

COMMENT UTILISER UN RECENSEMENT COLLECTÉ SUR CINQ ANNÉES SUCCESSIVES POUR ÉTABLIR LA PYRAMIDE DES ÂGES DE RÉFÉRENCE À UNE DATE DONNÉE ?¹

Stéphane JUGNOT (*)

(*) Insee, Division des enquêtes et études démographiques

Introduction

Le recensement sert habituellement de base à la comptabilité démographique. Il détermine le niveau de population de référence pour les séries démographiques. La structure par sexe et âge du recensement sert également de référence pour construire la pyramide des âges, nécessaire au calcul de la plupart des indicateurs démographiques : taux de fécondité, espérance de vie, etc.

Entre deux recensements, le niveau de population et la pyramide des âges sont actualisés annuellement à partir des statistiques d'état civil et d'une estimation du solde migratoire. Ces estimations sont d'abord provisoires car dès que les résultats d'un nouveau recensement sont disponibles, une révision des estimations annuelles est habituellement opérée de façon à caler le niveau de population et la pyramide des âges sur ces nouveaux résultats.

Jusqu'en 2008, les résultats des enquêtes annuelles de recensement n'ont pas été utilisés pour établir la structure selon le sexe et l'âge des estimations de population publiées. La disponibilité, fin 2008, des résultats définitifs du nouveau recensement a conduit à procéder à une révision des estimations réalisées depuis le précédent recensement de 1999. Ces premiers résultats définitifs ont pour date de référence le 1er janvier de l'année médiane du cycle de cinq ans, soit le 1er janvier 2006. Une rétopolation sera réalisée sur les pyramides des âges datées du 1er janvier 2000 au 1er janvier 2006.

Toutefois, 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 et il est 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 5 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 (par génération). Le nouveau recensement oblige donc à développer une méthodologie adaptée au nouveau mode de collecte du recensement.

Après avoir présenté rapidement le nouveau recensement (mode de collecte et estimation), nous présenterons les difficultés à résoudre pour élaborer la pyramide de référence et la solution retenue. Nous mettrons notamment l'accent sur trois aspects : le choix entre utiliser l'année de naissance ou l'âge au recensement, la nécessité d'utiliser également des informations externes (en l'occurrence les

¹ Ce papier présente des travaux réalisés avec l'aide Marie Anguis et Catherine Beaumel, au sein de la division des enquêtes et études démographiques, dans le cadre d'un groupe de travail auquel ont également pris part Laurent Toulemon (Ined) ; Anne-Thérèse Aerts, Noëlle Serruys (Insee, département de l'action régionale) et Olivier Léon (Insee, pôle de service à l'action régionale « Emploi et Population »).

données d'état civil), la question de la déformation de la structure moyenne liée à une pondération qui doit d'abord permettre de se rapporter à la population totale de milieu de période.

1. Quelques rappels sur le nouveau recensement

Jusqu'au recensement de 1999, le dénombrement de la population s'appuyait sur une collecte aussi exhaustive que possible, concentrée sur un mois, au printemps. Les recensements étaient alors espacés de plusieurs années, avec une durée intercensitaire croissante : 1968, 1975, 1982, 1990 et 1999. Ce type de recensement nécessite de mobiliser ponctuellement des moyens importants, humains et budgétaires, parfois dans un contexte économique et budgétaire difficile (ce qui a déjà conduit à un report). Dans le même temps, avec le développement des sources administratives, il est plus difficile de justifier une collecte exhaustive d'information sociodémographique sur la population. L'idée d'une collecte répartie sur plusieurs années sur des échantillons du territoire permettent de pallier à ces difficultés. Il faut ensuite construire une méthode d'estimation permettant de calculer des populations communales se rapportant à une même date. Il est en effet indispensable que les populations légales des communes, dont l'estimation est l'un des principaux objectifs du recensement, aient la même référence. Les populations légales servant de base à de nombreux textes juridiques, il s'agit de garantir l'égalité de traitement entre les communes.

1.1. Une collecte étalée de le temps

Les principes généraux du nouveau mode de collecte du recensement ont été fixés par la loi n°2002-276 du 27 février 2002². Bertrand, Chauvet, Christian, et Grosbras [1], ou Dumais et Isnard [2] les présentent rapidement. Dumais, Bertrand, Kauffmann [3] présente de façon plus détaillée la stratégie d'échantillonnage. On peut également se rapporter à l'ouvrage collectif de l'Insee [4].

