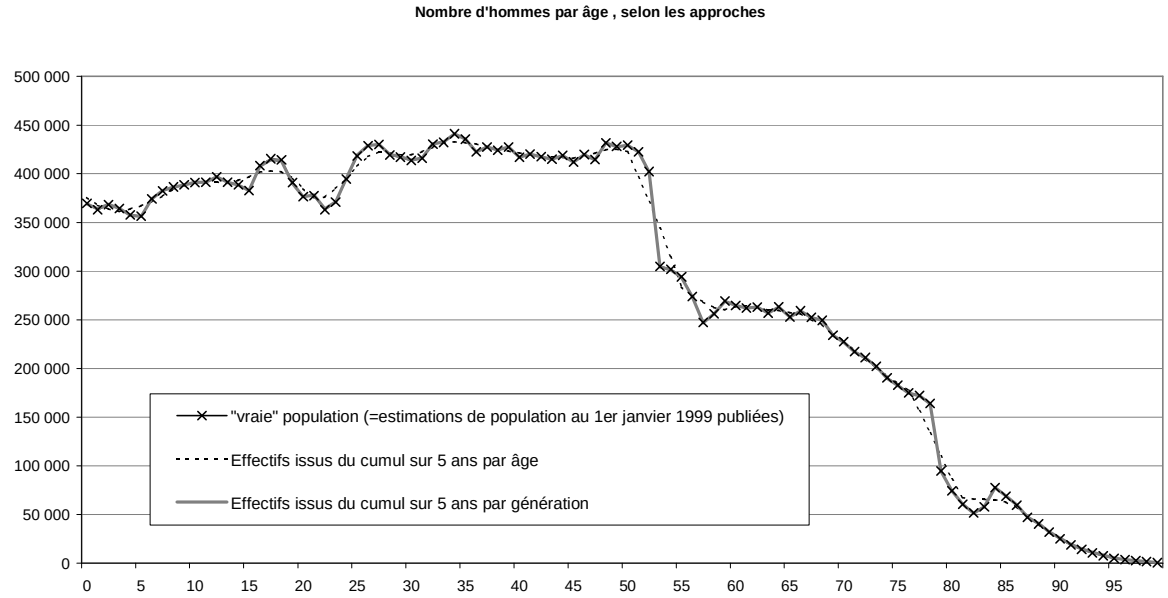
Les graphiques 1 montrent les résultats de la comparaison des niveaux absolus (graphique 1a) et l'écart en niveau relatif (graphique 1b) entre les effectifs par âge des hommes issus d'un cumul « par âge » et ceux issus d'un cumul « par génération », par rapport aux vrais effectifs, connus dans cette simulation.

Ces graphiques illustrent notamment le lissage opéré lorsque les informations sont cumulées sur la variable indiquant l'âge au recensement. En particulier, les déformations de la pyramide liées au passage aux classes creuses de 1914-1918 puis à leur écho sur les générations descendantes de ces classes creuses sont atténuées. Les écarts relatifs d'effectifs par âge peuvent être importants aux âges correspondant à des changements dans l'histoire démographique : ils peuvent dépasser 10%, voire 20% à certains âges.

[5] La partie I-B, sur la déformation des pondérations, explique ce « grosso modo ».

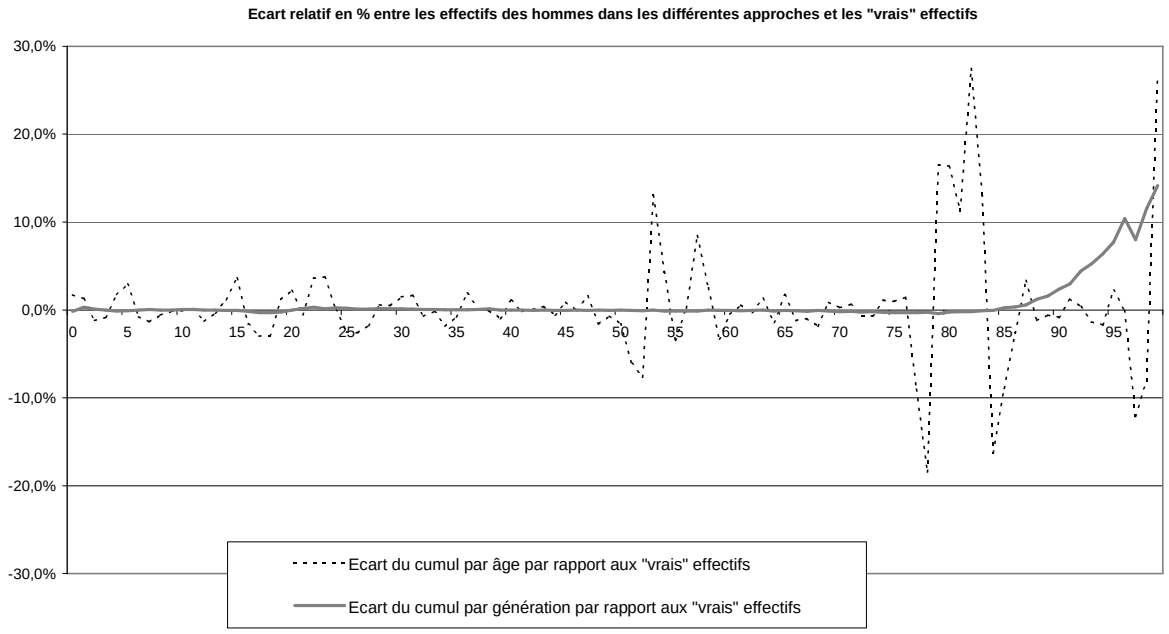
[6] *Insee Résultats – Société* n°84 (août 2008). *La Situation démographique* est une publication annuelle de l'Insee qui diffuse les indicateurs démographiques détaillés et les estimations de population au 1er janvier (population et pyramide des âges).

Graphique 1a



Source : Simulation à partir des pyramides des âges des 1er janvier 1997 à 2001 de « La situation démographique en 2006 », *Insee résultats* n°84, Août 2008 (pyramides des âges au 1er janvier 2000 et 2001 provisoires).
Champ : France métropolitaine.

Graphique 1b



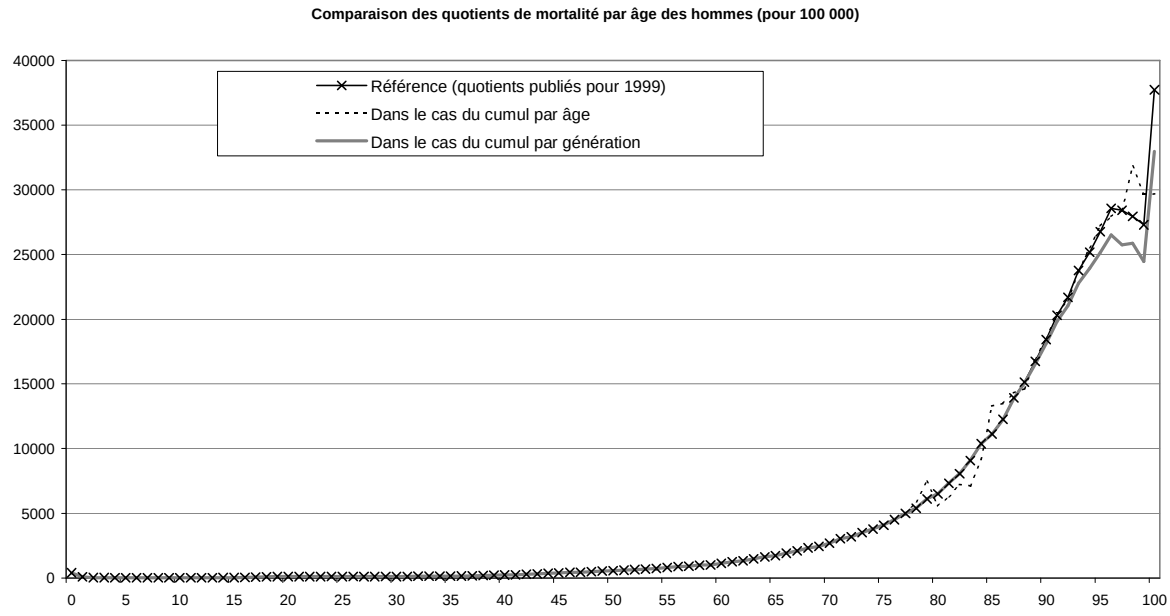
Source : Simulation à partir des pyramides des âges des 1er janvier 1997 à 2001 de « La situation démographique en 2006 », *Insee résultats* n°84, Août 2008 (pyramides des âges au 1er janvier 2000 et 2001 provisoires).
Champ : France métropolitaine.

Cette comparaison montre, en revanche, que le cumul par génération permet d'aboutir à une pyramide reflétant beaucoup mieux les aléas démographiques, sauf aux âges élevés, pour lesquels

elle surestime les effectifs de façon croissante avec l'âge. Le calcul des effectifs par génération en moyenne sur cinq ans est en effet d'autant plus efficace pour approcher les effectifs de la génération l'année médiane que ces effectifs sont relativement stables d'une année sur l'autre, donc pour les âges où les migrations et les décès affectent relativement peu les effectifs. Or, aux âges élevés, le taux de mortalité devient relativement très important, si bien que les décès ne sont plus négligeables dans l'évolution des effectifs d'une année sur l'autre - nous reviendrons plus loin sur la façon de tenir compte de la mortalité.

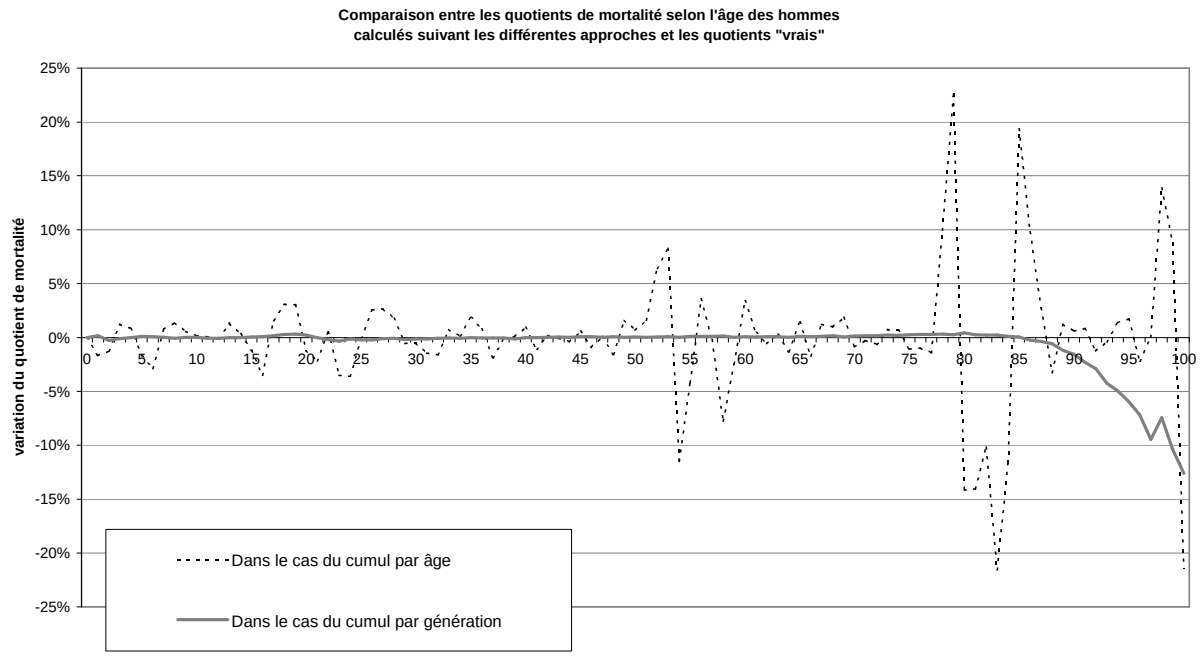
Du fait de l'existence de classes creuses ou de reprise brutale de la natalité, le lissage des effectifs opéré avec le cumul selon l'âge au recensement biaise les estimations des effectifs des différentes générations et perturbe, de ce fait, le calcul des indicateurs démographiques en créant des *artefacts*. Les graphiques suivants montrent ainsi les différences observées pour les quotients de mortalité par âge pour les hommes, en niveau absolu (graphique 2a) et en écart relatif, par rapport aux quotients réels, connus dans la simulation (graphique 2b). L'utilisation du cumul par âge aboutit à une courbe heurtée aux âges pour lesquels l'effet de lissage est important, heurts qui ne reflètent pas la réalité de l'évolution de la mortalité avec l'âge. D'une année sur l'autre, ces artefacts se déplaceraient, en suivant le vieillissement des générations qui marquent les « accidents » démographiques. Pour sa part, le cumul par génération évite ces artefacts mais il conduit à une sous-estimation des risques de décès aux âges élevés, qui croît avec l'âge.

Graphique 2-a



Source : Simulation à partir des pyramides des âges des 1er janvier 1997 à 2001 et des statistiques d'état-civil de « La situation démographique en 2006 », *Insee résultats* n°84, Août 2008 (pyramides des âges au 1 er janvier 2000 et 2001 provisoires).
Champ : France métropolitaine.
.

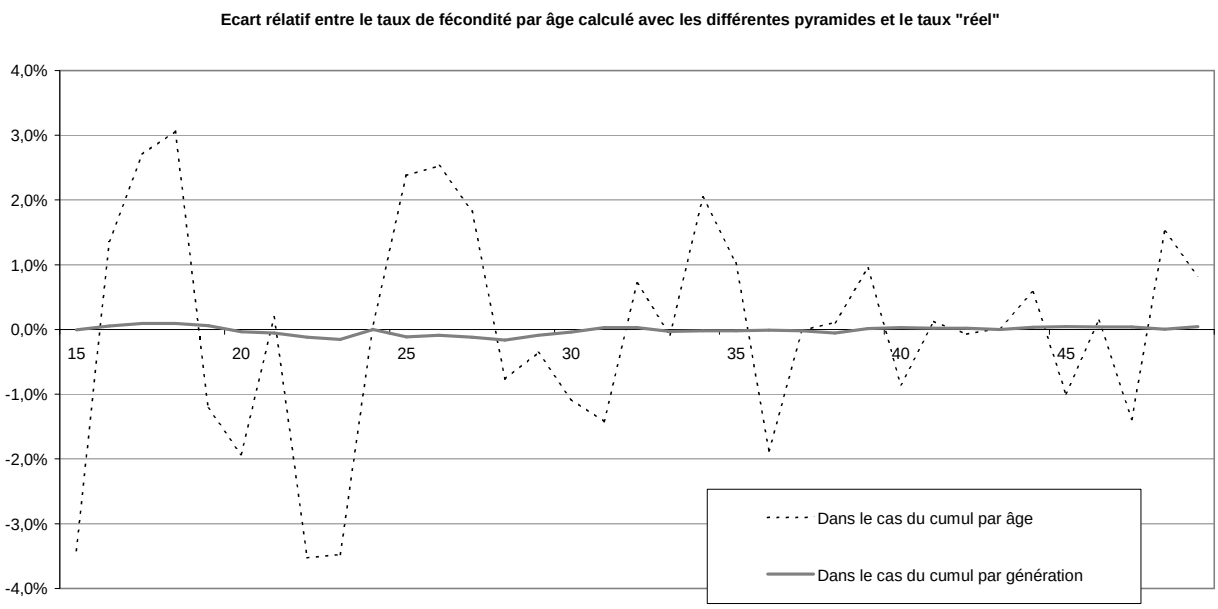
Graphique 2b



Source : Simulation à partir des pyramides des âges des 1er janvier 1997 à 2001 et des statistiques d'état-civil de « La situation démographique en 2006 », *Insee résultats* n°84, Août 2008 (pyramides des âges au 1 er janvier 2000 et 2001 provisoires).
Champ : France métropolitaine.

Le constat fait pour les hommes vaut également pour les femmes. Pour celles-ci, on peut également calculer l'effet du choix de la méthode de cumul sur le taux de fécondité par âge. La comparaison des différentes approches (graphique 3) conduit ici aussi à privilégier un cumul par génération. Compte tenu de la période de procréation des femmes, le calcul du taux de fécondité par âge n'est pas affecté par la surestimation des effectifs aux âges élevés de la pyramide des âges issue du cumul par génération.

Graphique 3



Source : Simulation à partir des pyramides des âges des 1er janvier 1997 à 2001 et des statistiques d'état-civil de « La situation démographique en 2006 », *Insee résultats* n°84, Août 2008 (pyramides des âges au 1 er janvier 2000 et 2001 provisoires).
Champ : France métropolitaine.

Formalisation

Les conséquences d'une collecte étalée dans le temps peuvent être facilement formalisées, sous la double hypothèse simplificatrice déjà énoncée : en plaçant d'emblée les informations de collecte au 1er janvier et en supposant que les résultats obtenus par cumul sur cinq ans résultent d'une moyenne simple. Le raisonnement peut être effectué genre par genre.

Soit *P(g,n)*, la population réelle d'un genre donné, née l'année *g,* présente au 1er janvier de l'année *n* (population d'âge *x*, en âge atteint dans l'année, avec *x=n-g*), population que l'on cherche à estimer (*g<n)* ;
P_r(g,n), la population du genre considéré, née l'année *g* estimée à partir de l'enquête annuelle de recensement de l'année *n* ;
D(*g*,*n*), les décès intervenus l'année *n* de personnes du genre considéré nées l'année *g* ;
N(*g*), les naissances dans l'année *g* (du genre considéré) ;
SM(g,n), le solde migratoire (entrées – sorties) dans l'année *n* de personnes nées l'année *g* (du genre considéré) *;*
e(g,n)= P(g,n)- P_r(g,n), l'écart entre la population du genre considéré née l'année *g* réelle et celle estimée à partir de l'enquête annuelle de recensement de l'année *n*.

L'écart *e(g,n)* résulte des conditions de collecte sur le terrain (facilité d'accès aux logements, intensité et qualité du suivi de collecte, …), mais aussi des conditions de retraitement des informations. Avec la nouvelle méthode de collecte du recensement, il intègre également l'aléa de sondage des petites communes et des immeubles des grandes communes enquêtés dans l'année. Cet aléa n'existait pas lorsque le recensement était exhaustif mais, *a contrario*, l'utilisation d'un répertoire d'immeubles exhaustif dans les grandes communes et la routinisation progressive des opérations de recensement, désormais régulières, permettent de mieux contrôler les conditions de collecte et de retraitement, ce qui peut réduire les écarts qui leur sont imputables.

La population d'une génération *g* donnée présente au 1er janvier *n+1* se déduit de la population de la même génération présente au 1er janvier de l'année *n,* en ajoutant à cette dernière le solde migratoire concernant cette génération pour l'année *n* et en lui retirant les personnes décédées. On retrouve l'équation usuelle, qui sert de base à la comptabilité démographique pour actualiser la dernière pyramide des âges de référence calée sur le recensement et disposer ainsi d'estimations de population plus récentes :

$$P(g,n+1)=P(g,n) - D(g,n) + SM(g,n) \qquad (pour\ g<n)$$

Dans la suite, l'année *n désigne* l'année médiane des cinq années de collecte consécutives, qui s'étalent donc de l'année *n-2* à *n+2*. La pyramide des âges étudiée porte donc sur la situation au 1er janvier de l'année *n*.

Cas de la pyramide des âges obtenue par cumul des informations selon l'âge l'année du recensement

Quand la pyramide des âges est obtenue par le cumul des informations collectées sur cinq ans selon l'âge l'année du recensement, l'effectif d'âge *x=n-g >0*, est estimé dans notre cas simplifié, par *P*(x)* :

$$P^*(x)= 1/5.\left[P_r(g-2,n-2) + P_r(g-1,n-1) + P_r(g,n) + P_r(g+1,n+1) + P_r(g+2,n+2)\right]$$

$$= 1/5.\left[P(g-2,n-2) + P(g-1,n-1) + P(g,n) + P(g+1,n+1) + P(g+2,n+2)\right]$$
$$- 1/5.\left[e(g-2,n-2) + e(g-1,n-1) + e(g,n) + e(g+1,n+1)+ e(g+2,n+2)\right]$$

$$= 1/5.[\ P(g-2,n) + D(g-2,n-2) + D(g-2,n-1) - SM(g-2,n-2) - SM(g-2,n-1)$$
$$+ P(g-1,n) + D(g-1,n-1) - SM(g-1,n-1)$$
$$+ P(g,n)$$
$$+ P(g+1,n) - D(g+1,n) + SM(g+1,n)$$
$$+ P(g+2,n) - D(g+2,n) - D(g+2,n+1) + SM(g+2,n) + SM(g+2,n+1)]$$
$$- 1/5.[\ e(g-2,n-2) + e(g-1,n-1) + e(g,n) + e(g+1,n+1) + e(g+2,n+2)]$$

$$= 1/5.[\ P(g-2,n) + P(g-1,n) + P(g,n) + P(g+1,n) + P(g+2,n)]$$
$$+1/5.[\ D(g-2,n-2) + D(g-2,n-1) + D(g-1,n-1) - D(g+1,n) - D(g+2,n) - D(g+2,n+1)]$$
$$-1/5.[\ SM(g-2,n-2) + SM(g-2,n-1) + SM(g-1,n-1) - SM(g+1,n) - SM(g+2,n) - SM(g+2,n+1)]$$
$$- 1/5.[\ e(g-2,n-2) + e(g-1,n-1) + e(g,n) + e(g+1,n+1) + e(g+2,n+2)]$$

Tant que les décès aux âges considérés sont faibles au regard des effectifs (ou si le nombre de décès est du même ordre de grandeur à âge donné d'une année sur l'autre), le deuxième terme peut être négligé. Pour le moment, les quotients migratoires sont faibles à chaque âge, ce qui rend négligeable le 3e terme. Si les écarts entre les données de collecte et la réalité sont faibles, ils peuvent être également négligés. Dans ces conditions, lorsque l'on cumule selon l'âge les résultats de cinq enquêtes annuelles de recensement successives, l'effectif attribué à un âge donné est proche du premier terme, donc d'une moyenne des effectifs de cinq âges successifs. La pyramide issue d'un cumul par âge correspond donc à un lissage de la vraie pyramide.