La collecte du recensement est désormais étalée sur cinq ans. Elle s'appuie sur une partition du territoire en cinq parties, constituant autant de « groupes de rotation ». Chaque année, une partie du territoire est enquêtée. La collecte commence mi janvier et dure quatre à cinq semaines. Sur les bases du recensement de 1999, 14% des logements sont enquêtés chaque année. Sur cinq ans, l'ensemble du territoire est couvert.

Pour constituer les groupes de rotation, les 36 700 communes de France ont été réparties en deux strates selon leur taille : les communes de moins de 10 000 habitants d'une part ; les communes de plus de 10 000 habitants d'autre part.

1.1.1. Dans les communes de moins de 10 000 habitants

Les communes de moins de 10 000 habitants ont été réparties entre les cinq groupes de rotation de façon à ce qu'une commune sur cinq soit enquêtée chaque année. Le recensement y est exhaustif et a donc lieu tous les cinq ans. Les cinq groupes de communes ont été constitués de façon à ce que chaque groupe reflète autant que possible la structure de l'ensemble des communes de moins de 10 000 habitants de chaque région, pour un certain nombre de critères. Pour cela, la méthode des échantillons équilibrés a été utilisée avec les critères suivants : nombre de logements dans la commune, nombre de logements en immeuble collectif, nombre d'habitants de la commune de moins de 20 ans, de 20 à 39 ans, de 40 à 59 ans, de 60 à 74 ans, de 75 ans ou plus, population des femmes et des hommes et population départementale totale.

² Loi relative à la démocratie de proximité, publiée au *Journal Officiel* du 28 février 2002, titre V, « Des opérations de recensement ».

1.1.2. Dans les communes de plus de 10 000 habitants

Dans les communes de plus de 10 000 habitants, un échantillon d'environ 8% des logements est enquêté chaque année. Sur cinq ans, 40% des logements sont donc enquêtés.

L'ensemble du dispositif s'appuie sur un répertoire d'immeubles localisés (RIL) mis en place pour toutes les communes de plus de 10 000 habitants. Il s'agit d'une liste d'adresses géocodées. Initialisée par le recensement de 1999, elle est mise à jour chaque année à partir de sources administratives (taxe d'habitation, permis de construire, fichiers de La Poste,...), en lien avec les communes. La liste est également validée à partir d'observations sur le terrain réalisées lors de la préparation de chaque enquête annuelle de recensement³.

Ensuite, le tirage de l'échantillon d'une enquête annuelle de recensement s'effectue en considérant trois strates : les adresses comportant un nombre important de logements, ou « grandes adresses » ; les « nouvelles adresses » et les « autres adresses ».

Afin réduire les effets de grappe dans une enquête annuelle de recensement, les « grandes adresses » d'une commune⁴, ont été réparties entre les cinq groupes de rotation, en cherchant à égaliser leur nombre de logements et en évitant autant que possible que les « grandes adresses » enquêtées une année donnée soit géographiquement concentrées sur les mêmes quartiers. Chaque groupe de rotation est enquêté l'année correspondante, de façon exhaustive.

Les « nouvelles adresses »⁵ repérées lors de la mise à jour du répertoire d'immeubles localisés sont ventilées entre les cinq groupes de rotation. Elles ont été enquêtées de façon exhaustive dans les cinq années suivant leur introduction, lors de l'enquête annuelle concernant leur groupe de rotation.

Enfin, les « autres adresses » à enquêter, ont été réparties de façon à constituer cinq groupes équilibrés, avec des critères d'équilibrage analogues à ceux utilisés pour constituer les groupes de rotation des petites communes. Un échantillon d'adresses est tiré dans le groupe de rotation à enquêter, chaque adresse étant recensée exhaustivement. Le poids de tirage est ajusté de façon à aboutir à un taux de sondage de 8% au niveau de la commune, en tenant compte des trois strates.

1.1.3. Pour les personnes « hors ménages »

Les internats d'établissements d'enseignement, les casernes, les prisons, les maisons de retraite, les couvents, etc., constituent des « communautés ». Leur population est recensée de façon exhaustive une fois tous les cinq ans, en mars.