Cas de la pyramide des âges obtenue par cumul des informations selon la génération

Quand la pyramide des âges est obtenue par le cumul des informations collectées sur cinq ans selon la génération et pour les générations ayant trois ans ou plus au 1er janvier de l'année médiane, l'effectif à chaque âge $x=n-g>2$, correspond à l'effectif de la génération g, estimée, dans notre cas simplifié, par $P^*(g)$ [7]:

$$P^*(g) = 1/5.[\ P_r(g,n-2) + P_r(g,n-1) + P_r(g,n) + P_r(g,n+1) + P_r(g,n+2)]$$

$$= 1/5.[\ P(g,n-2) + P(g,n-1) + P(g,n) + P(g,n+1) + P(g,n+2)]$$
$$- 1/5.[\ e(g,n-2) + e(g,n-1) + e(g,n) + e(g,n+1) + e(g,n+2)]$$

$$= 1/5.[\ P(g,n) + D(g,n-2) + D(g,n-1) - SM(g,n-2) - SM(g,n-1)$$
$$+ P(g,n) + D(g,n-1) - SM(g,n-1)$$
$$+ P(g,n)$$
$$+ P(g,n) - D(g,n) + SM(g,n)$$
$$+ P(g,n) - D(g,n) - D(g,n+1) + SM(g,n) + SM(g,n+1)]$$
$$- 1/5.[\ e(g,n-2) + e(g,n-1) + e(g,n) + e(g,n+1) + e(g,n+2)]$$

$$= P(g,n)$$
$$+ 1/5.[\ D(g,n-2) + 2.D(g,n-1) - 2.D(g,n) - D(g,n+1)]$$
$$- 1/5.[\ SM(g,n-2) + 2.SM(g,n-1) - 2.SM(g,n) - SM(g,n+1)]$$
$$- 1/5.[\ e(g,n-2) + e(g,n-1) + e(g,n) + e(g,n+1) + e(g,n+2)]$$

[7] Dans le cas du cumul par génération, la relation $g<n$ n'est pas toujours vérifiée. Un effet de bord joue pour les générations les plus jeunes, nées au cours du cycle des cinq années de collecte. Ce point est abordé plus bas, donc les relations présentées ne font intervenir que la composante « décès » du solde naturel, comme dans le cas du cumul selon l'âge.

Sous les mêmes conditions que précédemment, lorsque l'on cumule selon la génération, c'est-à-dire selon l'année de naissance, les résultats de cinq enquêtes annuelles de recensement successives l'effectif attribué à une génération donnée correspond aux effectifs exacts de la génération.

Pour les générations ayant un ou deux ans au 1er janvier de l'année médiane, l'estimation des effectifs se heurte à un effet de bord :

- pour la génération née l'année *g=n-2*, l'effectif recensé l'année *n-2*, est nul[8]. Par contre, cette génération est recensée les quatre années suivantes. L'estimation de cette génération par le cumul des informations collectées sur cinq ans sous-estime les effectifs réels d'environ un facteur 1/5, dans la mesure où la mortalité et les migrations aux bas âges sont négligeables,.

- pour la génération née l'année *g=n-1*, l'effectif recensé l'année *n-2* et l'année *n-1*, est nul. Par contre, cette génération est recensée les trois années suivantes. Dans le même contexte que précédemment, l'estimation de cette génération par le cumul des informations collectées sur cinq ans sous estime les effectifs réels d'environ un facteur 2/5.

Dans cette approche, ces générations nécessitent donc une méthode d'estimation spécifique. Les générations nées les années *n, n+1* et *n+2*, sont également recensées partiellement au cours du cycle des cinq enquêtes annuelles de recensement mais, nées postérieurement à la date de référence, elles n'entrent pas en considération.

Ampleur des biais selon l'âge

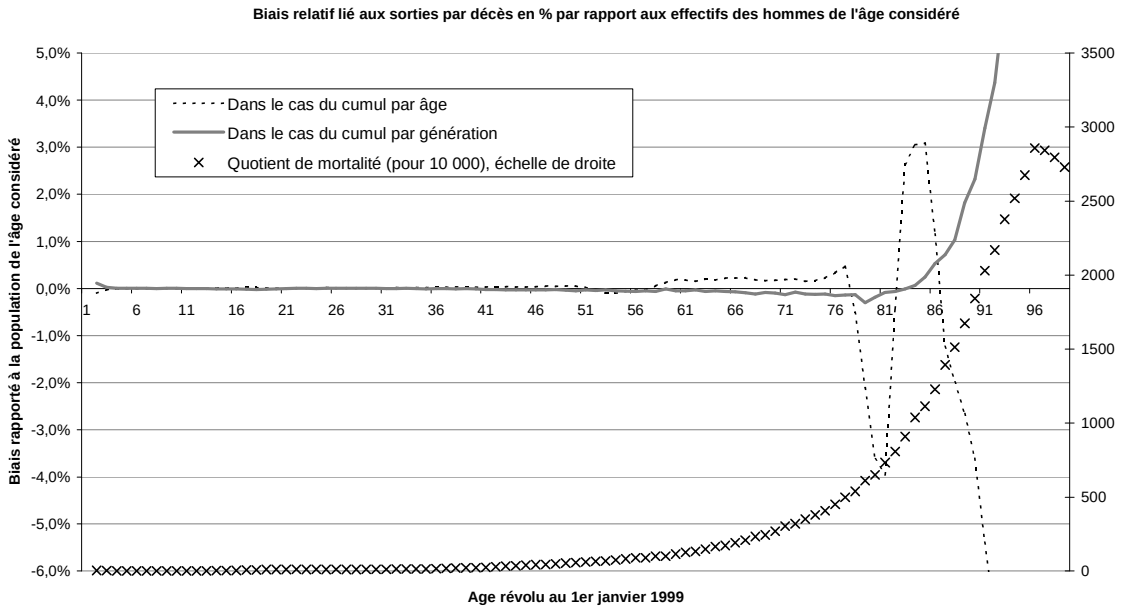
La formalisation ci-dessus montre qu'en cumulant les informations en moyenne sur cinq ans, selon l'âge ou selon la génération, on obtient une approximation, respectivement des effectifs lissés par âge ou des effectifs exacts par âge, avec deux types d'écart possibles : l'un lié à la non-prise en compte des flux affectant la population considérée d'une année sur l'autre, l'autre lié aux aléas de sondage, de collecte et de traitements du recensement. Dans le cas présent des estimations de population, le premier facteur se décompose en deux : le mouvement naturel et le mouvement migratoire.

L'ampleur des biais peut être évaluée dans le cadre de la simulation déjà présentée, sur la période 1997-2001. Le graphique 4 montre ainsi l'ampleur du biais lié à la non-prise en compte des décès par âge - ce biais est rapporté à la population réelle de l'âge considéré. Il montre que ce biais est négligeable jusqu'aux alentours de 80 ans dans les deux approches. Il progresse ensuite aux âges élevés, lorsque les quotients de mortalité deviennent élevés.

Le graphique 5 montre l'ampleur du biais lié à la non-prise en compte du solde migratoire par âge - ce biais est rapporté à la population réelle de l'âge considéré. Il montre que ce biais est négligeable à tout âge. Il s'élève en valeur absolu aux très grands âges mais, d'une part, il reste très faible au regard du biais lié à la non-prise en compte des décès ; d'autre part, il peut s'agir d'un *artefact*, car si les décès par âge sont bien connus grâce aux statistiques d'état civil, les estimations de solde migratoire par âge reposent sur des informations limitées et sont donc moins robustes. L'erreur de mesure sur le solde migratoire intervient aussi, sans qu'elle soit connue.

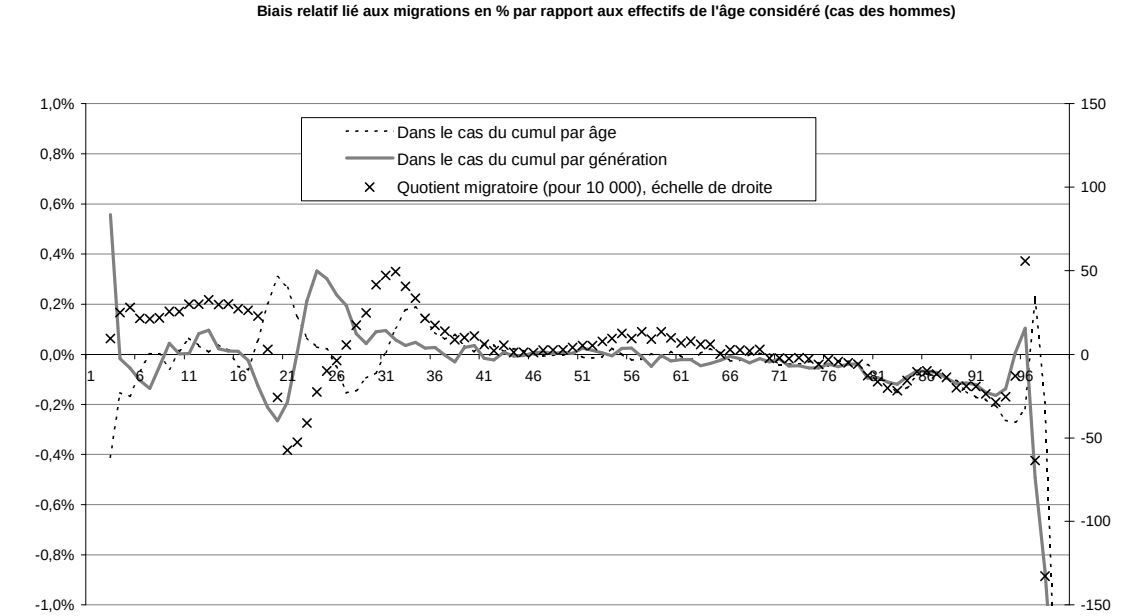
[8] Dans la réalité, avec les enquêtes annuelles de recensement, il est presque nul en fait, les personnes nées entre le 1er janvier et la date de référence de l'enquête annuelle, aux alentours de mi janvier, étant comptabilisées.

Graphique 4



Source : Simulation à partir des pyramides des âges des 1er janvier 1997 à 2001 et des statistiques d'état-civil de « La situation démographique en 2006 », *Insee résultats* n°84, Août 2008 (pyramides des âges au 1er janvier 2000 et 2001 provisoires).
Champ : France métropolitaine.
Lecture : A 80 ans, le quotient de mortalité des hommes en 1999, c'est-à-dire leur risque de décès à cet âge, s'établit à environ 650 pour 10 000. Ne pas tenir compte du biais lié au décès pour construire une pyramide des âges issue d'un cumul selon l'âge au recensement, conduit à sous-estimer les effectifs d'hommes de cet âge de 3,6%. La sous-estimation est de 0,2% dans le cas de la pyramide des âges issue d'un cumul selon la génération.

Graphique 5



Source : Simulation à partir des pyramides des âges des 1er janvier 1997 à 2001 et des statistiques d'état-civil de « La situation démographique en 2006 », *Insee résultats* n°84, Août 2008 (pyramides des âges au 1er janvier 2000 et 2001 provisoires).
Champ : France métropolitaine.

I-B) Des populations communales ramenées à la même date, le 1er janvier de l'année médiane (passage de pondérations annuelles à une pondération de cumul)

La simulation présentée dans la partie I-A faisait l'hypothèse que le cumul des résultats des cinq enquêtes annuelles de recensement successives était obtenu par simple moyenne des informations collectées chaque année, moyenne calée ensuite sur la population totale de référence, le 1er janvier 1999 dans la simulation. Dans le cas du recensement, la méthode de cumul est plus complexe de façon à fournir une population légale pour chaque commune relative à la même date, le 1er janvier de l'année médiane des cinq collectes annuelles successives. La « pondération du cumul », utilisée pour publier les résultats du recensement, ne peut pas se déduire des pondérations annuelles de chaque enquête annuelle (permettant d'extrapoler au niveau national un résultat annuel : on parle de « pondération annuelle »). Comme nous allons le voir, **le passage des pondérations annuelles à la pondération de cumul déforme légèrement la structure estimée de la population ; c'est pourquoi *in fine*, il sera préférable, pour calculer la pyramide des âges de référence pour le calcul des indicateurs démographiques, d'utiliser les résultats des enquêtes annuelles estimés avec les pondérations annuelles, plutôt que les résultats estimés avec la pondération de cumul.**

La méthode d'estimation de la population utilisée dans le recensement

Population des ménages dans les communes de moins de 10 000 habitants

Pour les communes recensées l'une des deux années précédant l'année médiane[9], une extrapolation est réalisée, en combinant une estimation du taux d'évolution du parc de logement et une estimation d'évolution de la taille moyenne des ménages : en raison d'une tendance structurelle à la réduction de la taille moyenne des ménages, la population progresse moins vite que le nombre de logements.

De façon plus précise, à partir des deux derniers recensements disponibles[10], on calcule le ratio égal au rythme annuel d'évolution de la population des ménages rapporté au rythme annuel d'évolution du nombre de résidences principales. Ce coefficient est ensuite supposé stable sur la période d'extrapolation (qui n'excède pas deux ans). Les fichiers de la taxe d'habitation permettent de calculer l'évolution du parc de logements sur cette période. L'évolution de la population sur la période d'extrapolation est ensuite obtenue en appliquant à l'évolution issue de la taxe d'habitation, le coefficient correcteur calculé à partir des deux derniers recensements. Cette première étape conduit à attribuer une même pondération p1 à tous les logements d'une petite commune donnée, p1 variant d'une commune à l'autre.

Pour les communes recensées l'année médiane (2006 pour les premiers résultats disponibles), la population est celle recensée dans l'enquête annuelle. La pondération du cumul des logements recensés et des individus est donc égale à 1.

Pour les communes recensées l'une des deux années suivant l'année médiane[11], une interpolation simple est réalisée entre le dernier chiffre collecté, postérieure à la date de référence, et le dernier résultat disponible du recensement, par définition antérieure à cette date. Pour 2006, il s'agissait des résultats du recensement de 1999. Pour les résultats du recensement de 2007, ce sont les résultats de 2006. En régime permanent, l'interpolation sera donc au plus de deux ans. L'interpolation conduit à attribuer une même pondération p2 à tous les logements d'une petite commune donnée, p2 variant d'une commune à l'autre.

[9] 2004 et 2005 pour le premier cycle de cinq ans de la nouvelle méthode de recensement.

[10] Recensement de 1999 et enquête annuelle de 2004 ou 2005 pour les communes enquêtées respectivement en 2004 et 2005 pour le premier exercice. Pour les résultats du recensement de 2011, l'extrapolation s'appuiera sur les résultats des enquêtes annuelles de 2004 et 2009 pour les communes enquêtées en 2009 et sur les résultats des enquêtes annuelles de 2005 et 2010 pour les communes enquêtées en 2010.

[11] 2007 et 2008 pour le premier cycle de cinq ans de la nouvelle méthode de recensement.

Dans les petites communes extrapolées[12], la population est rapportée au 1er janvier de l'année médiane, mais aussi le nombre de logements. Or généralement les deux n'évoluent pas au même rythme. Pour qu'un logement et les individus qui y habitent aient la même pondération de cumul, un calage sur marge est réalisé, avec le nombre de logements et la population des ménages comme marge. Ce calage conduit à corriger la pondération p1 d'un facteur p3, variable d'un logement à l'autre dans une même commune. La pondération de cumul est alors p1xp3. Pour des raisons techniques, ce calage n'est effectué que dans les petites communes de plus de 2000 habitants. Il n'a pas lieu d'être dans les communes recensées l'année médiane. Le calage n'est pas non plus effectué dans les communes recensées après l'année médiane, parce qu'il n'y a pas d'informations externes sur l'évolution du nombre de logements. De ce fait, le nombre de logements est supposé évoluer comme le nombre d'habitants.

Population des ménages dans les communes de plus de 10 000 habitants

Pour les grandes communes, l'estimation de population des logements s'appuie sur un répertoire d'immeubles localisés daté du 1er janvier de l'année médiane, donc sur une base exhaustive des logements à cette date[13].

Chaque année, la multiplication de l'effectif de la population enquêtée par le poids de sondage donne une estimation de la population du groupe d'adresses[14]. La somme des cinq groupes de rotation donne une population moyenne des ménages et un nombre moyen de logements pour la période, donc une taille moyenne des ménages sur la période. Cette taille moyenne des ménages est ensuite appliquée au nombre de logements du répertoire d'immeubles localisés daté du 1er janvier de l'année médiane.

Pour le calcul de la pondération de cumul, le calage sur le répertoire d'immeubles localisés ne s'effectue pas directement au niveau de la commune mais à un niveau infra-communal, souvent l'IRIS[15]. Ce calage modifie les pondérations annuelles associées aux individus recensés.

Population « hors ménages »

Pour les populations des communautés (maisons de retraite, casernes, prisons, etc.), le principe est le même que celui retenu pour les populations des ménages en petites communes : la population des communautés recensées après l'année médiane est interpolée ; la population des communautés recensées avant l'année médiane est extrapolée. L'interpolation se fait en répartissant, à parts égales sur chaque année, l'écart entre le dernier chiffre publié et celui fourni par le recensement de la communauté. L'extrapolation est faite à partir de sources administratives externes. En s'appuyant sur le répertoire des communautés, les communautés disparues entre leur recensement et l'année médiane sont supprimées ; les capacités des communautés créées sont ajoutées.

Pour les populations des habitations mobiles, les personnes sans abris et les mariniers, les effectifs sont maintenus constants entre deux opérations de recensements de ces catégories, donc sur cinq ans.

[12] Donc celles recensées en 2004 et 2005, pour le premier cycle de cinq ans de la nouvelle méthode de recensement.

[13] Le répertoire d'immeubles localisés étant actualisé au 1er juillet, le répertoire au 1er janvier est obtenu en faisant la moyenne des deux versions successives du répertoire qui l'encadrent. En d'autres termes, le nombre de logements d'une adresse au 1er janvier *N* est estimé par la moyenne du nombre de logements à cette adresse au 1er juillet *N-1* et au 1er juillet *N*.

[14] Compte-tenu des objectifs du recensement, le traitement de la non-réponse, ou « logements non enquêtés », est intégré au processus de recensement en amont des opérations d'actualisation et de calage. Des bulletins fictifs sont ainsi intégrés par duplication de bulletins réels. Ce type de traitements existait déjà avec les recensements généraux de population et concerne toujours les communes de moins de 10 000 habitants. Pour les communes de plus de 10 000 habitants, la disponibilité d'un répertoire d'immeubles permet de mieux gérer la non-réponse.

[15] Zonage infra communal constitué par l'Insee pour diffuser des résultats infra communaux du recensement, constitué selon des critères géographiques et démographiques, notamment pour assurer une homogénéité de l'habitat et une stabilité des limites dans le temps.

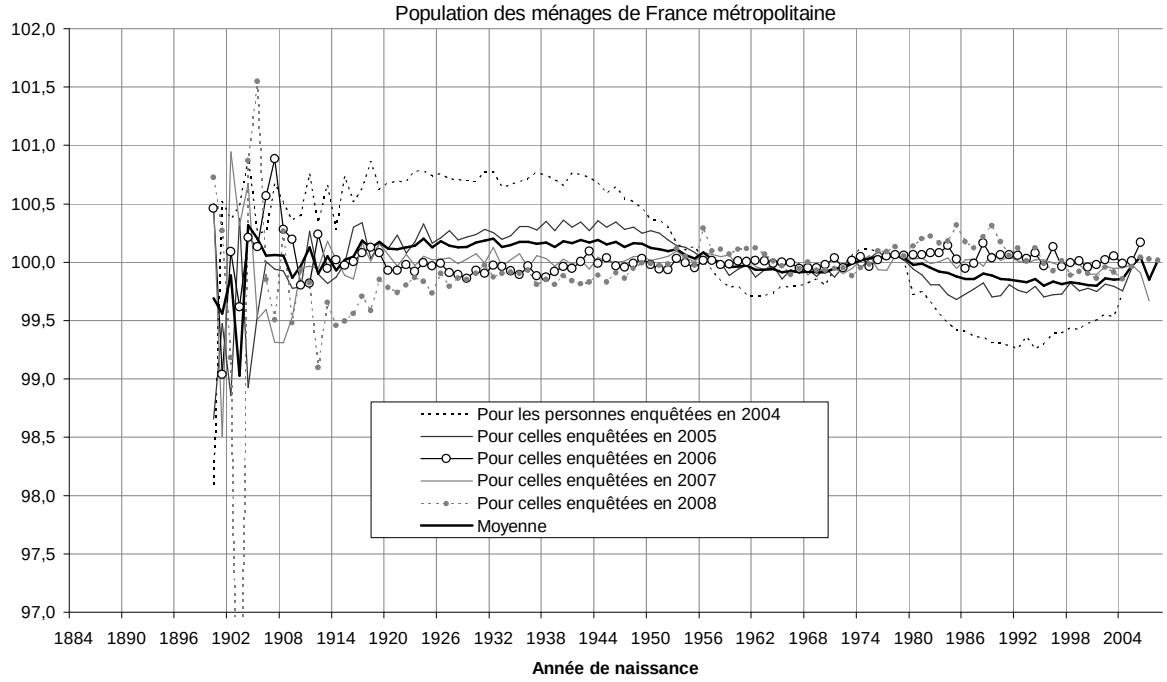
Conséquences sur la mesure des caractéristiques de la population recensée

Le cas des petites communes, collectées de façon exhaustive une seule année sur les cinq années utilisées, suffit à illustrer les conséquences de la méthode de calcul de la pondération de cumul sur la mesure des caractéristiques de la population recensée. Pour les petites communes recensées l'année médiane ou l'une des deux années suivantes, la repondération modifie le niveau de population totale de la commune concernée pour la ramener à la bonne date mais elle ne modifie pas sa structure, notamment pas sa pyramide des âges. Comme le coefficient de redressement appliqué au niveau de population recensée n'est pas identique d'une petite commune à l'autre, la structure de la population de tout agrégat de communes se retrouve déformée par rapport à celle qui résulte directement des collectes annuelles, même dans le cas fictif où l'on regarderait la structure par âge de l'ensemble des petites communes recensées une même année donnée. Pour les petites communes recensées l'une des deux années précédent l'année médiane, la correction p3 effectuée lors de l'extrapolation, parce qu'elle est variable d'un logement à l'autre, déforme la structure de la population par rapport à celle recensée. De même, au sein d'une grande commune, la structure de la population mesurée avec la pondération de cumul n'est pas exactement la structure moyenne sur les cinq années de collecte[16].

Le graphique 6 illustre les écarts de la structure par âge et sexe selon que l'on utilise les résultats d'une enquête annuelle donnée avec la pondération « annuelle », ou les résultats du cumul 2004-2008 sur le champ des seules personnes recensées cette année là pour la France métropolitaine. L'exercice est ici limité aux personnes vivant dans des logements ordinaires, donc la population des ménages.

Graphique 6

Déformation de la structure liée au passage de la pondération annuelle à la pondération de cumul
(Poids de la génération donnée avec la pondération de cumul / poids avec la pondération annuelle, en indice)
Population des ménages de France métropolitaine



Source : Insee, Enquêtes annuelles de recensement 2004 à 2008.
Champ : Population des ménages, France métropolitaine.
Lecture : Concernant les personnes enquêtées en 2004 (courbe en pointillés noirs), la courbe prend des valeurs supérieur à 100 pour les personnes nées avant le milieu des années cinquante. La proportion de chacune des générations concernées dans la population totale estimée avec la pondération de cumul sur le champ de personnes enquêtées en 2004 est donc supérieure à celle estimée avec la pondération annuelle de 2004. C'est le contraire pour les générations nées après 1980. Pour les personnes enquêtées en 2006, la courbe n'est pas constante à 100 en raison de la méthode de calcul des pondérations dans les grandes communes.

[16] Ce constat reste théorique car au niveau d'une grande commune, la marge d'imprécision liée à l'échantillonnage des logements enquêtés chaque année rend l'utilisation des résultats annuels beaucoup moins robuste que l'utilisation directe du cumul.

Lorsque l'on se restreint au champ des personnes enquêtées une année donnée, le poids d'une génération d'un sexe donné dans la population, donc ses effectifs, peut être très différent selon qu'il est obtenu avec la pondération de cumul (celle des populations légales) ou avec la pondération annuelle pour les âges les plus avancés, mais les effectifs concernés (donc la part dans la population totale) sont alors très faibles. En dehors des âges élevés, l'écart du poids, donc la déformation de structure sur une année de collecte, est pour de nombreux âges de l'ordre de 0,4 à 0,8% en valeur absolue, soit davantage que le biais lié à la non-prise en compte des décès ou celui lié à la non prise en compte des flux migratoires. Pour disposer des données de cadrage dont ont besoin la plupart des utilisateurs, ces écarts restent toutefois minimes. La répartition entre hommes et femmes est elle aussi concernée mais de façon encore plus négligeable.

	Champ des personnes enquêtées en :				
Part des femmes	2004	2005	2006	2007	2008
Avec la pondération annuelle	51,70%	51,66%	51,62%	51,65%	51,62%
Avec la pondération de cumul	51,71%	51,66%	51,61%	51,65%	51,61%

Champ : Population des ménages de France métropolitaine.

On peut essayer de formaliser un peu plus précisément cela dans le cas de la pyramide des âges obtenue en cumulant les informations sur cinq ans selon la génération (en mettant de côté le genre).

Considérons le cycle de cinq ans centré sur l'année *n.* Pour simplifier certaines formules, on appelle *i* la i-ème année des cinq années utilisées dans le cumul sur cinq ans. En d'autres termes, pour l'année *n-2, i=1* ; pour l'année *n-1, i=2* ; pour l'année *n, i=3* ; pour l'année *n+1, i=4* et pour l'année *n+2, i=5*.

Soit : - $P_r(i)$, la population totale à l'enquête annuelle collectée l'année *i* (il s'agit ici de l'effectif total obtenu en utilisant le poids annuel).
- $P_r(g,i)$, la population totale de la génération *g* à l'enquête annuelle collectée l'année *i* (poids annuel)
- $P^*(i)$, les effectifs enquêtés la i-ème année des cinq années utilisées dans le cumul sur cinq ans (il s'agit ici de l'effectif obtenu en pondérant les personnes avec le poids de cumul,).
- $P^*(g,i)$, les effectifs de la génération *g* enquêtés la i-ème année des cinq années utilisées dans le cumul sur cinq ans (poids de cumul).

Soit $\alpha(g,i)$, le poids de la génération *g* dans la population totale, dans l'enquête annuelle *i* (poids annuel). $\alpha(g,i)= P_r(g,i)/ P_r(i).$

Soit $\alpha^*(g,i)$, le poids de la génération *g* dans l'ensemble de la population estimée à partir des personnes enquêtées l'année *i* avec le poids de cumul. $\alpha^*(g,i)= P^*(g,i)/ P^*(i)$

On appelle $k(g,i)= \alpha^*(g,i)/ \alpha(g,i)$: c'est donc un coefficient de déformation de la structure par âge induit par le passage de la pondération annuelle à la pondération de cumul pour les personnes recensées l'année *i* (dans le graphique 6, les courbes de parts relatives correspondent à des courbes de $k(g,i)$-1 en pourcentage).

On a donc *:* $P^*(g,i)= k(g,i).(P^*(i)/P_r(i)).P_r(g,i)$.

Dans le cumul, l'effectif de la génération *g* collecté l'année *i* est donc corrigé par un coefficient de déformation de la structure par âge - $k(g,i)$ - et du poids de l'enquête annuelle *i* dans le cumul - $P^*(i)/P_r(i)$, qui serait exactement de 1/5 si la population était stationnaire et l'échantillon des groupes de rotation parfaitement équilibré.

Pour simplifier l'écriture dans l'expression suivante, on appelle : $\alpha_1(g,n)= k(g,n-2).(P^*(n-2)/P_r(n-2))$, la valeur du produit de ces deux coefficients pour la génération g, collectée en *n-2*, dans le cumul centré sur l'année *n* (*n-2* correspond donc à la 1ère année du cumul). De même, on définit $\alpha_2(g,n)$, $\alpha_3(g,n)$, $\alpha_4(g,n)$, $\alpha_5(g,n)$.

Les effectifs de la génération *g* dans le cumul sur cinq ans sont donc :

$P^*(g) = P^*(g,n-2)+ P^*(g,n-1)+ P^*(g,n)+ P^*(g,n+1)+ P^*(g,n+2)$

$= (\alpha_1(g,n) + \alpha_2(g,n) + \alpha_3(g,n) + \alpha_4(g,n) + \alpha_5(g,n)).\ P(g,n)$
$+ \alpha_1(g,n).D(g,n-2) +(\alpha_1(g,n)+ \alpha_2(g,n)).D(g,n-1) -(\alpha_4(g,n)+\alpha_5(g,n)).D(g,n)- \alpha_5(g,n).D(g,n+1)$
$- \alpha_1(g,n).SM(g,n-2) + (\alpha_1(g,n)+ \alpha_2(g,n)).SM(g,n-1) - (\alpha_4(g,n)+\alpha_5(g,n)).SM(g,n) -\alpha_5(g,n).SM(g,n+1)$
$- \alpha_1(g,n).e(g,n-2) + \alpha_2(g,n).e(g,n-1) + \alpha_3(g,n).e(g,n) + \alpha_4(g,n).e(g,n+1)+ \alpha_5(g,n).e(g,n+2)$

Lorsque la déformation de la structure par âge induite par les repondérations réalisées pour le calcul des populations légales est limitée et s'il y a une répartition relativement équilibrée des effectifs pondérés entre les différentes années utilisées dans le cumul, les $\alpha_i(g,n)$ sont proches d'1/5. On retrouve alors le cas simplifié présenté en I-A.

Prendre en compte la déformation de la structure par âge peut être utile parce que d'une année sur l'autre, la déformation appliquée aux personnes recensées la même année et appartenant à la même génération ne sera pas identique. De ce fait, la comparaison de deux pyramides des âges successives issues du cumul de cinq enquêtes annuelles successives, sans prise en compte de cette déformation perturbe le profil du solde migratoire apparent par âge (et genre). Le recours aux pondérations annuelles plutôt qu'à la pondération de cumul est un moyen de faire l'économie de cette difficulté pour le calcul de la pyramide des âges de référence pour le calcul des indicateurs démographiques.

I-C) Les changements de concepts

L'âge révolu à la date du recensement

Pour l'établissement de la pyramide des âges de référence et pour les comparaisons entre les différents recensements, il est essentiel de préciser la façon de définir l'âge des personnes. La date de naissance figurant dans les informations demandées dans le bulletin individuel du recensement, deux âges sont disponibles :

- l'âge atteint dans l'année du recensement, obtenu en faisant la différence de millésime entre l'année de l'enquête de recensement et l'année de naissance (c'est-à-dire également l'âge révolu au 31 décembre de l'année de l'enquête) ;

- l'âge en années révolues à la date du recensement, c'est-à-dire, l'âge atteint au dernier anniversaire, avant l'enquête de recensement.

Dans les résultats diffusés de façon standard à partir du recensement, l'âge atteint était privilégié dans le recensement de 1999. Avec les enquêtes annuelles, c'est plutôt l'âge révolu à la date du recensement.

Pour les calculs d'une pyramide des âges par génération, c'est l'année de naissance qui est prise en compte. Par souci d'homogénéité dans les travaux présentés dans ce document, notamment l'exercice de simulation, la pyramide des âges issue d'un cumul par âge est également construite selon l'âge atteint dans l'année, approximation de la pyramide des âges selon l'âge au recensement (cette approximation ne modifie pas les analyses et permet de les centrer sur les difficultés premières).

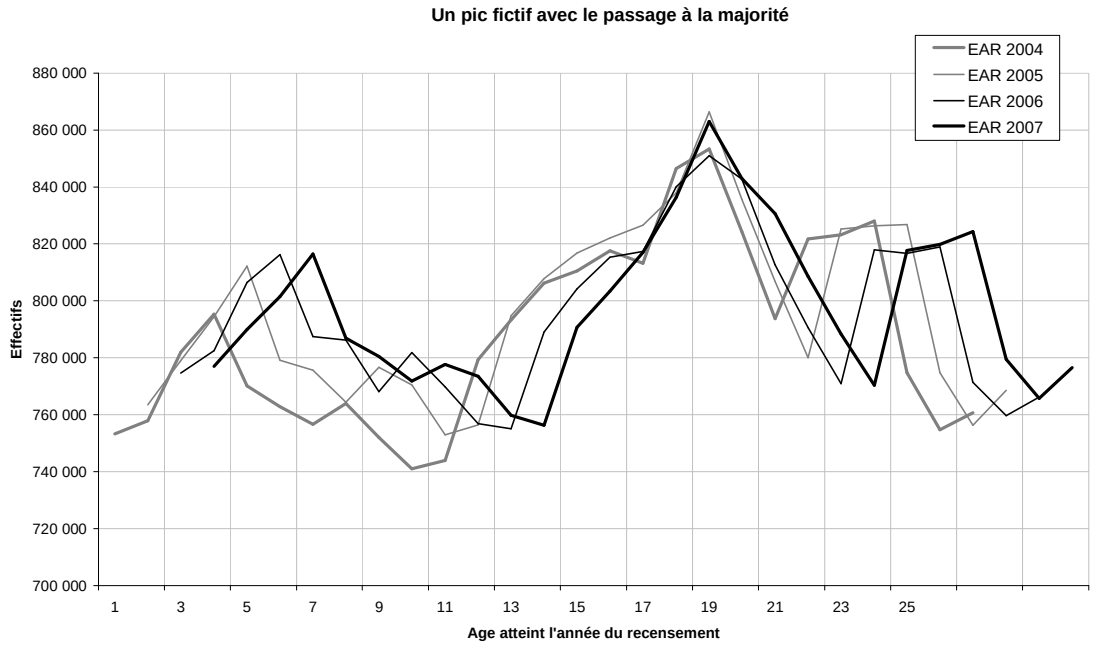
Comptabilisation des élèves internes majeurs

Jusqu'en 1999, les élèves internes majeurs (au nombre de 165 000 cette année-là) étaient comptabilisés chez leurs parents, donc parmi les personnes vivant dans les ménages, alors que les personnes majeures des établissements militaires (au nombre de 35 500) l'étaient dans leur résidence personnelle. Ils sont tous désormais comptabilisés dans la population des communautés,

donc hors ménages. *A contrario*, les mineurs sont désormais tous comptés chez leurs parents lorsqu'ils logent en cité universitaire. Ils étaient 3 500 en 1999.

Ces conventions créent un *artefact* dans les comparaisons avec le recensement de 1999 : une forte hausse de la proportion de jeunes vivant hors ménage à partir de 18 ans. Par ailleurs, ce changement de comptabilisation semble générer des doubles comptes pour une partie de ces élèves internes.

Graphique 7



Source : Insee, Enquêtes annuelles de recensement 2004 à 2008.
Champ : France métropolitaine.

De fait, les résultats des enquêtes annuelles montrent l'existence d'un pic anormal à 19 ans en âge atteint dans l'année du recensement (donc pour l'essentiel, des personnes de 18 ans en âge révolu au moment du recensement). Le graphique 7, *supra*, montre les effectifs par âge atteint dans l'année dans chacune des enquêtes annuelles 2004 à 2007. Alors que les courbes, avec leurs pics et leurs creux, se déplacent d'un an à chaque enquête successive, traduisant le vieillissement des générations, peu perturbé par les migrations et les décès, un pic n'est pas marqué par ce glissement annuel, à l'âge de 19 ans.

En pratique, les ménages recensées remplissent d'une part une feuille de logement, où ils listent les « habitants permanents » du logement (« liste A »), les « enfants majeurs logés ailleurs pour leurs études » (« liste B ») et les « autres habitants du logement » (« liste C »). Les bulletins individuels ne doivent être remplis que pour les « habitants permanents ». Toutefois, certains parents peuvent considérer leur enfant majeur interne dans un établissement comme un habitant permanent, surtout s'il revient régulièrement au domicile familial. Il y a alors un risque de double compte de l'enfant, comptabilisé à la fois chez ses parents, en ménage ordinaire, et, hors ménage, là où il est interne.

Le décalage de date de référence entre les ménages et les communautés peut aussi induire un double compte puisqu'un jeune interne qui atteint 18 ans entre les deux dates de référence (*grosso modo* entre mi janvier et début mars) peut être à raison déclaré chez ses parents lors de la collecte auprès des ménages et en communauté lors de la collecte des communautés. La période de

collecte des communautés va être alignée sur celle des ménages à partir de l'enquête annuelle de recensement de 2010, ce qui supprimera tout risque de ce point de vue.

I-D) La datation des informations

Lors des recensements généraux de population, les informations se rapportaient à la date de référence du recensement, soit le 8 mars 1999, pour le dernier d'entre eux. Les pyramides des âges diffusées se rapportant à la situation au 1er janvier, la pyramide des âges de référence déduite du recensement était obtenue en ramenant au 1er janvier la pyramide issue du recensement. Les personnes nées l'année du recensement (personnes d'âge zéro) étaient exclues. *A contrario*, les personnes décédées entre le 1er janvier et la date du recensement, n'étant pas recensées alors qu'elles étaient vivantes au 1er janvier, étaient réintégrées dans les effectifs par sexe et âge. Les estimations annuelles du solde migratoire par sexe et âge étaient également utilisées pour ré imputer les flux migratoires par sexe et âge survenus entre le 1er janvier et la date du recensement.

Avec le nouveau recensement, la date de référence tourne autour de mi janvier, pour l'essentiel. :

- pour les ménages (y compris les personnes résidant dans les habitations mobiles et les personnes sans abri) résidant hors du département de La Réunion, la date de référence correspond au troisième jeudi du mois de janvier ;

- pour les ménages (y compris les personnes résidant dans les habitations mobiles et les sans abri) du département de La Réunion, la date de référence correspond au cinquième jeudi de l'année ;

- pour les communautés hors du département de La Réunion, la date de référence est le 1er mars ;

- pour les communautés du département de La Réunion, la date de référence est le troisième lundi de l'année en 2004 et le quatrième lundi à partir de 2005.

Seules les personnes nées avant la date de référence sont prises en compte. Pour mémoire, pour 1999, la date de référence utilisée est le 8 mars 1999.

Comme pour les recensements généraux de population, la collecte décrit une situation moyenne postérieure au 1er janvier, surtout si l'on considère que les ménages tendent à décrire davantage la situation au moment où ils remplissent les bulletins qu'à la date de référence. Comme la collecte s'étale de mi janvier à fin février, les informations collectées donneraient donc plutôt une situation moyenne au début du mois de février, aux naissances postérieures à la date de référence près (ces naissances sont supprimées lors des contrôles). Les personnes décédées entre le 1er janvier et la date à laquelle elles auraient été enquêtées manquent alors, de même que les personnes émigrantes sur la même période. *A contrario*, les immigrants sur la période seraient à exclure. Un chiffrage précis n'est pas facile : manque d'information sur les flux migratoires, nécessité de prise en compte des dates de collecte différentes des ménages et des établissements, ceux-ci n'accueillant pas des populations avec le même risque de mortalité, à âge donné. L'ordre de grandeur se situerait autour de 30 000 à 60 000, beaucoup moins, naturellement, que lorsque le recensement avait lieu en mars.

Ce léger écart joue sur la structure par âge, puisque le risque de décès et les quotients migratoires varient selon le sexe et l'âge. Il peut jouer également sur les niveaux globaux de population mais de façon complexe. En effet, pour les grandes villes, le nombre de logements s'appuie sur le répertoire d'immeubles localisés décrivant une situation datée du 1er janvier de l'année médiane. De ce fait, dans les grandes villes, le nombre de ménages est exactement relatif au 1er janvier de l'année médiane, la taille moyenne des ménages étant celle mesurée au recensement. L'écart entre les effectifs moyens collectés et les effectifs « réels » au 1er janvier serait donc de l'ordre de l'évolution de la taille moyenne des ménages sur un mois appliquée au nombre de ménages concernés. Pour les petites communes, l'interpolation ou l'extrapolation pour se rapporter à cette date s'effectue à partir d'évolutions annuelles de données administratives et non sur l'évolution entre la date de référence de l'enquête annuelle et le 1er janvier de l'année médiane. Un écart de l'ordre de l'évolution de la

population sur un mois est donc possible ; cet écart devant être relativisé par le fait que l'extrapolation ou l'interpolation est elle-même une estimation, donc soumise à un aléa.

***In fine*, compte tenu de la complexité des calculs à réaliser et des hypothèses sous-jacentes à faire pour rapporter au 1er janvier les informations collectées sur quelques semaines à partir de mi janvier, il n'est pas certain que la correction marginale qui serait apportée pour rapporter les effectifs par sexe et âge au 1er janvier ne déformerait pas autant (à la marge) la pyramide des âges que l'absence de correction.**

II- La nouvelle pyramide des âges de référence pour le calcul des indicateurs démographiques

Pour éviter des *artefacts* dans les analyses des indicateurs démographiques annuels, notamment ceux calculés par âge, l'utilisation d'une pyramide cumulant les informations selon l'âge au recensement n'apparaît pas satisfaisante. Deux options sont dès lors envisageables :

- soit l'utilisation d'une pyramide des âges par génération à partir des cinq années de collecte en la corrigeant des décès (sans correction du faible biais lié au solde migratoire, pour les raisons évoquées dans la partie I-A, et en utilisant les pondérations annuelles plutôt que les pondération de cumul pour éviter les déformations évoquées dans la partie I-B) ;

- soit l'utilisation d'une pyramide des âges telle qu'elle résulte d'une seule enquête annuelle de recensement[17].

Concernant le « pic des 18 ans », dans les deux cas, en l'absence d'informations pour estimer l'ampleur de l'*artefact* sur les effectifs des personnes de 18 ans révolus au recensement, il est difficile de le corriger. Une pyramide par génération obtenue par cumul de cinq enquêtes annuelles lisse toutefois par construction cet *artefact*. Dans les deux cas également, aucun traitement particulier n'est réalisé pour se rapporter au 1er janvier pour les raisons expliquées dans la partie I-D.

II-A) Les options

Option 1 : Construire une pyramide des âges par génération à partir des cinq années de collecte en la corrigeant des décès

Avec les mêmes notations que précédemment, les effectifs de la génération *g* sont estimés par l'estimateur *P'(g) :*

$$P'(g)= \left(1/5.\left[P_r(g,n-2) + P_r(g,n-1) + P_r(g,n) + P_r(g,n+1) + P_r(g,n+2)\right] \right.$$
$$\left. - 1/5.[D(g,n-2) + 2.D(g,n-1) -2.D(g,n)- D(g,n+1)]\right).(P^*/P')$$

où : - P_r *est obtenu avec les pondérations annuelles de l'année considérée ;*
- *P* est la population légale totale, toutes générations confondues, datée du 1er janvier de l'année médiane ;*
- $P'=\Sigma_g \left(1/5.\left[P_r(g,n-2) + P_r(g,n-1) + P_r(g,n) + P_r(g,n+1) + P_r(g,n+2)\right] \right.$
$\left. - 1/5.[D(g,n-2) + 2.D(g,n-1) -2.D(g,n)- D(g,n+1)]\right).$

L'application du coefficient multiplicateur (P^*/P')[18] permet de caler la somme totale des effectifs par âge issue du cumul par génération sur les effectifs du recensement, c'est-à-dire la population légale au 1er janvier de l'année médiane. Ce recalage permet de conserver le rôle de référence qu'a le recensement. Effectué globalement, il permet de ne pas déformer la structure par sexe et âge issue de l'estimation faite à partir du cumul des âges par génération.

[17] Age révolu à la date de référence du recensement (mi-janvier)
[18] Il conviendra de s'assurer de la relative stabilité de ce rapport d'une année sur l'autre parce que les évolutions de ce rapport généreront du solde migratoire apparent, par comparaison de deux pyramides des âges, si cette option est mise en oeuvre.