Toutes les communautés d'une commune sont normalement recensées la même année⁶. Pour les communes de moins de 10 000 habitants, la collecte a lieu la même année que les logements ordinaires. Pour les communes de plus de 10 000 habitants, une répartition des communes en cinq groupes de rotation a été effectuée, de façon à ce que la charge de collecte soit, au niveau régional, la même d'une année sur l'autre.

Par ailleurs, l'Insee a mis en place un répertoire des communautés, mis à jour tous les ans à partir de sources administratives comme le fichier FINESS des établissements sanitaires et sociaux, le fichier des établissements scolaires avec internat, le fichier des établissements pénitentiaires, etc..⁷ Le répertoire est expertisé par les communes une fois tous les cinq ans, avant le recensement exhaustif des communautés de la commune.

Les personnes qui résident habituellement dans des « habitations mobiles terrestres » et les personnes sans abri sont recensées de façon exhaustive les deux premiers jours de collecte, avec

³ Validation des adresses entrant dans le groupe de rotation à enquêter, soit environ 20% des logements de la commune. Sur cinq ans, 100% du territoire est donc couvert.

⁴ Adresses réunissant chacune au moins 60 logements et, ensemble, au plus 10% des logements de la commune.

⁵ En moyenne, environ 1% des logements d'une commune.

⁶ Une vingtaine de grandes communes font exception compte tenu du nombre important de communautés qu'elles hébergent. La collecte est dans leur cas étalée sur plusieurs années.

⁷ Ces sources couvrent 80% des communautés dont la quasi totalité des plus grandes.

comme date de référence, le premier jour de collecte à zéro heure⁸. Dans les communes de moins de 10 000 habitants, elles sont recensées l'année du recensement des logements ordinaires. Dans les autres communes, le recensement a été effectué en 2006. Il sera renouvelé tous les cinq ans.

Les marinières et les personnes vivant sur leur bateau ont été recensés de façon exhaustive en 2006. Ce recensement sera renouvelé tous les cinq ans.

1.2. Une population « rapportée » au 1er janvier de l'année médiane

Chaque enquête annuelle dispose d'un poids permettant d'extrapoler au niveau national un résultat annuel⁹. Par la suite, on parlera de « pondération annuelle ». Toutefois, les résultats définitifs du recensement sont ceux qui sont obtenus en cumulant les informations collectées sur cinq enquêtes annuelles consécutives, à partir d'une pondération spécifique. Par la suite, on parlera de « pondération du cumul ».

Pour assurer l'égalité de traitement des communes, il est en effet nécessaire d'estimer la population légale des communes à date donnée, identique pour toutes. Pour cela, la méthode d'estimation rapporte la population des communes au 1er janvier de l'année médiane de la période de collecte. La date de référence est donc le 1er janvier 2006 pour les résultats du « recensement 2006 », construits à partir des enquêtes annuelles collectées de 2004 à 2008.

L'estimation obtenue par cumul des cinq années de collecte n'est donc pas une simple moyenne des cinq années de collecte. La pondération du cumul ne peut pas se déduire des pondérations annuelles.

1.2.1. La pondération de cumul

La note de l'Insee du 16 décembre 2008 [6] présente la méthode de façon pédagogique. Bertrand P., Chauvet G., Christian B. Grosbras J.-M. [5] présentent quelques travaux de simulation sur la précision.

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

Pour les communes recensées les deux années précédant l'année médiane (2004 et 2005), 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 disponibles¹⁰, on calcule le ratio égal au rythme annuel d'évolution de la population des ménages rapporté au rythme annuel d'évolution du nombre de résidence principale. 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

⁸ Soit la date de référence pour le recensement des logements.

⁹ L'utilisation régulière des résultats annuels suppose une expertise préalable. Quelques chiffres ont toutefois été présentés sur l'emploi et le chômage lors d'une réunion du Conseil national de l'information statistique, en mars 2007.