Il conviendra de s'assurer régulièrement que le biais lié au solde migratoire reste négligeable en s'appuyant sur les estimations du solde migratoire par âge. Dans le cas contraire, les effectifs de la génération *g* pourront être estimés par *:*

$$P'(g)= \Big(1/5.\big[P_r(g,n-2) + P_r(g,n-1) + P_r(g,n) + P_r(g,n+1) + P_r(g,n+2)\big]$$
$$- 1/5.\big[D(g,n-2) + 2.D(g,n-1) -2.D(g,n)- D(g,n+1)\big]$$
$$+ 1/5.\big[SM(g,n-2) + 2.SM(g,n-1) - 2.SM(g,n) - SM(g,n+1)\big] \Big). (P^*/P')$$

Cette solution n'est pas privilégiée d'emblée dans la mesure où les estimations du solde migratoire par âge sont pour le moment fragiles en dehors de sources d'informations solides sur les flux d'entrées et de sorties du territoire national, hormis les entrées légales d'étrangers ressortissant de pays non membres de l'Union européenne.

Par ailleurs, l'utilisation des enquêtes annuelles, donc des pondérations des enquêtes annuelles, est privilégiée par rapport à une tabulation directe selon l'année de naissance des résultats du cumul de cinq enquêtes annuelles successives (c'est-à-dire la pondération du cumul). Cette solution permet d'éviter les effets de déformation de la structure par âge induite par le calcul des populations légales au 1er janvier de l'année médiane. Le fait de corriger le biais lié aux sorties par décès suffit dans tous les cas à modifier la structure par sexe et année de naissance par rapport à celle obtenue dans le recensement avec la pondération de cumul.

Pour prendre en compte les effets de bord pour les générations nées l'année *n-2* et *n-1*, plusieurs solutions sont envisageables :

- Utiliser la même démarche générale, en se limitant aux seules enquêtes annuelles pertinentes (utilisation de quatre enquêtes annuelles pour la génération née l'année *n-2* et de trois enquêtes annuelles pour celle née l'année *n-1*), soit :

$$P'(n-2)= \Big(1/4.\big[P_r(n-2,n-1) + P_r(n-2,n) + P_r(n-2,n+1) + P_r(n-2,n+2)\big]$$
$$- 1/4.\big[D(n-2,n-1) -2.D(n-2,n)- D(n-2,n+1)\big]\Big).(P^*/P')$$

$$P'(n-1)= \Big(1/3.\big[P_r(n-1,n) + P_r(n-1,n+1) + P_r(n-2,n+2)\big]$$
$$- 1/3.\big[-2.D(n-1,n)- D(n-1,n+1)\big]\Big).(P^*/P')$$

- Déduire les effectifs de ces générations du nombre de naissances enregistrées dans les statistiques d'état civil à ces âges en les corrigeant des décès enregistrés dans les statistiques d'état civil pour ces générations entre le 1er janvier de l'année *n-2* et le 1er janvier de l'année *n* pour la génération née en *n-2* (entre le 1er janvier de l'année *n-1* et le 1er janvier de l'année *n* pour la génération née en *n-1*). Cette deuxième estimation est plus élevée que la première. De fait, pour des raisons qui resteraient à mieux cerner, les enquêtes annuelles de recensement, comme auparavant les recensements généraux de population, donnent une estimation du nombre d'enfants de très bas âges plus faible que le nombre qui se déduit des statistiques d'état-civil, à savoir les naissances corrigées des décès[19].

Option 2 : Utiliser la pyramide des âges fournie par l'enquête annuelle

Dans ce cas, la pyramide au 1er janvier de l'année *n* résulterait de l'application de la structure par âge de la pyramide de l'enquête annuelle collectée l'année *n* (avec sa pondération annuelle) aux effectifs totaux issus du cumul de cinq enquêtes annuelles.

[19] Cette option a été testée et écartée début 2009 lors du calcul de la pyramide des âges du 1er janvier 2006, à l'aide des cinq premières enquêtes annuelles de recensement.

Par rapport à la solution précédente, sa mise en œuvre serait plus simple puisqu'il n'y aurait plus à tenir compte des biais liés à une collecte étalée dans le temps.

En revanche, même si la taille de l'échantillon des enquêtes annuelles permet de disposer de résultats précis au niveau national, le biais lié aux traitements du recensement pourrait être plus important qu'en utilisant le cumul sur cinq ans. En reprenant la formule de la partie I-A relative au cas d'un cumul par génération, cela serait le cas si :

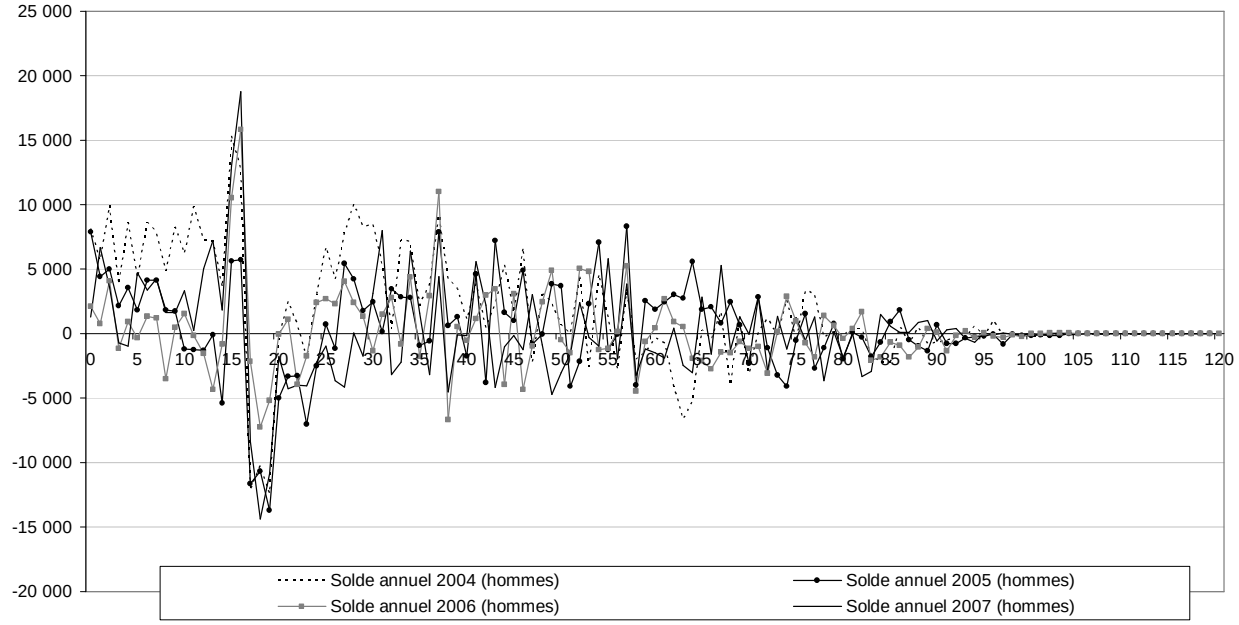
$$e(g,n) > 1/5.\left[\, e(g,n-2) + e(g,n-1) + e(g,n) + e(g,n+1) + e(g,n+2)\right]$$

Par ailleurs, **un aléa faible et acceptable si l'on s'intéresse au niveau des effectifs d'une génération donnée peut être significatif pour le solde migratoire apparent déduit de la comparaison des résultats de deux enquêtes annuelles successives.** De ce point de vue, l'utilisation des enquêtes annuelles de recensement conduit à des soldes apparents par sexe et âge instables d'une année sur l'autre (graphique 8), et beaucoup plus heurtés que ceux implicitement publiés actuellement (graphique 9), sans qu'on ne dispose d'éléments pour juger si cette instabilité reflète la réalité du solde migratoire.

Graphique 8

Soldes migratoires apparents des hommes

Source EAR (pondérations annuelles). Champ : France entière

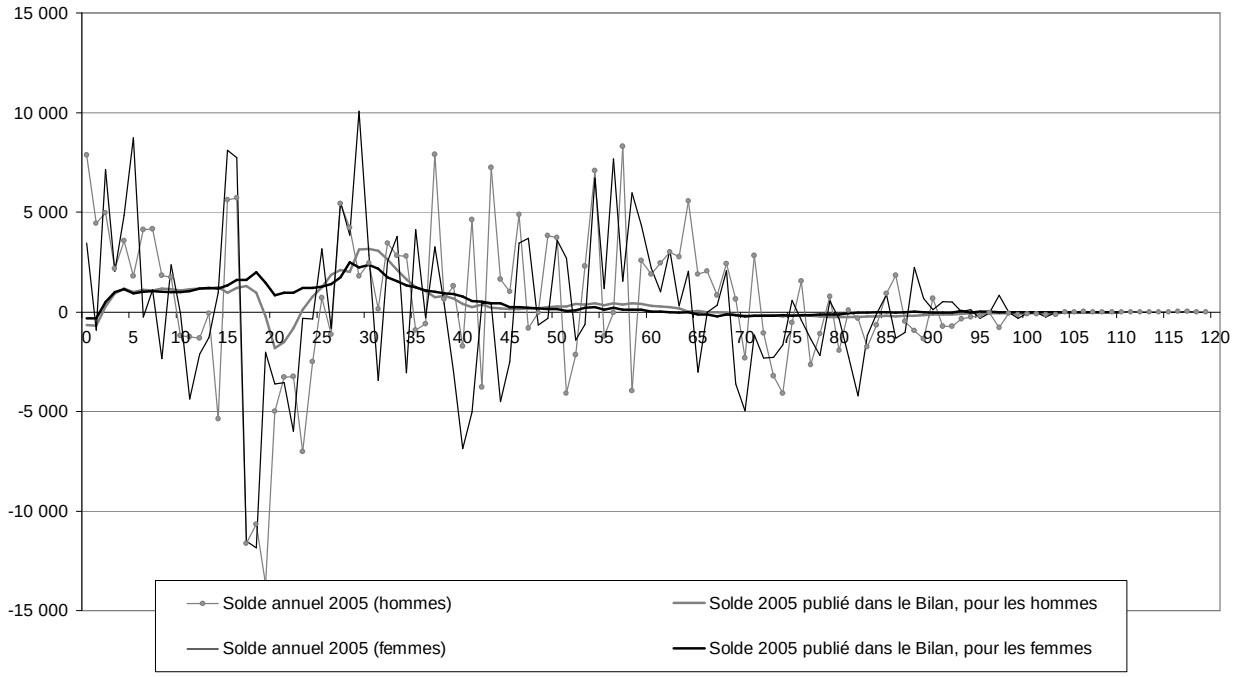


Source : Insee, Enquêtes annuelles de recensement 2004 à 2008. Statistiques d'état civil.
Champ : France entière.
Lecture : Le solde migratoire apparent qui se déduit de la comparaison des résultats annuels des enquêtes de recensement et du solde naturel est, pour les hommes de 28 ans, de près de 10 000 en 2004, moitié moins en 2005 et nul en 2007.

Graphique 9

Comparaison des soldes migratoires estimés par les enquêtes annuelles de recensement avec ceux du Bilan démographique en 2005

Source EAR (pondérations annuelles). Champ : France entière.



Source : Insee, Enquêtes annuelles de recensement 2004 à 2008. Statistiques d'état civil.
Champ : France entière.

Impact de l'option choisie sur le solde migratoire apparent par âge

En repartant des formulations utilisées précédemment, dans le cas de l'option 2, le solde migratoire apparent pour la génération *g* (d'un genre donné) au cours de l'année serait estimé en comparant deux enquêtes annuelles de recensement successives. En dehors du cas particulier de l'âge zéro, il serait donc égal à :

$$SM\ apparent = P_r(g,n+1) - P_r(g,n) - D(g,n)$$

$$= P(g,n+1) - P(g,n) - D(g,n) - e(g,n+1) + e(g,n)$$

$$= SM(g,n) - e(g,n+1) + e(g,n)$$

Le solde apparent ainsi estimé correspond donc au solde réel aux erreurs de mesure et aux aléas de collecte du recensement près.

Dans le cas de l'option 1, le solde apparent estimé correspond en revanche approximativement à un solde migratoire moyen sur cinq ans aux erreurs de mesure. Ces termes d'erreurs jouent avec la même ampleur, indéterminée, qu'avec l'option 2, puisqu'il s'agit encore d'une différence de deux aléas annuels (on rappelle que pour la pyramide des âges, l'aléa joue en moyenne sur cinq ans, et doit donc être plus stable et moins élevé).

Dans le cas de l'option 1, le solde migratoire apparent pour la génération *g* au cours de l'année *n*, s'écrit en effet :

$$
\begin{aligned}
SM\ apparent\ &= P'_{+1}(g) - P'(g) + D(g,n) \\
&= \Big(1/5.\big[P_r(g,n-1) + P_r(g,n) + P_r(g,n+1) + P_r(g,n+2) + P_r(g,n+3)\big] \\
&\quad - 1/5.\big[D(g,n-1) + 2.D(g,n) - 2.D(g,n+1) - D(g,n+2)\big]\Big).(P^*_{+1}/P_{+1}') \\
&\quad - \Big(1/5.\big[P_r(g,n-2) + P_r(g,n-1) + P_r(g,n) + P_r(g,n+1) + P_r(g,n+2)\big] \\
&\quad - 1/5.\big[D(g,n-2) + 2.D(g,n-1) - 2.D(g,n) - D(g,n+1)\big]\Big).(P^*/P') + D(g,n)
\end{aligned}
$$

Si le ratio *P*/P'* est stable dans le temps :

$$
\begin{aligned}
SM\ apparent\ &\sim \Big(1/5.\big[P_r(g,n-1) + P_r(g,n) + P_r(g,n+1) + P_r(g,n+2) + P_r(g,n+3)\big] \\
&\quad - 1/5.\big[D(g,n-1) + 2.D(g,n) - 2.D(g,n+1) - D(g,n+2)\big] \\
&\quad - 1/5.\big[P_r(g,n-2) + P_r(g,n-1) + P_r(g,n) + P_r(g,n+1) + P_r(g,n+2)\big] \\
&\quad + 1/5.\big[D(g,n-2) + 2.D(g,n-1) - 2.D(g,n) - D(g,n+1)\big]\Big).(P^*/P') + D(g,n) \\
&= 1/5.(P^*/P').\big[P_r(g,n+3) - P_r(g,n-2) \\
&\qquad + D(g,n-2) + D(g,n-1) + D(g,n) + D(g,n+1) + D(g,n+2)\big] \\
&\qquad + \big[1 - P^*/P'\big].D(g,n) \\
&= 1/5.(P^*/P').\big[P(g,n+3) - P(g,n-2) - e(g,n+3) + e(g,n-2) \\
&\qquad + D(g,n-2) + D(g,n-1) + D(g,n) + D(g,n+1) + D(g,n+2)\big] \\
&\qquad + \big[1 - P^*/P'\big].D(g,n) \\
&= 1/5.(P^*/P').\big[SM(g,n+2) + SM(g,n+1) + SM(g,n) + SM(g,n-1) + SM(g,n-2)\big] \\
&\qquad - 1/5.(P^*/P').\big[e(g,n+3) + e(g,n-2))\big] \\
&\qquad + \big[1 - P^*/P'\big].D(g,n))
\end{aligned}
$$

Si le ratio *P*/P'* est proche de 1 :

$$
\begin{aligned}
SM\ apparent\ &\sim 1/5.\big[SM(g,n+2) + SM(g,n+1) + SM(g,n) + SM(g,n-1) + SM(g,n-2)\big] \\
&\quad - 1/5.\big[e(g,n+3) + e(g,n-2))\big]
\end{aligned}
$$

Ce calcul résulte de deux approximations liées au calage de la population estimée sur la population légale (ratio *P*/P'*), chacune pouvant être source de biais : la stabilité du ratio *P*/P'*, d'une part ; sa proximité avec 1, d'autre part. Ainsi, dans le cas de l'option 1, si le coefficient de redressement est proche de 1 et stable dans le temps, le solde migratoire apparent sera légèrement sous-évalué (si le facteur est inférieur à 1) ou sur-évalué (si le facteur est supérieur à 1), mais l'évolution de la structure par sexe et âge du solde migratoire apparent ne sera pas perturbée. Elle peut en revanche être légèrement perturbée dans le cas contraire.

II-B) La méthode retenue et ses effets sur les indicateurs démographiques

La méthode retenue

En photographie à une date donnée, l'option 1 apparaît plus robuste que l'option 2, notamment parce que les termes d'erreurs sont davantage lissés. Par ailleurs, en évolution, les termes d'erreurs sont du même ordre de grandeur dans les deux options, l'option 1 conduisant toutefois également à un solde migratoire apparent lissé, si les aléas de collecte sont faibles au regard des soldes migratoires par sexe et âge. Il est donc par construction plus stable qu'avec l'option 2, ce qui peut être préférable en terme de communication alors que les sources statistiques manquent pour disposer d'estimations fiables du solde migratoire réel. Finalement, c'est donc l'option 1 qui a été retenue pour les estimations de population nationale utilisée pour le calcul des principaux indicateurs démographiques.

Pour obtenir la pyramide des âge de référence à la date du recensement (par exemple au 1er janvier 2006), on commence donc par calculer les effectifs d'une génération et d'un sexe donnés en faisant la moyenne des effectifs de cette génération estimés dans chacune des cinq enquêtes annuelles de recensement correspondantes avec leurs pondérations annuelles (les enquêtes annuelles 2004 à 2008 pour la pyramide des âges au 1er janvier 2006). Ces effectifs sont ensuite corrigés du biais des décès, conformément à la relation indiquée dans la partie II-A. Puis un calage est effectué de façon à ce que la population totale de la pyramide des âges corresponde à la population légale totale. En d'autres termes, la structure par sexe et âge calculée est appliquée à la population légale totale.

Pour prendre en compte les effets de bord pour les générations nées l'année *n-2* et *n-1*, on utilise la même démarche générale, en se limitant aux seules enquêtes annuelles pertinentes (utilisation de quatre enquêtes annuelles pour la génération née l'année *n-2* et de trois enquêtes annuelles pour celle née l'année *n-1*). Il n'y a pas de traitement particulier pour se rapporter au 1er janvier, parce que la complexité du traitement pourrait créer autant de biais que ceux qu'il s'agit de corriger, qui restent minimes. Il n'y a pas non plus de traitement du « pic des 18 ans », qui subsiste mais lissé par la moyenne sur cinq ans.

En reprenant les formules précédentes, on a donc pour les générations nées avant l'année n-2 :

$$P'(g)= \big(1/5.\big[P_r(g,n-2) + P_r(g,n-1) + P_r(g,n) + P_r(g,n+1) + P_r(g,n+2)\big] - 1/5.\big[D(g,n-2) + 2.D(g,n-1) -2.D(g,n)- D(g,n+1)\big]\big).(P^*/P')$$

Pour la génération née l'année n-2 :

$$P'(n-2)= \big(1/4.\big[P_r(n-2,n-1) + P_r(n-2,n) + P_r(n-2,n+1) + P_r(n-2,n+2)\big] - 1/4.\big[D(n-2,n-1) -2.D(n-2,n)- D(n-2,n+1)\big]\big).(P^*/P')$$

Pour la génération née l'année n-1:

$$P'(n-1)= \big(1/3.\big[P_r(n-1,n) + P_r(n-1,n+1) + P_r(n-1,n+2)\big] - 1/3.\big[-2.D(n-1,n)- D(n-1,n+1)\big]\big).(P^*/P')$$

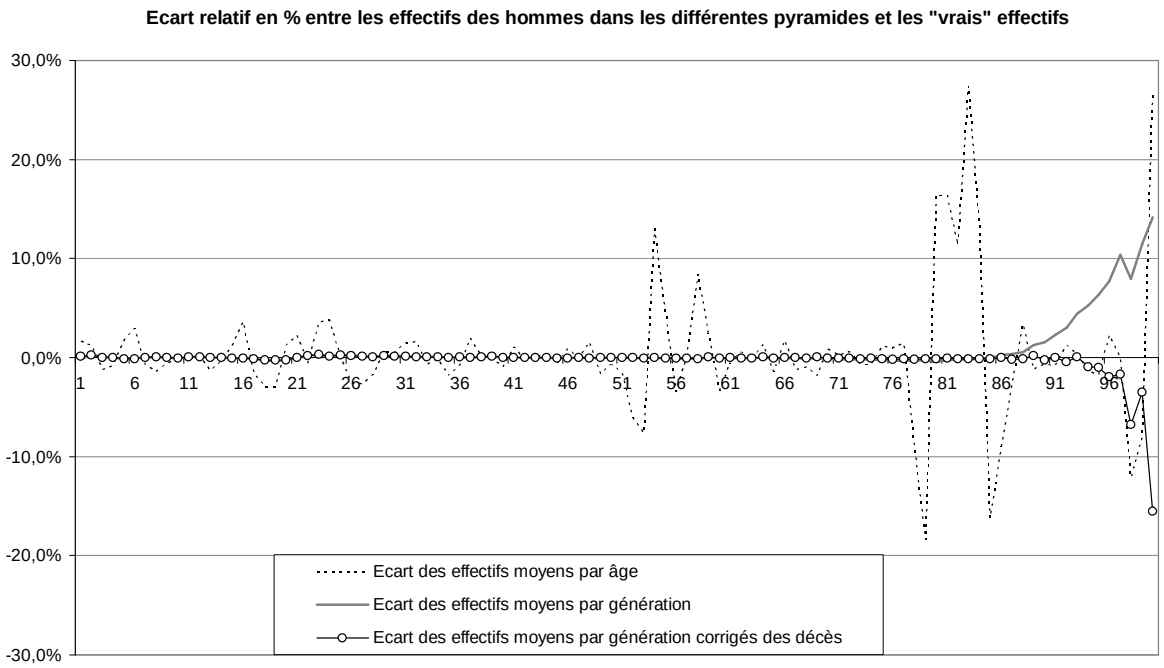
Une illustration des effets de la correction des décès à partir d'une simulation

Pour illustrer les effets du choix effectué sur les indicateurs démographiques, la simulation utilisée dans la première partie peut être à nouveau mobilisée. On peut alors comparer différents indicateurs selon que l'on utilise la « vraie » pyramide des âges, dans le cadre de la simulation, ou une pyramide

utilisant un cumul d'information selon la génération et avec correction des décès. On présentera également pour mémoire les indicateurs utilisant la pyramide des âges issue d'une moyenne selon l'âge et celle issue d'une moyenne selon la génération sans correction des décès.

Le graphique 10 (page suivante) montre que la correction des décès permet de supprimer la surestimation croissante avec l'âge qui était observée aux grands âges dans le cas de l'utilisation d'un cumul par génération sans correction. Il reste cependant des écarts importants dans quelques cas, en raison d'effectifs faibles. Ces légers écarts sont donc sans grands effets sur la plupart des indicateurs démographiques. Ils jouent surtout sur les quotients de mortalité aux grands âges, sans effet significatif sur l'espérance de vie à la naissance. A ces âges, la mesure des quotients de mortalité est une difficulté récurrente, à l'origine de travaux de recherche spécifiques[20].

Graphique 10 - Les effectifs des hommes selon l'âge (simulation sur un cumul 1997-2001)



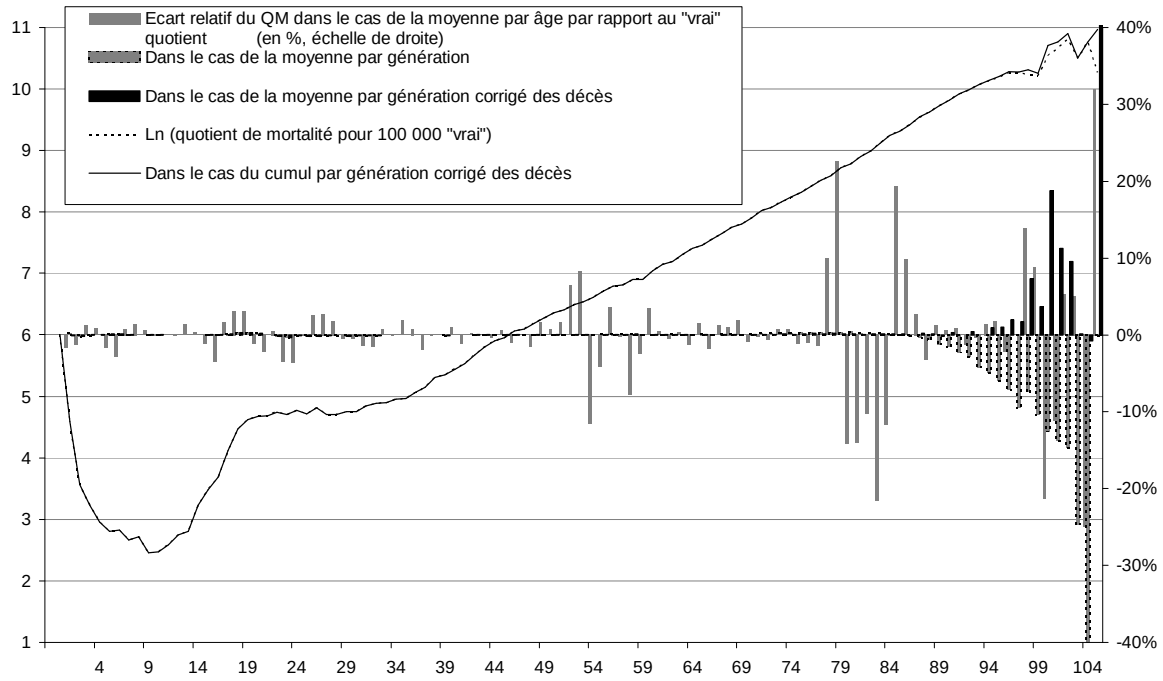
Champ : France métropolitaine.

Les graphiques 11a et 11b montrent les quotients de mortalité par âge des hommes en 1999, obtenus dans le cadre de la simulation. La correction des décès permet mécaniquement de supprimer la sous-estimation croissante avec l'âge des quotients de mortalité qui était observée aux grands âges dans le cas de l'utilisation d'un cumul par génération sans correction. Les très légers écarts observés sur les moins de 40 ans subsistent.

[20] Voir par exemple France Meslé et Jacques Vallin, « Comment améliorer la précision des tables de mortalité aux grands âges », *Population*, 2002, n°4-5, volume n° 57, Ined. Des échantillons spécifiques ont été mis en place pour étudier la mortalité aux grands âges. En France, un échantillon « très grands âges » et un échantillon « complément grands âges » ont ainsi été constitués à partir du recensement de 1999 (voir Isabelle Robert-Bobbée, Christian Monteil, Olivier Cadot, « La mortalité aux grands âges en France : nouvelles données, nouveaux résultats », *Document de travail*, n°F0701, février 2007, Insee).

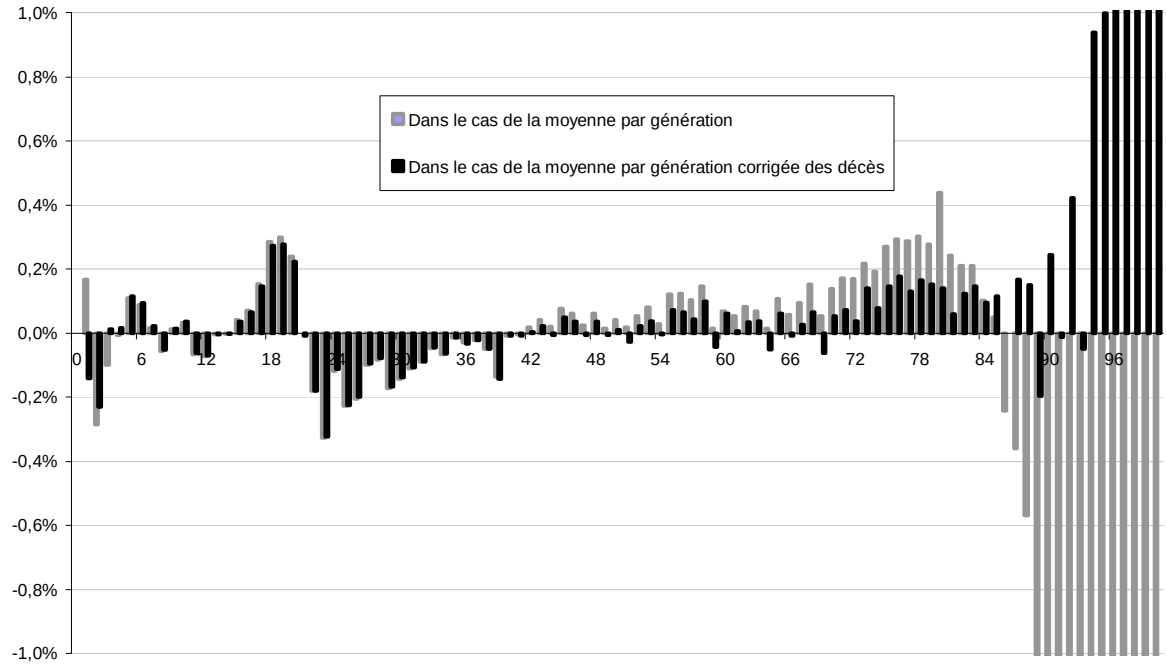
Graphiques 11a et 11b - Les quotients de mortalité des hommes

Quotient de mortalité des hommes 1999 (simulation)



Champ : France métropolitaine.

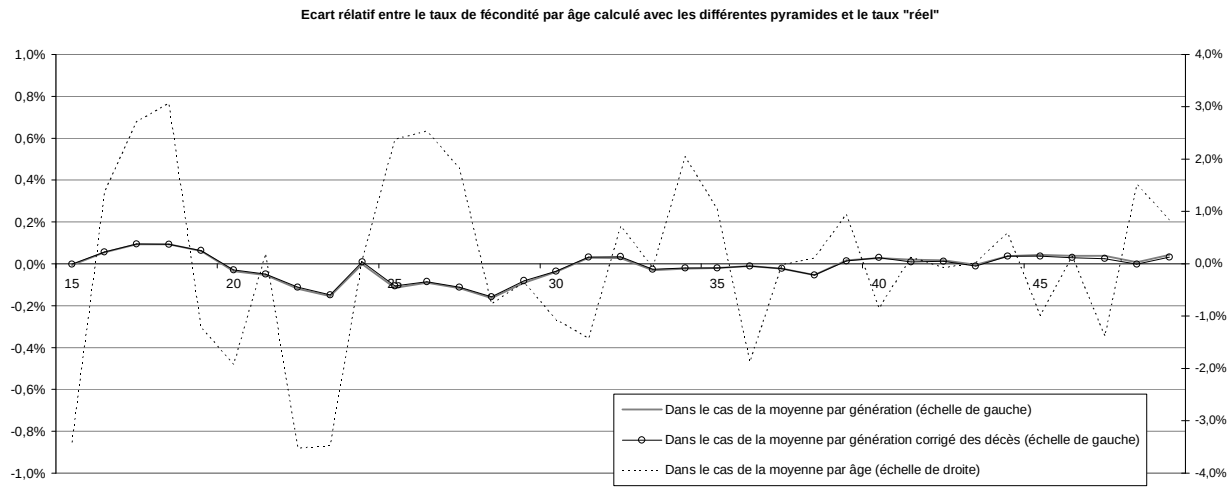
Ecart relatif en % du quotient de mortalité simulé par rapport au quotient "vrai" (cas des hommes)



Champ : France métropolitaine.

S'agissant du taux de fécondité par âge (graphique 12), la simulation illustre le faible impact de la correction des décès. Ce résultat était attendu puisque, pour la tranche d'âges concernée, les effectifs sont peu affectés par la correction des décès.

Graphique 12- Les taux de natalité par âge de la mère



Champ : France métropolitaine.

Enfin, le tableau ci-dessus présente deux des principaux indicateurs démographiques synthétiques : l'espérance de vie (selon le sexe) et l'indicateur conjoncturel de fécondité. Le choix de la méthode a peu d'effet sur leur niveau.

Tableau 1 : Simulation de l'espérance de vie et l'indicateur conjoncturel de fécondité 1999

		Indicateurs calculés à partir de :		
	la « vraie » pyramide au 1er janvier 1999	la pyramide issue d'un cumul par génération corrigée des décès	la pyramide issue d'un cumul par génération	la pyramide issue d'un cumul par âge
Espérance de vie à la naissance des hommes	75,0	75,0	75,0	75,1
Espérance de vie à la naissance des femmes	82,5	82,5	82,6	82,6
Indicateur conjoncturel de fécondité	1,78	1,78	1,78	1,78

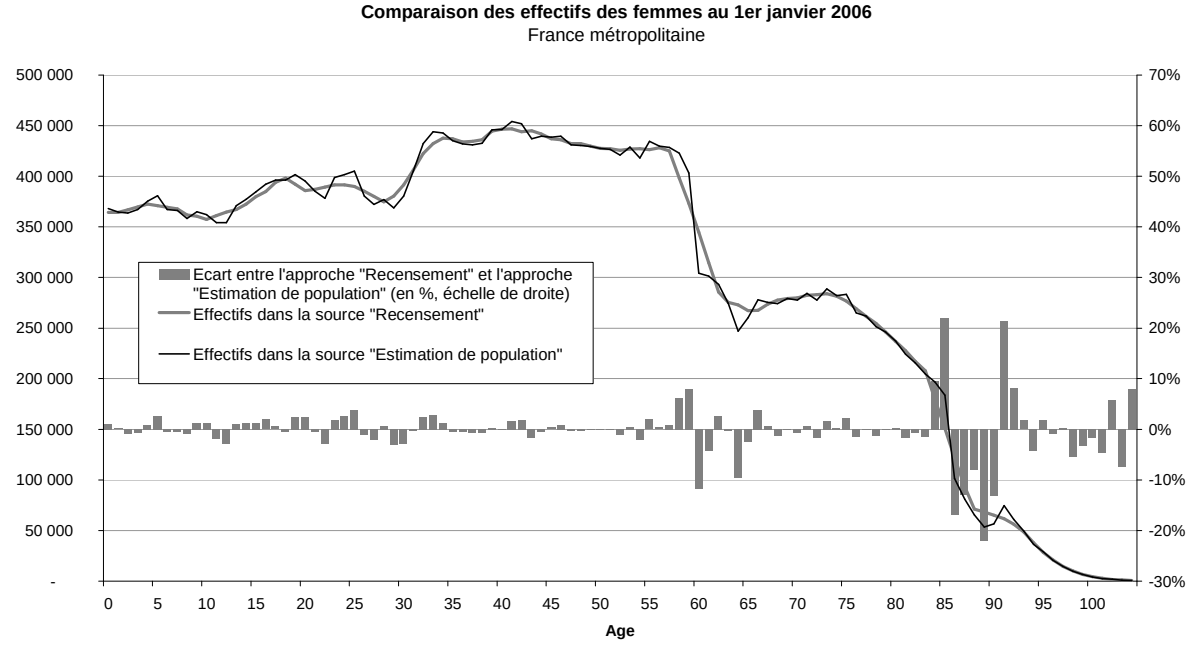
Champ : France métropolitaine.

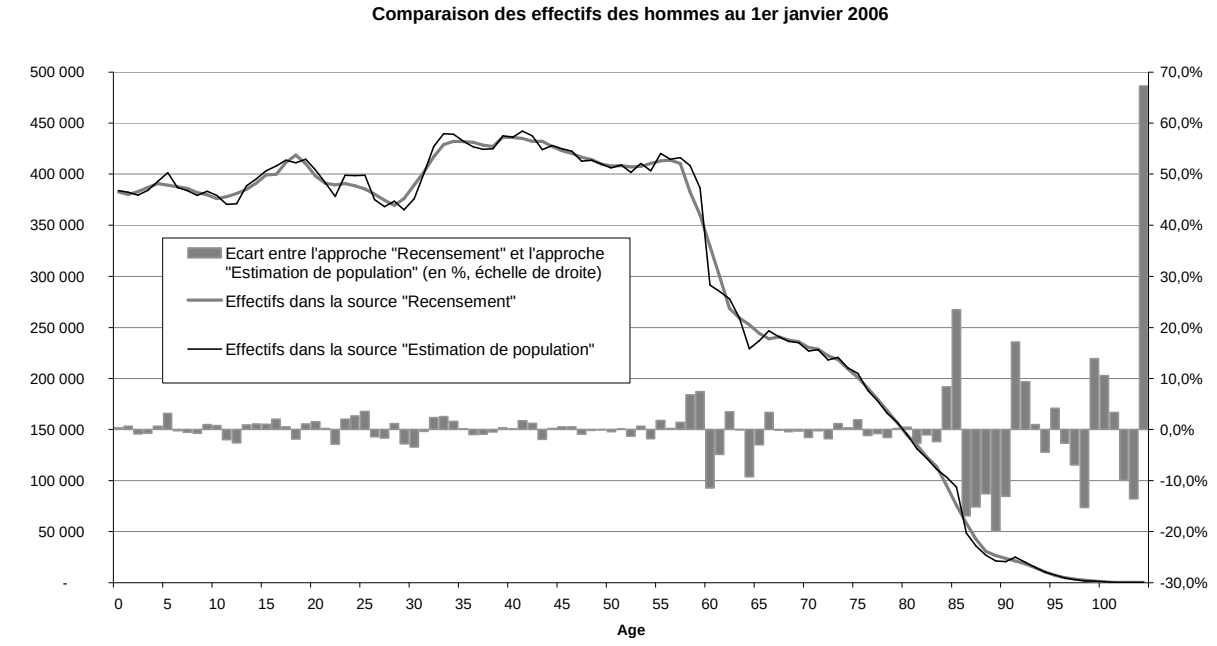
Comparaison de la pyramide des estimations de population et de la pyramide du recensement

Avec les résultats du recensement de 2006, il est également possible de comparer les indicateurs démographiques selon qu'ils utilisent la pyramide des âges issue directement des résultats du recensement ou la pyramide des âges issue des estimations de population. La première correspond à une moyenne des informations collectées sur cinq ans selon l'âge à la date de l'enquête annuelle et intègre les légers effets de déformation de structure liés à l'utilisation de la pondération de cumul du recensement. La seconde est calculée selon la méthode retenue *supra* et correspond donc à une moyenne des informations collectées sur cinq ans selon la génération, avec les pondérations

annuelles des enquêtes annuelles, une correction des décès et un calage sur la population légale totale.

Graphiques 13a et 13b - Effectifs des femmes et des hommes selon l'âge, au 1er janvier 2006





Champ : France métropolitaine.

Les graphiques 13a et 13b illustrent, à partir des effectifs des femmes et des hommes selon l'âge, l'effet de lissage induit par un cumul selon l'âge, qui n'existe pas avec la seconde approche. A la différence de la simulation présentée plus haut sur la période 1997-2001, il n'est ici évidemment pas possible de comparer l'une et l'autre à une situation « vraie », par nature inconnue.

Le tableau 2 présente la structure de la population au 1er janvier 2006 par âge selon les deux approches. Les résultats apparaissent très proches.

Tableau 2 - Structure de la population par âge au 1er janvier 2006

	Hommes		Femmes	
	Source "Recensement"	Source "Estimation de population"	Source "Recensement"	Source "Estimation de population"
0-4 ans	6,5%	6,5%	5,8%	5,8%
5-19 ans	19,8%	19,8%	18,8%	18,9%
20-29 ans	12,9%	13,0%	13,0%	13,0%
30-39 ans	14,2%	14,2%	14,4%	14,4%
40-49 ans	14,3%	14,3%	14,8%	14,8%
50-54 ans	6,9%	6,8%	7,2%	7,1%
55-59 ans	6,7%	6,9%	6,9%	7,1%
60-64 ans	4,7%	4,5%	5,0%	4,8%
65-69 ans	4,0%	4,0%	4,6%	4,6%
70-79 ans	4,5%	4,5%	6,3%	6,3%
80-89 ans	2,8%	2,8%	5,3%	5,3%
90 ans et plus	0,4%	0,4%	1,2%	1,2%
Moins de 20 ans	26,2%	26,3%	24,6%	24,7%
60 ans et plus	16,4%	16,2%	22,4%	22,2%
75 ans et plus	6,2%	6,2%	10,9%	10,9%

Champ : France métropolitaine.

III- Faut-il décliner la même méthode à tous les échelons géographiques et pour toutes les analyses par âge ?

La méthodologie présentée pour construire une pyramide des âges « exacte » plutôt qu'une pyramide lissée à partir des informations du nouveau recensement cherche à répondre à un besoin spécifique : il s'agit de disposer d'une pyramide conceptuellement cohérente avec les pyramides des âges utilisées jusqu'à présent pour le calcul des indicateurs démographiques et, plus particulièrement, d'éviter que les analyses démographiques soient perturbées par des *artefact* liés à la méthode d'estimation, la première partie permettant de soulever notamment le cas des quotients de mortalité détaillés.

Ce n'est pas pour autant qu'il serait indispensable de retravailler les données issues du recensement pour construire des indicateurs démographiques infra nationaux ou, plus généralement, pour tout indicateur décliné selon l'âge. Certes, le fait d'avoir explicités les biais pouvant exister entre la pyramide des âges issue du nouveau recensement et la pyramide « réelle » inconnue, met en avant des difficultés. Toutefois, il en existait également avec l'utilisation de recensements exhaustifs ponctuels.

D'une part, il existait déjà des aléas de collecte. Les enquêtes post censitaires réalisées après le recensement de 1990 ont montré que les double-comptes (0,7% de la population recensée) et les omissions (1,8%) se combinaient pour conduire à un taux d'omission net de l'ordre de 1,1%, soit 600 000 personnes. L'introduction d'un ajustement de 480 000 personnes sur la période 1990-1998 pour compenser une variation de population entre les recensements de 1990 et 1999 plus faible que celle déduite du solde migratoire et du solde naturel, l'illustre également.

D'autre part, les biais liés aux flux migratoires et aux soldes naturels existaient aussi *de facto* lorsque l'on utilisait les données du dernier recensement disponible pour conduire des analyses ou établir des diagnostics valant pour le présent. De ce point de vue, le fait de disposer chaque année de données moyennes à partir d'informations collectées sur les cinq dernières années permettra de disposer d'informations en moyenne plus récentes pour réaliser ces analyses et ces diagnostics.

III-a) La question des pyramides des âges infra nationales

L'utilisation d'une pyramide des âges moyenne de cinq années de collecte selon « la génération » et corrigée des décès peut techniquement être envisagée à des échelons infra nationaux, les informations de l'état civil étant disponibles à la commune de façon exhaustive. L'opportunité de cette démarche doit cependant être étudiée en tenant compte du fait que, plus l'échelon géographique est fin, moins les sources de biais imputables aux flux migratoires sont négligeables et plus la précision des résultats des enquêtes annuelles s'atténue.

Autrement dit, la source d'écart suivante, qu'on a pu négliger dans la pyramide des âges nationale issue du cumul par génération et corrigée des décès, pourrait être moins négligeable :

$$- \alpha_1(g,n).SM(g,n-2) + (\alpha_1(g,n)+ \alpha_2(g,n)).SM(g,n-1) - (\alpha_4(g,n)+\alpha_5(g,n)).SM(g,n) -\alpha_5(g,n).SM(g,n+1)$$
$$- \alpha_1 (g,n).e(g,n-2) + \alpha_2 (g,n).e(g,n-1) + \alpha_3 (g,n).e(g,n) + \alpha_4 (g,n).e(g,n+1)+ \alpha_5(g,n).e(g,n+2)$$

De plus, si des estimations nationales permettent d'envisager, quand cela deviendra nécessaire, de corriger des flux migratoires, cela serait beaucoup moins envisageable au niveau local, faute d'estimations disponibles sur les soldes migratoires infra nationaux. Comme au niveau national, les recensements permettent de calculer à ces échelons un solde migratoire apparent, somme du solde migratoire réel inconnu et des erreurs de mesure du recensement. Mais au niveau national, il existe d'autres sources statistiques sur les flux migratoires qui, même si elles sont parcellaires, contribuent à l'estimation et à la validation des estimations de solde migratoire. Or elles ne sont pas disponibles aux échelons infra nationaux.

A des échelons géographiques infra nationaux, les autres sources possibles de biais ne sont plus de second ordre, qu'il s'agisse du terme d'erreur intégrant l'imprécision lié à l'échantillonnage ou des biais liés aux flux migratoires, deux termes dont l'importance augmente quand l'échelon géographique devient plus fin. Dans ce contexte, cumuler les informations selon la génération et corriger des décès n'améliore plus nécessairement la précision de la pyramide des âges. Le faire malgré tout reviendrait à afficher auprès des utilisateurs non spécialistes qu'il existe un travail méthodologique pour construire une pyramide des âges « exacte » plutôt que « lissée » qui serait en partie illusoire, car la pyramide des âges ainsi construite n'en serait pas forcément plus exacte pour autant.

Au cas par cas, les deux écueils soulevés pourraient être levés pour certains territoires en utilisant un critère de taille pour la précision et en examinant la contribution des flux migratoires à la croissance du territoire considéré. Mais cet examen au cas par cas supprimait toute homogénéité de traitement au sein d'un même échelon territoriale ou statistique, comme l'aire urbaine ou le département. Or, l'utilisation d'une pyramide lissée telle que celle issue simplement d'un cumul des enquêtes annuelles selon l'âge peut alors suffire pour la plupart des utilisateurs. Des travaux réalisés sur les indicateurs démographiques régionaux, présentés dans l'annexe 2, suggèrent d'ailleurs une faible influence du choix de la pyramide pour les principaux constats démographiques, comme l'espérance de vie, l'indicateur conjoncturel de fécondité ou la répartition de la population par tranche d'âges. Les quotients démographiques détaillés par âge, plus sensibles au choix de la pyramide utilisée, ne sont habituellement pas diffusés par l'Insee à ces échelons, en raison de leur relative instabilité d'une année sur l'autre : les probabilités de survenue des évènements d'état civil (mariage, maternité, décès) se réalisent en effet avec un aléa d'autant plus grand qu'il porte sur des populations de faibles effectifs. Les évolutions aléatoires qui en résultent ne reflètent alors pas une évolution tendancielle des risques.

En pratique, lors de la publication des résultats du premier cycle de cinq enquêtes annuelles de recensement successives, relatifs au 1er janvier 2006, il a toutefois été décidé de procéder de manière analogue à la méthode nationale pour le calcul des pyramides des âges départementales et régionales de façon à assurer une cohérence interne à la source « estimation de population », les pyramides lissées issues directement du recensement étant par ailleurs disponibles avec les résultats du recensement.

III-b) Étudier un comportement sociodémographique lié à l'âge

Les parties I et II ont montré comment utiliser le recensement pour produire une pyramide des âges, soit directement issue d'une tabulation du recensement selon l'âge au recensement, soit en le tabulant par génération et en le corrigeant des décès. Le calcul de certains indicateurs et certains usages nécessitent une pyramide des âges ventilée selon des critères supplémentaires : par exemple, pour construire une pyramide des âges selon la situation matrimoniale, afin de calculer certains indicateurs démographiques comme le taux de nuptialité selon l'état matrimonial antérieur ou le taux de primo-nuptialité ; par exemple, pour construire une pyramide des personnes vivant en ménage ordinaire pour caler les enquêtes réalisées auprès des ménages. Plus généralement, le recensement permet de décrire la population, pour les différents échelons géographiques, selon de nombreuses caractéristiques sociodémographiques où une analyse par âge peut intervenir. Outre la situation matrimoniale, on peut citer le type de famille, la situation sur le marché du travail, ...

Pour ces différents usages, les résultats du recensement pourraient, comme pour la pyramide des âges elle-même, être utilisés en cumulant les informations de cinq enquêtes annuelles successives, soit selon l'âge au recensement, soit selon l'année de naissance ou la génération. S'agissant de comportements sociodémographiques liés à l'âge, l'utilisation directe de l'information selon l'âge au recensement est la plus pertinente.

Prenons l'exemple de la pyramide des âges nationale des personnes vivant dans des ménages ordinaires, utilisée pour le calage de l'enquête Emploi (d'autres enquêtes l'utilisent aussi indirectement en se calant sur des marges calculées avec l'enquête Emploi). Pour les estimations à des dates

postérieures au recensement, elle s'obtient en appliquant aux estimations de population par sexe et âge, la part des personnes vivant en « ménages » observée lors du dernier recensement.

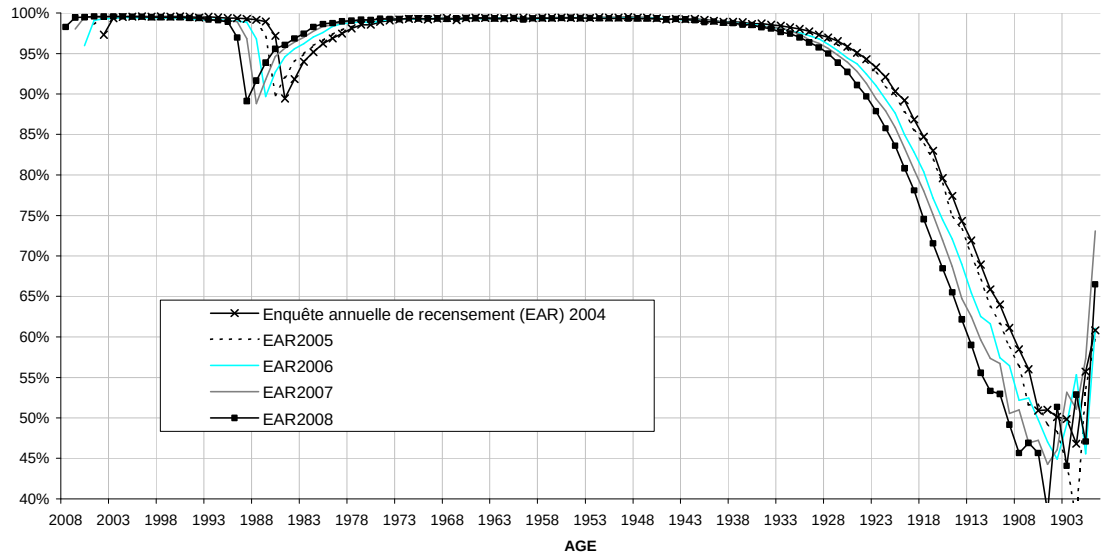
Dans le cadre du nouveau recensement, au cours de l'année *n*, on dispose désormais :

- d'estimations annuelles de la part des personnes vivant en ménages ordinaires par sexe et âge, à partir des enquêtes annuelles (pondération annuelle), jusqu'au 1er janvier de l'année *n-1* ;
- d'une estimation de cette part, par sexe et âge, à partir du cumul de cinq années de collecte (pondération du cumul), datée du 1er janvier *n-3*.

Dans tous les cas, les informations disponibles sont plus récentes que celles disponibles auparavant avec les recensements généraux de population.

Graphique 14

Part des Femmes vivant en ménages, selon leur année de naissance
Résultats par enquête annuelle, pondération annuelle, France métropolitaine



Champ : France métropolitaine.

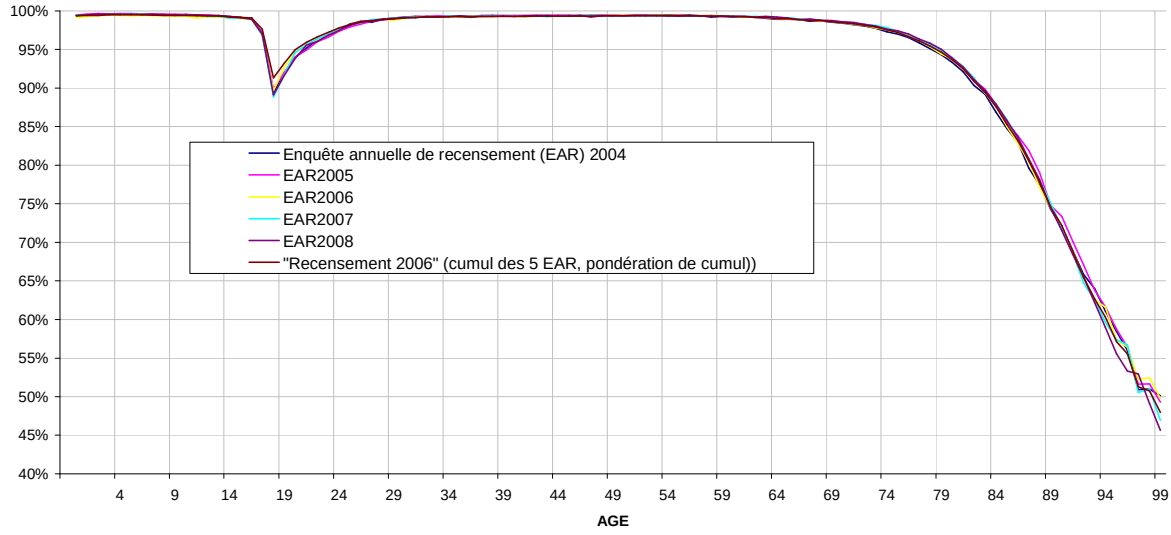
Le graphique 14 montre la part des femmes vivant dans un logement ordinaire dans les enquêtes annuelles 2004 à 2007, selon l'année de naissance (les résultats sont similaires pour les hommes). Cette part diminue aux âges d'études, avec les passages en internat et en cité universitaire. Le creux brutal au passage de la majorité est un *artefact* lié au mode de comptabilisation par le recensement les personnes scolarisées vivant en internat dans un établissement d'enseignement mais ayant leur résidence familiale dans une autre commune. Quand il s'agit de personnes majeures, elles sont comptées dans la commune d'études, donc « hors ménages ». *A contrario*, les élèves ou étudiants mineurs logés dans une cité universitaire (ou un foyer d'étudiants ou un logement) situé dans une autre commune que leur résidence familiale sont comptés dans la population des ménages de la commune de leur résidence familiale. Aux âges avancés, la part des personnes en ménage diminue régulièrement, avec les passages en maison de retraites ou autre institution spécialisée. Ces phénomènes sont, par nature, très liés à l'âge. De ce fait, les courbes par génération se décalent d'une enquête annuelle sur l'autre, témoignant de cette forte relation du phénomène étudié avec l'âge.

Cumuler les cinq enquêtes annuelles selon la génération conduirait à lisser ces courbes. *L'artefact* qui apparait de façon très nette à la majorité se retrouverait étalé aux âges alentours et, aux grands âges, le taux de personnes en ménages commencerait à diminuer plus tardivement. Or l'utilisation des résultats des enquêtes annuelles de recensement selon l'âge au recensement (graphique 15) montrent une certaine stabilité de la part de personnes vivant en ménages selon l'âge, traduisant une

réalité sociodémographique structurelle qui évolue lentement. De ce fait, utiliser le cumul sur cinq ans « par âge » est préférable pour estimer la part des personnes en ménages à appliquer aux effectifs de l'âge considéré calculés par ailleurs, plutôt que de calculer la pyramide des âges par génération en distinguant les personnes en ménages des autres (cette solution supposerait par ailleurs de poser des hypothèses sur la situation des personnes décédées pour la correction des décès).

Graphique 15

Part des Femmes vivant en ménages, selon leur âge au recensement
Résultats des enquêtes annuelles avec pondération annuelle, le cumul avec pondération de cumul,
France métropolitaine



Champ : France métropolitaine.

La même démarche peut être appliquée pour calculer une pyramide des âges nationale selon la situation matrimoniale utilisée pour le calcul de certains indicateurs démographiques comme le taux de nuptialité selon l'état matrimonial antérieur ou le taux de primo-nuptialité.

Bibliographie

Bertrand P., Chauvet G., Christian B. Grosbras J.-M., « Les plans de sondage du nouveau recensement », *Actes des VIIIe journées de méthodologie statistiques 2002*, INSEE.

Bertrand P., Chauvet G., Christian B. Grosbras J.-M., « Données produites par le recensement rénové de la population », *Actes des VIIIe journées de méthodologie statistiques 2002*, INSEE.

Desplanques G., « Avantages et incertitudes des enquêtes annuelles de recensement », *Population*, volume 63, n°3, 2008, INED.

Dumais J., Isnard M. « Le sondage de logements dans les grandes communes dans le cadre du recensement rénové de la population (RRP) », *Actes des VIIe journées de méthodologie statistiques 2000*, INSEE.

Dumais J., Bertrand P., Kauffmann B., « Sondage, estimation et précision dans la rénovation du recensement de la population », *Actes des VIIe journées de méthodologie statistiques 2000*, INSEE.
« Pour comprendre le recensement de la population », *Insee Méthodes*, hors série, mai 2005.

Jugnot S. « Comment utiliser un recensement collecté sur cinq années successives pour établir la pyramide des âges de référence à une date donnée ? », *Actes des Xe journées de méthodologie statistiques 2002*, INSEE.

« Recensement de la population : détermination de la population légale des communes », note d'information datée du 16 décembre 2009, mise en ligne sur www.insee.fr

ANNEXE 1 - Note de Guy Desplanques du 12 juin 2007 sur « Pyramide d'âges et recensement »

Direction des Statistiques Démographiques et Sociales
Département de la Démographie

NOTE à l'attention de
Stéfan Lollivier
Directeur de la DSDS

Dossier suivi par :
Guy DESPLANQUES
Tél. : 01 41 17 53 86
Fax : 01.41.17.62.79.
Mél. : Guy DESPLANQUES

Paris, le 12 juin 2007
N° 1445/DG75-F101

Objet : **Pyramide d'âges et recensement**

La synthèse de cinq EAR va produire une pyramide d'âges lissée, sous forme d'une moyenne mobile sur cinq ans. Compte tenu des variations annuelles des naissances, l'effectif à un âge donné pourra être de ce fait assez différent de l'effectif réel. Ainsi, si l'on calcule une pyramide lissée au 1er janvier 2005 à partir des pyramides 2003 à 2007 issues du bilan démographique 2006, en se limitant aux générations dont l'effectif est supérieur à 50000 (93 ans ou moins), on trouve pour 12 générations un écart dépassant 5% (40000 pour une génération d'environ 800000) et pour 29 générations un écart supérieur à 2% (16000). Même sur les groupes d'âge quinquennaux, l'écart dépasse plusieurs fois 2% (voir annexe 1).

Ce qu'on mesure avec une pyramide lissée

En gros, l'effectif d'un âge donné issu des résultats détaillés est la moyenne pondérée des effectifs des cinq collectes successives. C'est ce qui produit l'effet de lissage. D'autre part, comme le recensement a lieu vers le 15 janvier, la population d'une génération diffère légèrement entre le recensement et le 1er janvier.

Les raisons de produire une pyramide « exacte »

Une pyramide lissée présente l'avantage de limiter les effets d'une mesure incertaine des niveaux absolus et de préserver les ordres de grandeur. La forme globale de la pyramide d'âges n'est pas altérée par le lissage. De la même façon, les indicateurs synthétiques de fécondité ou de mortalité (indicateur conjoncturel de fécondité, âge moyen à la maternité, espérance de vie) sont peu percutés par un lissage : il y a une certaine compensation entre différents âges et la précision des effectifs globaux utilisés dans le calcul est certainement inférieure au léger biais qui résulte du lissage. Ainsi, pour 2004, les deux calculs de l'indicateur conjoncturel de fécondité, l'un basé sur la population lissée, l'autre sur la population « exacte », différent d'environ 0,012 (pour un niveau de 1,9), soit 0,7%. Pour l'espérance de vie féminine, on trouve un écart de 0,1 an.

Mais l'intérêt d'une pyramide d'âges réside d'abord dans la mise en évidence des ruptures entre certaines générations, comme celle qui sépare les générations 1945 et 1946.

Dans l'évolution démographique, on distingue le solde naturel et le solde migratoire. Cette décomposition présente un intérêt à la fois globalement et par sexe et âge. Si on utilise des pyramides lissées, la différence des effectifs d'une même génération à deux 1er janvier successif ne sera plus la différence entre le solde migratoire pour cette génération et le nombre de décès dans cette

génération. La décomposition du solde migratoire par âge deviendra une boîte noire, dont les données seront très difficiles à interpréter.

Une autre raison qui justifie l'élaboration d'une pyramide d'âges « exacte » tient aux biais de collecte du recensement. A l'évidence, le nouveau recensement présente certains biais très liés à l'âge. Ainsi, l'effectif des personnes de 19 ans est très surestimé du fait de doubles comptes. Si l'expérience montre que ces biais sont difficiles à corriger et si on se cale sans correction sur les EAR, la pyramide d'âges construite sera erronée, indépendamment du lissage. Comme il y a peu de décès vers 18 ou 19 ans, le solde déduit des effectifs d'une même génération deux années successives sera affecté par le biais. Avec le lissage, ce solde sera étalé sur plusieurs âges et lui-même lissé. Le biais sera moins apparent, mais visible. Bien sûr, pour une tranche plus large, 15-24 ans par exemple, cela sera transparent, mais une telle tranche d'âges est très large pour certaines analyses.

Les conséquences du lissage sont d'autant plus importantes que les nombres de naissances varient fortement d'une année sur l'autre. Entre 2005 et 2006, le nombre de naissances a augmenté d'environ 30000. C'était également le cas entre 1999 et 2000. Nous ne sommes pas assurés que de telles variations, déjà fortes, ne seront pas dépassées.

Élaboration d'une pyramide « exacte »

Nous n'examinons pas ici la possibilité d'obtenir directement des effectifs « exacts » par détermination de poids individuels qui assureraient un calage sur une pyramide d'âges. Nous proposons plutôt de produire une pyramide exacte, qui ne fasse pas partie des résultats du recensement (même si, bien sûr, elle utilise les données des EAR).

Il est relativement aisé de mettre au point une méthode pour éviter les effets du lissage. Il suffit de reprendre les données par génération (voir annexe 2). Il est probablement plus difficile d'éliminer les biais liés à la méthode de recensement et aux problèmes de collecte, qui supposent que ces biais sont identifiés et quantifiés. En outre, le lissage ne modifie pas le total, alors que la prise en compte des biais est susceptible de le modifier.
Le département de la démographie se propose donc de mettre au point une méthode permettant d'estimer une pyramide d'âges qui ne soit pas lissée. Cette méthode utilisera à la fois les données des EAR et les données de l'état civil. Les problèmes qui découlent de l'absence de traitement des biais connus seront explicités.

Les effectifs par sexe et âge seraient diffusés et utilisés comme dénominateurs dans les calculs des indicateurs démographiques.

Problèmes connexes

L'élaboration d'une pyramide « exacte » au niveau national soulève plusieurs questions :

- Faut-il calculer et, éventuellement, diffuser, des pyramides « exactes » pour d'autres niveaux géographiques ? Lesquels ?
- A-t-on besoin d'autres répartitions non lissées par âge détaillé, par exemple faut-il avoir une répartition « exacte » par âge détaillé des actifs ? Comment la déterminer ?
- Y a-t-il d'autres variables que l'âge posant le même type de problème ?

<u>Jusqu'à quel niveau géographique calculer des pyramides « exactes » ?</u>
Si la proposition d'élaborer une pyramide d'âges « exacte » au niveau national est retenue, il faut certainement construire ces pyramides pour la France métropolitaine et pour la France incluant les DOM.

Actuellement, le DAR diffuse des populations départementales par sexe et âge quinquennal et, à certains utilisateurs, par âge détaillé. D'autre part, le département de la démographie calcule régulièrement des indicateurs démographiques par région, département et pour d'autres zonages supracommunaux. D'autre part, les projections démographiques s'appuient sur des structures par âge détaillé. L'intérêt de ces structures est moindre qu'au niveau national. Compte tenu de l''inconvénient d'avoir deux structures par âge, l'une « exacte », l'autre lissée, ce choix ne s'impose peut-être pas, d'autant que la correction des biais de collecte est malaisée à un niveau géographique détaillé.

A-t-on besoin d'une répartition exacte des actifs par âge détaillé ou d'autres répartitions « exactes » ?
Pour illustrer la question, on peut calculer l'évolution du nombre d'actifs qui résulterait de populations lissées, dans l'hypothèse de taux d'activité par âge constants. A titre indicatif, l'écart entre les nombres d'actifs calculé avec les taux d'activité par sexe et âge basés sur la population lissée et sur la population réelle est d'environ 1 pour 1000. Il pourrait être supérieur si les effectifs des générations où les taux évoluent fortement (en début et en fin de vie active) étaient très variables d'une génération à la suivante, ce qui n'est pas le cas vers 2005.

Y a-t-il d'autres variables que l'âge posant le même type de problème ?
A priori, pour toutes les questions s'appuyant sur des dates, la technique d'élaboration des données détaillées, en assurant un lissage, est susceptible d'altérer fortement la réalité en cas de pic conjoncturel important. Ce peut être le cas pour les données par année d'arrivée ou par année d'emménagement dans le logement.

Annexe 1

Ecart entre population lissée et population réelle, en 2005

age	écart relatif 2005	age	écart relatif 2005	age	écart relatif 2005	age	écart relatif 2005	age	écart relatif 2005
0	0,77	25	1,17	50	-0,56	75	1,06	0-4	-0,36
1	0,38	26	1,89	51	0,98	76	0,40	5-9	0,22
2	0,63	27	-0,52	52	-0,10	77	0,89	10-14	0,41
3	-1,07	28	3,23	53	1,49	78	-0,47	15-19	-0,39
4	-2,54	29	3,23	54	-1,55	79	-0,60	20-24	-0,57
5	0,04	30	-0,06	55	-0,72	80	0,58	25-29	1,81
6	1,38	31	-2,38	56	-1,33	81	1,20	30-34	-1,24
7	1,49	32	-2,55	57	-6,28	82	1,28	35-39	0,40
8	-1,08	33	-1,78	58	**-8,18**	83	**-9,91**	40-44	-0,25
9	-0,78	34	0,58	59	**11,62**	84	**-21,21**	45-49	0,04
10	1,69	35	0,32	60	4,13	85	**14,36**	50-54	0,05
11	2,78	36	1,18	61	-3,44	86	**14,38**	55-59	-1,22
12	-0,70	37	1,38	62	0,47	87	**10,00**	60-64	2,34
13	-1,16	38	-0,79	63	**8,06**	88	**21,55**	65-69	-0,57
14	-0,43	39	-0,06	64	2,59	89	**12,64**	70-74	-0,80
15	-0,38	40	-1,96	65	-3,76	90	**-21,28**	75-79	0,29
16	-0,06	41	-0,94	66	-0,59	91	**-8,56**	80-84	**-4,38**
17	0,09	42	1,80	67	0,52	92	-1,96	85-89	**14,51**
18	-1,33	43	0,00	68	-0,21	Total	-0,02	90-94	**-7,70**
19	-0,26	44	-0,11	69	1,15			95-99	0,56
20	1,09	45	-0,95	70	-1,44			100 ou p	-1,07
21	3,48	46	0,95	71	1,72			Total	-0,02
22	-1,43	47	-0,17	72	-1,21				
23	-2,85	48	0,10	73	-1,08				
24	-3,09	49	0,25	74	-2,04				

Source : bilan démographique 2006.

Annexe 2

Méthodes d'estimation de la population exacte d'une génération

Avec les seules données des EAR, deux méthodes peuvent être utilisées pour obtenir une structure par âge détaillé : soit faire la moyenne sur 5 années des populations d'un âge donné, soit faire la moyenne dans une génération donnée.

Moyenne par âge

Dans le 1er cas, la moyenne obtenue, en combinant 5 générations successives, est très sensible aux variations annuelles de la natalité. S'il y a un biais de collecte pour certains âges, ce biais subsiste avec la même ampleur dans le calcul. On obtient une population lissée. Sur ces populations, l'application de l'équation de la dynamique n'est plus applicable, sauf à moyenner aussi les autres éléments de l'évolution démographique : naissances et décès. Si on lisse les populations sans lisser les décès par exemple, on estime par différence des soldes migratoires très élevés en valeur absolue aux âges élevés, à des âges où les migrations avec l'extérieur sont rares.

Moyenne par génération

Avec la moyenne par génération, la moyenne obtenue reflète bien les variations de la natalité au fil des années. S'il y a un biais de collecte pour certains âges, ce biais est lissé sur 5 années d'âge dans le calcul. Par exemple, le biais observé à 19 ans va être étalé sur la tranche 17-21 ans. Le profil par âge des décès et du solde migratoire peut également provoquer des biais. Par exemple, pour la mortalité, la moyenne au sein d'une génération tend à sous-estimer aux âges où la courbe de survie présente une concavité vers le bas et à surestimer fortement les effectifs aux âges élevés, aux âges où la courbe de survie présente une concavité vers le haut, car le nombre de décès d'un âge au suivant est fortement croissant avant d'être décroissant. En ce qui concerne le solde migratoire, les variations avec l'âge sont fortes : le solde peut être légèrement négatif ou très faiblement positif avant 18 ans et devenir très fortement positif entre 20 et 25 ans. La moyenne dans une génération va conduire à lisser cet effet d'âge.

Prise en compte de la mortalité

La mortalité étant assez stable d'une année sur l'autre, on peut corriger les problèmes dus au lissage aux âges élevés en effectuant le lissage avec prise en compte des décès. Autrement dit, on ne lisserait pas directement les populations d'une même génération à différentes dates, mais des populations dont on déduit les décès (années avant la date de référence ou non compris les décès (années après la date de référence). Par exemple, pour la génération 1920, au lieu de prendre la population en 2004, on prend une population obtenue en enlevant les décès en 2004 et 2005 dans cette génération. De même la population en 2007 est remplacée par la population recensée à laquelle on ajoute les décès survenus en 2006 dans la même génération. Cette correction n'est utile qu'aux âges où la mortalité est importante et est de l'ordre de grandeur des migrations au même âge ou plus forte.

Il est plus difficile de corriger le biais lié aux migrations puisque, si les données administratives ou les données du recensement sur l'année d'entrée donnent une idée de la répartition par âge des migrants, aucune information directe n'est disponible sur la répartition des sortants. La structure par âge du solde migratoire est donc assez mal connue. De ce fait, le lissage introduit de facto par une moyenne par génération n'est probablement pas trop gênant pour l'estimation de ce solde par âge.

Au total, une population par âge obtenue par lissage des effectifs d'une même génération fournirait une bonne estimation de la population exacte de la génération, à condition d'introduire une prise en compte des décès aux âges élevés. Une simulation effectuée à partir des projections de population, qui, par construction, minimisent les aléas de mesure, montre que la prise en compte des décès est justifiée à partir d'un âge voisin de 40-44 ans.

Reste une difficulté aux âges les plus jeunes : à 2 ans ou avant. A 2 ans par exemple, le lissage dans une génération devrait surestimer légèrement l'effectif à cause des décès qui surviennent à 0 ans (dans l'année suivant la naissance). La population à 0 ou 1 an pose davantage de problèmes, puisque ces deux générations n'ont pu être observées dans les deux années précédant la date de référence. On peut donc envisager une moyenne lissée sur 3 années (l'année de référence et les deux suivantes), mais il est très possible de corriger de la mortalité survenue avant. Reste à examiner la

possibilité d'une correction liée à la mobilité. En effet, classiquement, le nombre de naissances domiciliées d'une année est bien supérieur à la population de la même génération recensée peu après, même si on tient compte de la mortalité.

Le Chef du Département de la Démographie

signé : G. DESPLANQUES

Copie :
S Jugnot, O Lefebvre, chefs de département de la DSDS

ANNEXE 2 - Propositions du groupe de travail sur le calcul de la pyramide des âges de référence pour les estimations de population

[Extrait du rapport présentant les travaux du groupe de travail « Pyramide des âges et indicateurs démographiques », en date du 23 février 2009]

A l'issue de ses travaux, le groupe de travail fait les propositions suivantes[21] :

P1) Au niveau national, pour le calcul des indicateurs démographiques (taux de fécondité, quotients de mortalité, espérance de vie, taux de nuptialité, etc.), le groupe de travail recommande de ne pas utiliser directement la pyramide des âges obtenue par tabulation simple des résultats définitifs du recensement. Il convient de privilégier une pyramide des âges retravaillée, « par génération », construite à partir des informations collectées dans les cinq enquête annuelles utilisées pour établir les populations légales.

Cette pyramide serait établie à partir des effectifs par sexe et année de naissance calculés avec les résultats (pondérations) des enquêtes annuelles, corrigés des décès par sexe et année de naissance intervenus sur la période : la pyramide des âges ainsi construite sera relative à la même date de référence que les populations légales, soit le 1er janvier de l'année médiane ; les décès intervenus les deux premières années du cycle seraient supprimés des effectifs comptabilisés dans les deux premières enquêtes annuelles ; *a contrario*, les décès intervenus dans les deux dernières années du cycle seraient réintroduits dans les effectifs comptabilisés dans les deux dernières enquêtes annuelles.

Cette pyramide donnerait une structure par sexe et âge qui serait ensuite appliquée aux effectifs totaux tels qu'ils sont calculés dans la population légale.

P1 bis) Les informations collectées sur la structure de la population sont relatives, *de facto*, à la situation moyenne sur la période de collecte, donc datable de fin janvier ou début février. On pourrait donc envisager de ramener ces informations au 1er janvier, en supprimant les nés de janvier de l'année de collecte, en ajoutant les décès de janvier de l'année de collecte et, dans l'absolu, en ajoutant également un douzième du solde migratoire par sexe et âge mais ce point présente des difficultés méthodologiques. Cette option permettrait de disposer d'une pyramide théoriquement plus en cohérence avec les informations d'état civil. Elle ne serait pas cohérente avec le fait que les chiffres du recensement seront datés « 1er janvier 2006 ». De ce fait le groupe de travail ne tranche pas entre les deux options et propose l'alternative suivante :

- Soit, on se limite à supprimer les personnes nées en janvier de l'année de collecte (pas d'âge zéro) [option finalement retenue];

- Soit on cherche à se rapporter à une structure au 1er janvier en supprimant les nés de janvier de l'année de collecte et en ajoutant les décès de janvier de l'année de collecte. Compte tenu de la fragilité des estimations du solde migratoire, celui-ci ne serait pas pris en compte.

P1 ter) Concernant le pic artificiel observé autour de l'âge de 18 ans, en l'absence d'informations pour estimer l'ampleur de la correction qu'il faudrait apporter, le groupe retient ne rien faire aucun traitement particulier, le cumul par génération lissant déjà en partie le pic.

[21] Propositions validées avec la validation du compte-rendu de la dernière réunion du groupe de travail : compte-rendu n°2676/DG75-F170, du 13 novembre 2008, de la réunion du groupe de travail du 9 octobre 2008.

P2) Pour le calcul des indicateurs régionaux et départementaux diffusés de façon standard, le groupe de travail propose d'utiliser directement la pyramide des âges qui résulte de la tabulation simple, « par âge », des résultats définitifs du recensement.

En effet, il n'apparaît pas souhaitable de calculer des pyramides des âges corrigées de façons standard à tous les échelons géographiques. De tels calculs pourraient donner aux utilisateurs l'illusion d'une précision des résultats qui n'existe pas du fait des autres sources d'imprécisions, qui jouent plus fortement à un niveau fin qu'au niveau national (imprécision lié à l'échantillonnage, difficultés ponctuelles de collecte). Plus le niveau géographique est fin, plus les autres sources d'écarts par rapport au réel sont fortes, rendant moins essentielle la correction du biais induit par l'usage de la pyramide « par âge ».

Le choix retenu privilégie donc la cohérence entre les communes ou groupement de communes et les départements et régions. De fait, les indicateurs agrégés actuellement publiés apparaissent peu sensibles à la méthode retenue. La pyramide issue du cumul « par âge » issue du recensement peut donc être utilisée.

Cette proposition générale n'est pas exclusive. Dans le cadre des produits spécifiques, pour répondre à certains besoins particuliers, il peut être utile de revenir à une méthodologie analogue à celle développée au niveau national. Cet aspect concerne notamment la réponse aux besoins locaux d'effectifs sur des tranches d'âges cibles précises, comme des effectifs d'enfants en âge scolaire, le nombre de personnes âgées au niveau départemental ou l'analyse de l'attrait du territoire, par tranche d'âges ; besoins qui mobilisent souvent l'outil Omphale.

P3) La pyramide des âges nationale « par génération » décrite dans la proposition P1 étant essentielle pour les utilisateurs démographes, il conviendrait de la mettre à disposition sur www.insee.fr dans *La Situation Démographique*. Elle pourrait y être proposée en même temps que la pyramide des âges lissée qui résulte d'une tabulation directe du recensement « par âge » de façon à afficher clairement l'existence des deux pyramides. Il faudrait l'accompagner d'un message simple et clair pour expliquer que les deux pyramides donnent des résultats très proches mais qu'elles ne répondent pas aux mêmes usages, la première étant destinée avant tout aux démographes, la seconde étant suffisante pour la plupart des utilisations et cohérente avec les séries régionales, départementales, et les informations communales issues du recensement.

Dans le *Bilan démographique,* les estimations avancées des indicateurs démographiques utiliseront une pyramide calée sur la pyramide retravaillée « par génération ». Par construction, il n'y aura pas de pyramide du recensement pour le 1er janvier N-2, N-1 et N. Il faut donc réfléchir à la visibilité à donner aux pyramides des âges provisoires « par génération » N-2, N-1 et N au moment du *Bilan.*

Pour les tableaux de la structure par âge figurant dans les chiffres clés et actualisés au moment de la publication du *Bilan démographique*, les chiffres du recensement datés du 1er janvier N-3 pourront être utilisés. Il faut souligner ici l'importance de bien faire apparaître la source adéquate pour chaque donnée publiée : « Recensement » ou « Estimations de population ».

P4) Il convient de documenter dès que possible la méthode utilisée pour construire la pyramide des âges, par exemple en mettant à disposition sur www.insee.fr un document de travail sur le sujet. Il conviendrait également de documenter davantage la méthode de rétropolation ainsi que la méthode utilisée chaque année pour construire les estimations de population par sexe et âge.

Cette documentation permettra par ailleurs aux utilisateurs experts qui souhaiteraient réaliser une pyramide « exacte » selon les mêmes principes que la pyramide nationale, mais à un autre échelon géographique, de le faire, les données d'état civil étant disponible par commune. Il conviendrait toutefois pour cela de s'assurer également que la date de collecte figure dans les fichiers détaillés du recensement qui seront diffusés. Aussi, le groupe rappelle l'importance d'en disposer dans les fichiers détail destinés aux chercheurs. Pour les autres utilisateurs experts, potentiellement intéressés par l'information sur la date de collecte, comme les agences d'urbanisme, la disponibilité de cette information dépend de la politique de diffusion de la DDAR et n'entre donc pas dans le champ du groupe de travail.

P5) Compte tenu des effets limités sur les indicateurs démographiques d'une correction du point de 1999 et des implications importantes qu'elle aurait pour l'INSEE, il n'appartient pas au groupe d'établir une préconisation ou une proposition sur ce point. Il serait toutefois important de mieux documenter l'introduction des ajustements.

Les questions méthodologiques posées par le point 1999 pourraient ainsi être utilement abordée dans le cadre d'un document plus général qui portera sur l'ensemble de la révision des séries démographiques pour présenter les rétropolations.

Par ailleurs, il conviendrait d'introduire systématiquement une colonne « ajustement » dans les tableaux qui présentent l'évolution annuelle de population et la décomposent entre le solde migratoire et le solde naturel. Actuellement, ce n'est le cas que dans un tableau du *Bilan démographique*. Cette introduction serait cohérente avec les tableaux diffusés par Eurostat. Elle faciliterait la compréhension de la notion de « solde migratoire apparent » aux échelons infra nationaux.

Seul un ajustement global serait affiché, sans décomposition, ni selon le sexe et l'âge, ni par région ou département. Il pourrait être utile d'envisager à terme un travail méthodologique sur la répartition de l'ajustement, pour montrer l'incidence sur les résultats des hypothèses faites (sur la répartition par région notamment).

P6) Il conviendra d'expertiser régulièrement les résultats issus du recensement pour s'assurer que les différents risques de biais restent négligeables. Il conviendra également d'être particulièrement vigilant lors de l'établissement de la pyramide définitive relative au 1er janvier 2007, qui permettra la première comparaison entre deux résultats consécutifs du nouveau recensement. Sur ce point, il est essentiel de poursuivre en 2009 les travaux sur les conséquences de la méthodologie retenue pour l'étude des évolutions annuelles.

P7) La pyramide des âges selon le statut matrimonial, nécessaire pour le calcul de certains indicateurs sera obtenu en appliquant à la pyramide des âges retravaillée « par génération », la répartition par sexe et âge issu de la tabulation directe du recensement par âge, le phénomène étudié étant plus lié à l'âge qu'à la génération.

Pour la même raison, si une demande de pyramide des âges des personnes vivant en ménages est faite pour caler, comme actuellement, l'enquête Emploi en continu, la même démarche sera effectuée (application des taux de ménages par sexe et âge dans le recensement à la pyramide « par génération »).

ANNEXE 3 - Le cas des pyramides des âges régionales et départementales

Cette annexe présente le travail réalisé par Anne Thérèse Aerts et Noëlle Serruys dans le cadre du groupe de travail « pyramide des âges et indicateurs démographiques ». Ce travail a été réalisé sur l'ensemble des régions. L'exemple de l'Ile-de-France est présenté ici à titre illustratif.

Le département de l'action régionale de l'Insee calcule et publie des indicateurs démographiques départementaux et régionaux, ainsi qu'une répartition de population par tranches d'âge à ces échelons. Par souci de cohérence, la question de la déclinaison de la méthode nationale à ces échelons a été examinée.

Pour cela, les travaux ont porté sur les pyramides régionales qui ont été estimées à partir du cumul des résultats des enquêtes annuelles de recensement (données pondérées par les poids du cumul) selon les différentes approches possibles (par âge, par génération, par génération avec correction des décès). Il n'a pas été envisagé de construire des pyramides à partir d'une moyenne simple des enquêtes annuelles de recensement (données pondérées par les poids annuels), les estimations de populations à partir de chaque enquête annuelle n'étant pas jugées aussi fiables qu'au niveau national pour les régions et surtout les départements : l'échantillon des enquêtes annuelles n'a pas été construit pour disposer de résultats départementaux.

Au moment de l'étude, on disposait de quatre années d'enquêtes annuelles du nouveau recensement. De façon à estimer une pyramide datée au 1er janvier, les pyramides ont été construites par tabulation des résultats des trois premières enquêtes (EAR 2004 à 2006), elles sont donc datées au 1er janvier 2005, ce qui facilite les comparaisons avec les indicateurs publiés.

Illustration des effets sur les structures par âge par région

Pour illustrer l'impact du choix de la méthode utilisée sur la structure par âge, quatre pyramides ont été comparées pour chaque région, pour les hommes et les femmes séparément :

- la pyramide par âge estimée à partir du cumul des enquêtes annuelles 2004 à 2006,
- la pyramide par génération estimée à partir du cumul des enquêtes annuelles 2004 à 2006,
- la pyramide par génération estimée à partir du cumul des enquêtes annuelles 2004 à 2006 et corrigée des décès,
- la pyramide au 1er janvier 2005 diffusée en janvier 2008.

Pour la pyramide estimée en cumul par génération et corrigée des décès, le principe de la correction à partir des statistiques d'état-civil a été le suivant : pour une année de naissance, les décès ayant eu lieu entre 2004 et 2005 ont été déduits de l'estimation de population du cumul ; inversement, les décès ayant eu lieu entre 2005 et 2006 ont été ajoutés à l'estimation du cumul.

Les constats généraux que l'on peut tirer de l'observation de ces quatre pyramides sont globalement les mêmes pour toutes les régions et sont illustrés ci-dessous par l'exemple de la région Île-de-France, au travers des deux graphiques qui suivent.

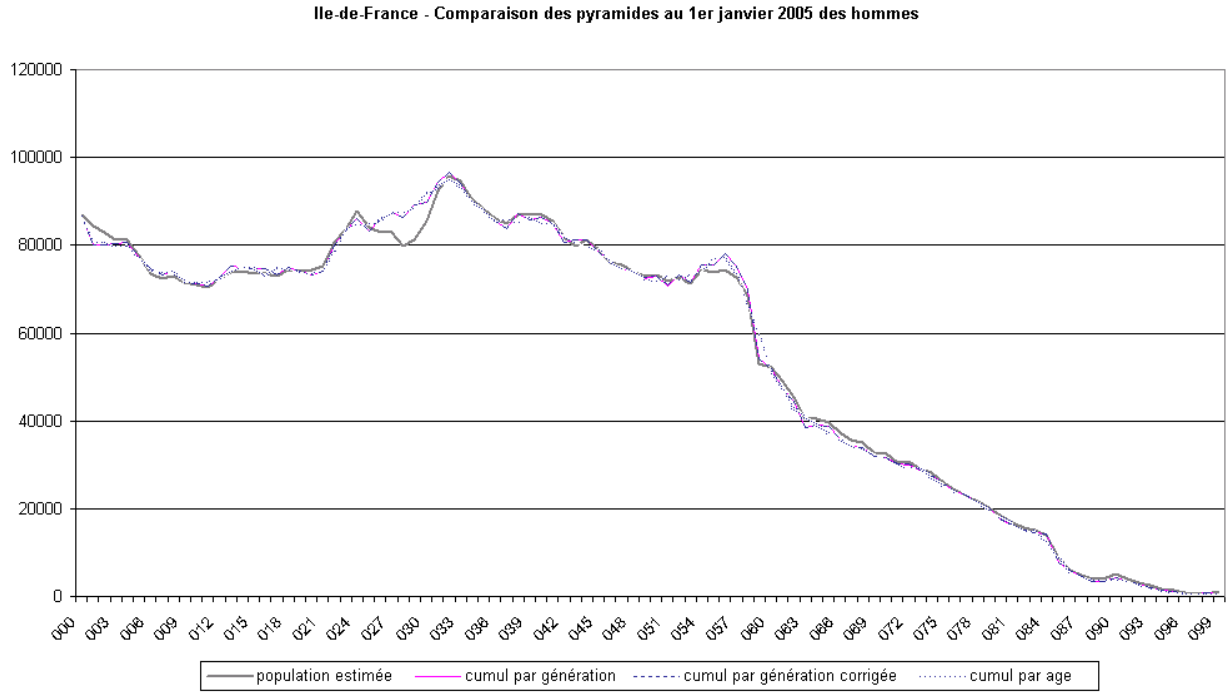
La comparaison des quatre pyramides montre qu'aussi bien pour les femmes que pour les hommes :

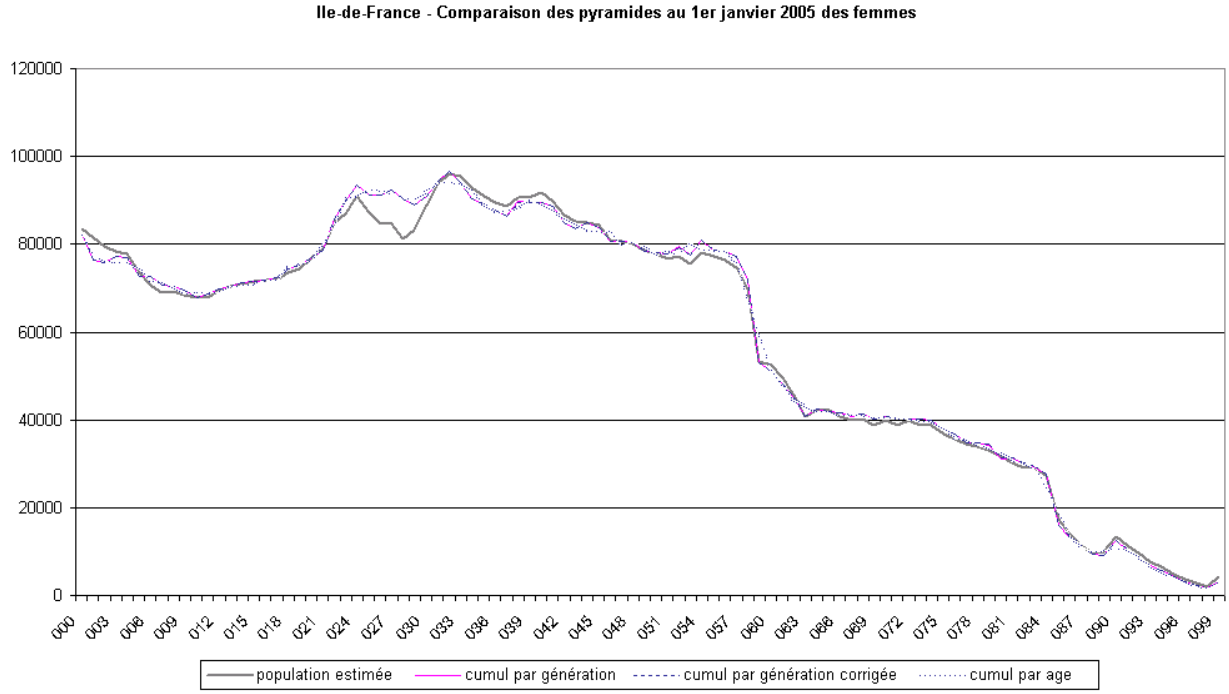
- Les pyramides sont généralement proches quelle que soit la méthode d'estimation retenue[22] (par âge, par génération, ou par génération avec correction des décès). L'estimation par génération donne

[22] Pour le cas particulier de la région Corse, la population estimée est peu importante, de ce fait les écarts entre les différentes méthodes d'estimations sont plus marqués et les pyramides obtenues sont également plus heurtées.

une structure par âge de la population un peu plus heurtée que l'estimation par âge qui tend à lisser les effectifs ; le fait de corriger ou non les décès modifie peu les résultats en niveau.

Graphiques 1a et 1b





- Il y a plus de différences entre les pyramides construites à partir des enquêtes annuelles de recensement et la pyramide au 1er janvier 2005 diffusée en janvier 2008. Les écarts peuvent être importants pour certaines tranches d'âges, le plus souvent les jeunes. Ces différences ne sont pas aussi visibles dans toutes les régions, elles sont particulièrement nettes dans certaines régions comme l'Île-de-France où pour les jeunes entre 25 et 30 ans, la population au 1er janvier 2005 diffusée en janvier 2008 est nettement sous-estimée par rapport à celle que l'on observe par tabulation des enquêtes annuelles du nouveau recensement.

En d'autres termes, les biais qu'il y a à utiliser des pyramides régionales et départementales basées sur un recensement général ancien, sont plus importants que ceux que l'on pourrait avoir en utilisant une méthode plutôt qu'une autre pour construire la pyramide des âges à partir des résultats de cumul des enquêtes annuelles de recensement.

Pour mieux comprendre ces différences, il faut revenir à la méthode utilisée actuellement pour prolonger les pyramides des âges régionales et départementales observées entre deux recensements.

Au cours d'une année donnée, l'évolution de la population d'une zone géographique résulte de deux facteurs :

- le solde naturel, différence au cours de l'année entre le nombre de naissances et le nombre de décès domiciliés dans la zone géographique ;
- le solde migratoire, différence, au cours de la même année, entre le nombre de personnes venues résider dans la zone (les entrants) et le nombre de personnes qui l'ont quittée pour résider ailleurs (les sortants).

Chaque année, le solde naturel des régions et des départements est déterminé grâce aux statistiques exhaustives de l'état civil. En revanche, le solde migratoire ne peut être qu'estimé de façon fragile puisque les flux migratoires échappent à toute procédure administrative d'enregistrement.

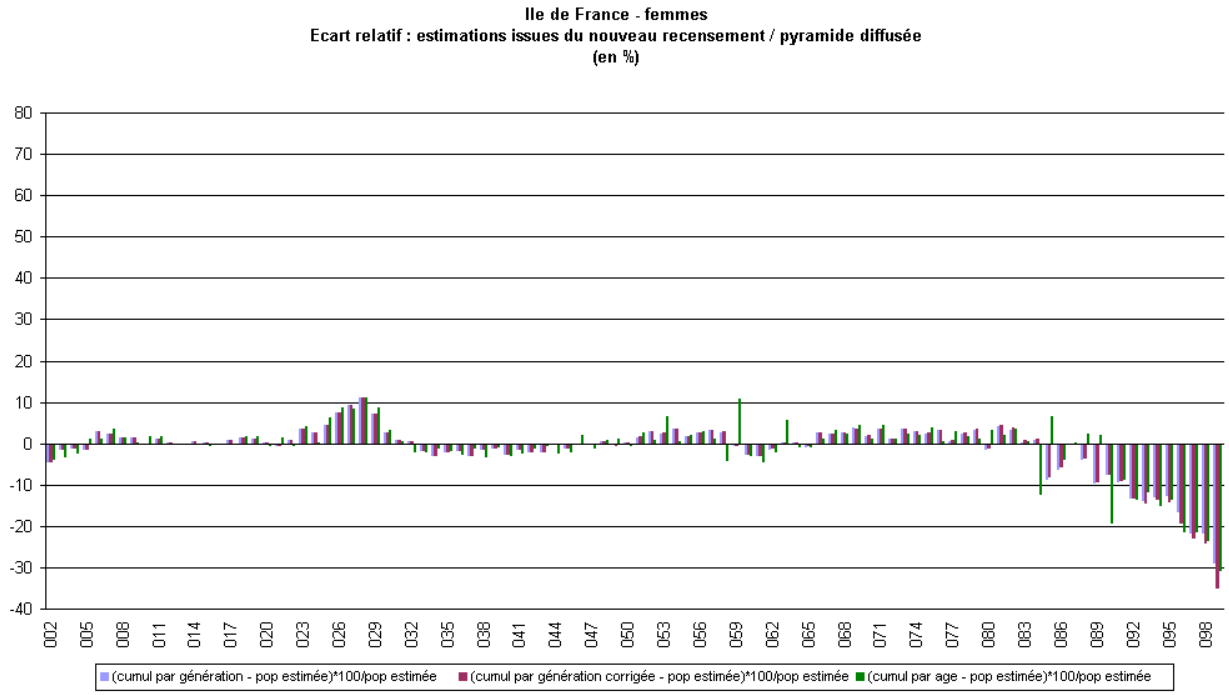
Sur la période 1999-2006, le taux de solde migratoire annuel (rapport du solde migratoire au cours de l'année sur la population en début d'année) a donc été estimé par prolongement du taux de solde migratoire annuel moyen observé entre les recensements de 1990 et 1999 : ce taux de solde migratoire annuel moyen sur 1990-1999 ayant été lui-même calculé pour les régions et les départements avec l'outil Omphale, construit pour proposer des projections de populations au niveau local.

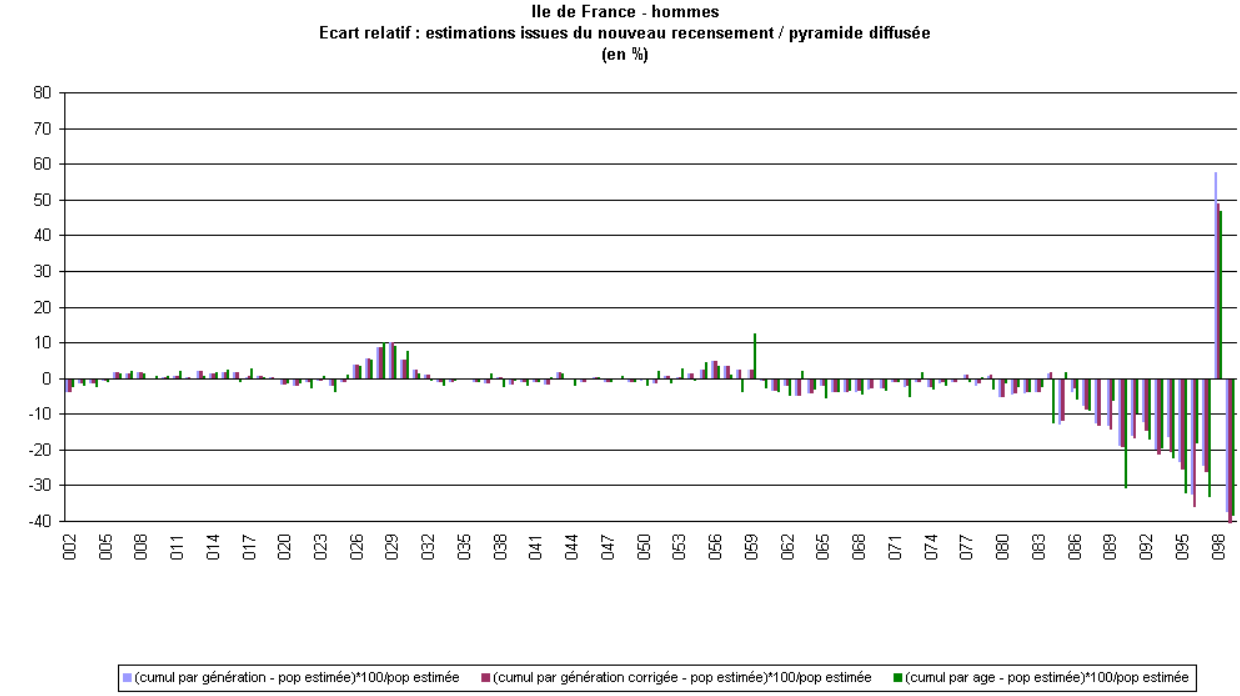
Les estimations de pyramides des âges qui ont été réalisées jusqu'à présent prolongent donc les phénomènes migratoires observés sur la période 1990-1999 et ne rendent pas compte des tendances migratoires plus récentes, postérieures à 1999, au contraire des pyramides des âges qui peuvent être tabulées à partir des enquêtes annuelles du nouveau recensement. Les différences sont en particulier visibles chez les jeunes qui sont les plus mobiles.

Par ailleurs, les taux de solde migratoires annuels moyens estimés dans l'outil Omphale sont par construction « lissés » : les bosses et les creux sont aplatis, de manière à disposer de quotients par âge qui soient pas trop erratiques, l'objectif étant de pouvoir réaliser ensuite des projections de population à des niveaux géographiques fins. Il est donc aussi possible que les estimations actuelles de population aient été un peu « lissées » par rapport aux évolutions réelles.

La comparaison des écarts relatifs entre la population estimée à partir du cumul des enquêtes annuelles de recensement (selon les différentes méthodes) et celle estimée selon la méthode actuelle confirme ces constats. Elle met en évidence des écarts non négligeables en valeur relative pour d'autres tranches d'âges, comme l'illustre l'exemple de l'Île-de-France (graphique 2a et 2b). Le constat vaut sur l'ensemble des régions.

Graphiques 2a et 2b



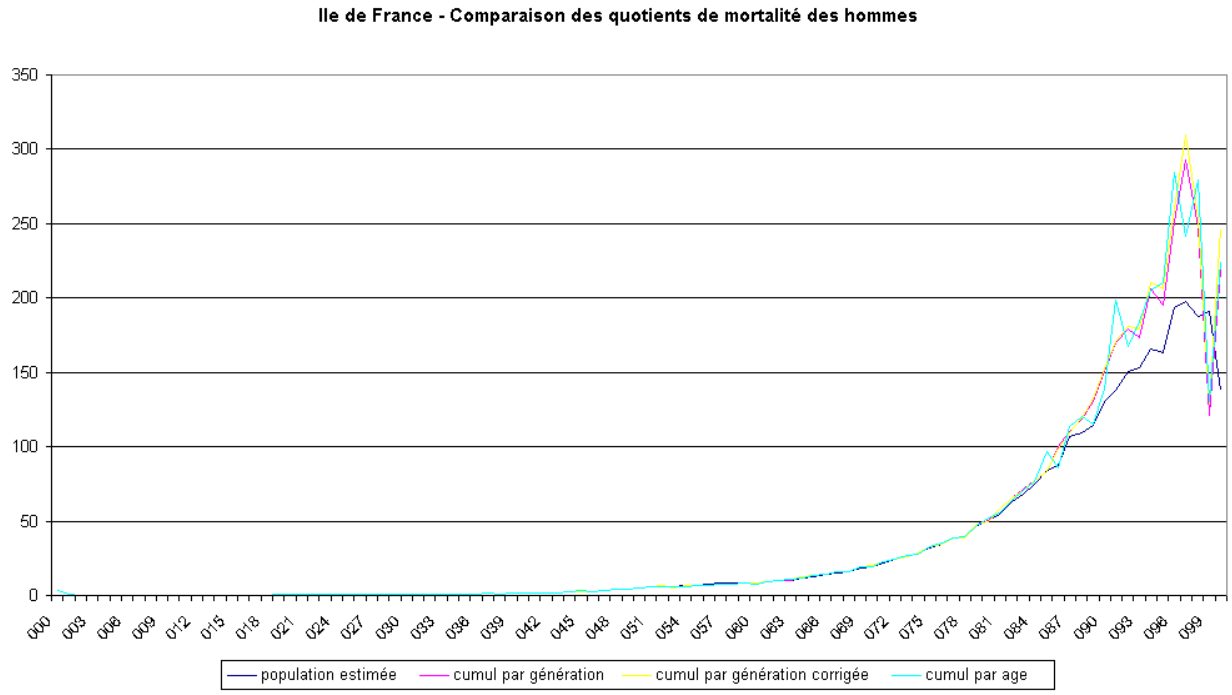


L'impact du choix de la méthode utilisée sur les principaux indicateurs démographiques a également été étudié au niveau régional, pour les quotients de mortalité, l'espérance de vie, le taux de fécondité part âge et l'indicateur conjoncturel de fécondité. Les principaux résultats présentés ici seront illustrés avec le cas de l'Île-de-France mais l'ensemble des régions donne des résultats similaires.

Impact sur les quotients de mortalité et les espérances de vie (à la naissance, à 1 an, 60 ans et 75 ans)

Les quotients de mortalité ont été calculés en rapportant le nombre de décès aux estimations de populations obtenues selon les différentes méthodes : par âge, par génération avec ou sans correction des décès, à partir des données en cumul du nouveau recensement ; et estimations selon la méthode actuelle. L'espérance de vie se déduit ensuite selon la formule habituelle.

Graphique 3



De façon générale, les quotients de mortalité en niveau sont peu impactés par la méthode d'estimation. Les divergences observées tiennent plus à des effets de taille (effectifs faibles aux grands âges) ou au fait que l'estimation en cumul par âge donne une population plus « lissée » par rapport aux autres estimations.

Sur la courbe des quotients de mortalité par âge, cela conduit à une estimation un peu plus heurtée de l'indicateur en cumul par âge. Les décès constatés aux âges élevés sont en effet directement en rapport avec l'effectif de la génération : ils sont moins importants si l'effectif de la génération est faible et vice-versa. Ainsi lorsqu'on rapporte les décès à la population de la génération, la courbe est relativement régulière, alors que lorsqu'on rapporte les décès à une population lissée sur plusieurs générations (estimation en cumul par âge), on fait apparaître de manière factice des risques de décès relativement plus faibles (ou plus élevés) en rapport à la population moyenne pour certains âges bien spécifiques.

Enfin, la qualité des fichiers d'état civil diminuant dans les grands âges, les données sur les décès et donc sur les quotients de mortalité ne sont pas considérées comme très fiables aux âges très élevés (approximativement : 98 ans pour les femmes et 94 ans pour les hommes). A titre indicatif, pour le calcul de l'espérance de vie au niveau régional et départemental, on se limite à 97 ans pour le calcul selon une formule exacte, ensuite on attribue une valeur approchée.

Les quotients de mortalité par âge ne sont pas publiés au niveau régional et départemental. Ce sont les indicateurs agrégés d'espérance de vie qui le sont. Ils sont très stables, quelle que soit la méthode d'estimation de population retenue utilisant les enquêtes annuelles de recensement. Ce constat vaut aussi bien pour l'espérance de vie à la naissance ou à un an que pour l'espérance de vie à 60 ou 75 ans.

Ile-de-France
Espérance de vie à la naissance, à un an, 20 ans, 40 ans , 60 ans et 75 ans

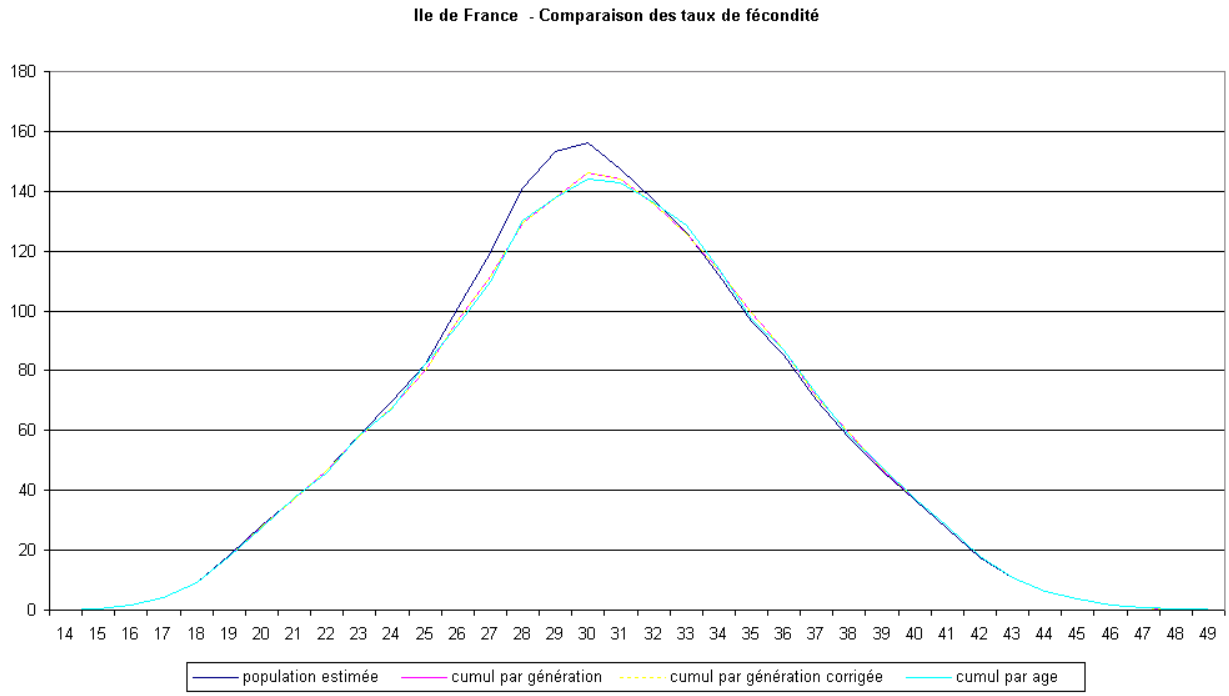
	Hommes					
Espérance de vie calculée	à la naissance	à un an	à 20 ans	à 40 ans	à 60 ans	à 75 ans
à partir de la population estimée	78,52	77,85	59,11	39,93	22,71	11,85
à partir du cumul par génération	78,22	77,55	58,81	39,62	22,34	11,49
à partir du cumul par génération corrigée	78,22	77,55	58,81	39,62	22,34	11,49
à partir du cumul par age	78,23	77,56	58,82	39,63	22,36	11,50

	Femmes					
Espérance de vie calculée	à la naissance	à un an	à 20 ans	à 40 ans	à 60 ans	à 75 ans
à partir de la population estimée	84,49	83,81	64,99	45,45	26,97	14,41
à partir du cumul par génération	84,40	83,72	64,91	45,36	26,86	14,24
à partir du cumul par génération corrigée	84,42	83,73	64,92	45,37	26,87	14,26
à partir du cumul par age	84,43	83,74	64,93	45,38	26,88	14,27

Impact sur le taux de fécondité et l'indicateur conjoncturel de fécondité

Les taux de fécondité ont été calculés en rapportant le nombre de naissance aux différentes estimations de populations : par âge, par génération avec ou sans correction des décès, à partir des données en cumul du nouveau recensement ; et estimations selon la méthode actuelle. Les constats généraux sont qualitativement les mêmes que pour les indicateurs de mortalité. Les différences sont surtout visibles aux âges entre 24 et 32 ans où les naissances sont les plus nombreuses (pour les indicateurs de mortalité, les écarts sont surtout visibles aux âges élevés où les décès deviennent quantitativement importants

Graphique 4



Généralement, l'indicateur en cumul par âge est un peu plus heurté que les indicateurs en cumul par génération (corrigés ou non des décès) en particulier pour les tranches d'âges 24-32 ans, en raison du phénomène de lissage de la population estimée en cumul par âge. Les estimations actuelles de la population conduit également à des taux de fécondité pour certaines tranches d'âge assez différents de ceux obtenus à partir du cumul des enquêtes annuelles de recensement. Ceci reflète les écarts déjà constatés sur les structures par âge : ces écarts, qui portent surtout sur des âges jeunes, sont logiquement visibles sur la courbe des taux de fécondité.

Les taux de fécondité par âge ne sont pas publiés au niveau régional et départemental. C'est l'indicateur conjoncturel de fécondité (ICF) qui l'est. Il est très stable quelle que soit la méthode d'estimation à partir des enquêtes annuelles de recensement.

Région		Indicateur conjoncturel de fécondité calculé à partir			
		de la population estimée	du cumul par génération	du cumul par génération corrigée	du cumul par age
11	Ile-de-France	2,01	1,96	1,96	1,96
21	Champagne-Ardenne	1,89	1,93	1,93	1,94
22	Picardie	2,03	2,03	2,03	2,03
23	Haute-Normandie	1,94	1,97	1,97	1,97
24	Centre	1,97	1,98	1,98	1,98
25	Basse-Normandie	1,93	1,96	1,96	1,96
26	Bourgogne	1,83	1,87	1,86	1,87
31	Nord-Pas-de-Calais	1,97	2,01	2,01	2,01
41	Lorraine	1,79	1,82	1,82	1,82
42	Alsace	1,79	1,78	1,79	1,78
43	Franche-Comté	2,00	2,02	2,02	2,02
52	Pays de la Loire	2,05	2,08	2,08	2,08
53	Bretagne	1,94	1,96	1,96	1,96
54	Poitou-Charente	1,85	1,88	1,88	1,88
72	Aquitaine	1,72	1,73	1,73	1,73
73	Midi-Pyrénées	1,76	1,75	1,75	1,75
74	Limousin	1,74	1,78	1,78	1,77
82	Rhône-Alpes	1,97	1,97	1,97	1,97
83	Auvergne	1,76	1,79	1,79	1,79
91	Languedoc-Roussillon	1,82	1,85	1,85	1,85
93	Provence-Alpes -Côte d'azur	1,88	1,91	1,91	1,91
94	Corse	1,65	1,66	1,66	1,66