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.

attribuer une même pondération p_1 à tous les logements d'une petite commune donnée, p_1 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 les deux années suivant l'année médiane (2007 et 2008), 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 antérieures à cette date. Pour 2006, il s'agissait des résultats du recensement de 1999. Pour les résultats du recensement de 2007, ce seront les résultats de 2006. En régime permanent, l'interpolation sera donc au plus de deux ans. L'interpolation conduit à attribuer une même pondération p_2 à tous les logements d'une petite commune donnée, p_2 variant d'une commune à l'autre.

Dans les petites communes extrapolées (donc celles recensées en 2004 et 2005), 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 p_1 d'un facteur p_3 , variable d'un logement à l'autre dans une même commune. La pondération de cumul est alors $p_1 \times p_3$. 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.

1.2.1.2. 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 date¹¹.

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'adresses¹². 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.

Le calage ne s'effectue pas directement au niveau de la commune mais à un niveau infra communal, souvent l'IRIS. Ce calage modifie les pondérations annuelles associées aux individus recensés.

Lors de ces traitements, il n'y a pas de distinction faite selon les trois strates de tirage (« grandes adresses », « nouvelles adresses », « autres adresses »).

¹¹ 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 .

¹² Compte-tenu des objectifs du recensement, le traitement de la non réponse, ou « logements non enquêtés », est intégrée 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 également 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.

1.2.1.3. Population « hors ménages »

Pour les populations des communautés, 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 des sources 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 marinières, les effectifs sont maintenus constants entre deux opérations de recensements de ces catégories, donc sur cinq ans.

1.2.2. La pondération annuelle

Dans les communes de moins de 10 000 habitants, les personnes recensées (et les logements) sont pondérées de façon à représenter l'ensemble des petites communes. Il n'y a pas de calage. Le poids des logements est le même au sein d'une même commune.

Dans les communes de plus de 10 000 habitants, la pondération de tirage est corrigé par un calage sur le répertoire d'immeubles localisés ayant servi de base de sondage à l'enquête annuelle¹³.

Aucune des pondérations n'est arrondie. Les populations légales sont arrondies après calcul. La somme des populations légales communales diffère donc de quelques unités de la population totale que l'on obtiendrait en faisant l'arrondi de la somme des poids sur l'ensemble de la France.

2. Les conséquences du nouveau recensement pour le calcul de la pyramide des âges de référence pour les indicateurs démographiques

La méthode d'estimation des résultats définitifs du recensement permet de disposer d'une estimation de population totale relative à l'année médiane, quel que soit le niveau géographique. En revanche, les informations relatives aux caractéristiques des personnes sont celles disponibles à la date de collecte du recensement. En d'autres termes, pour une petite commune collectée en 2004, sa population sera extrapolée pour correspondre à une population au 1er janvier 2006 ; elle différera donc du niveau de population collectée. En revanche, la proportion de femmes, de moins de 18 ans, de plus de 60 ans, de personnes en emploi ou la taille moyenne du ménage, seront *grosso modo* celles observées en 2004¹⁴. Dans les grandes communes, enquêtées chaque année, les caractéristiques reflèteront une situation moyenne sur la période.

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* 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.

¹³ Donc daté du 1er juillet précédant la collecte.

¹⁴ Voir 2.2

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 devront être réalisés en réalité sur le croisement sexe-génération ou sexe-âge.

2.1. Pyramide par âge ou pyramide par génération

Pour illustrer les différences, un exercice fictif simplifié¹⁵ peut être réalisé en s'appuyant sur les pyramides des âges déjà publiées de façon à simuler de façon plausible ce qu'aurait 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 de référence datée du « 1er janvier 1999 », c'est-à-dire la pyramide effectivement publiée.

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 1997 à 2001¹⁶, puis en recalant les effectifs ainsi obtenus par sexe sur le total publié pour le 1er janvier 1999 (les coefficients de recalage 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. 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. Enfin, parce que le raisonnement est effectué séparément pour les deux genres, afin de mettre l'accent sur la principale difficulté : l'âge.

Le graphique 1 montre les résultats de la comparaison des niveaux absolus (graphique 1a) et l'écart en niveau relatif (graphique 1b) pour les hommes. Ils illustrent le lissage opéré lorsque l'on cumule des informations sur la variable « âge ». En particulier, les passages aux classes creuses de 1914-1918 puis à son écho sur les générations descendantes de ces classes creuses sont atténués.

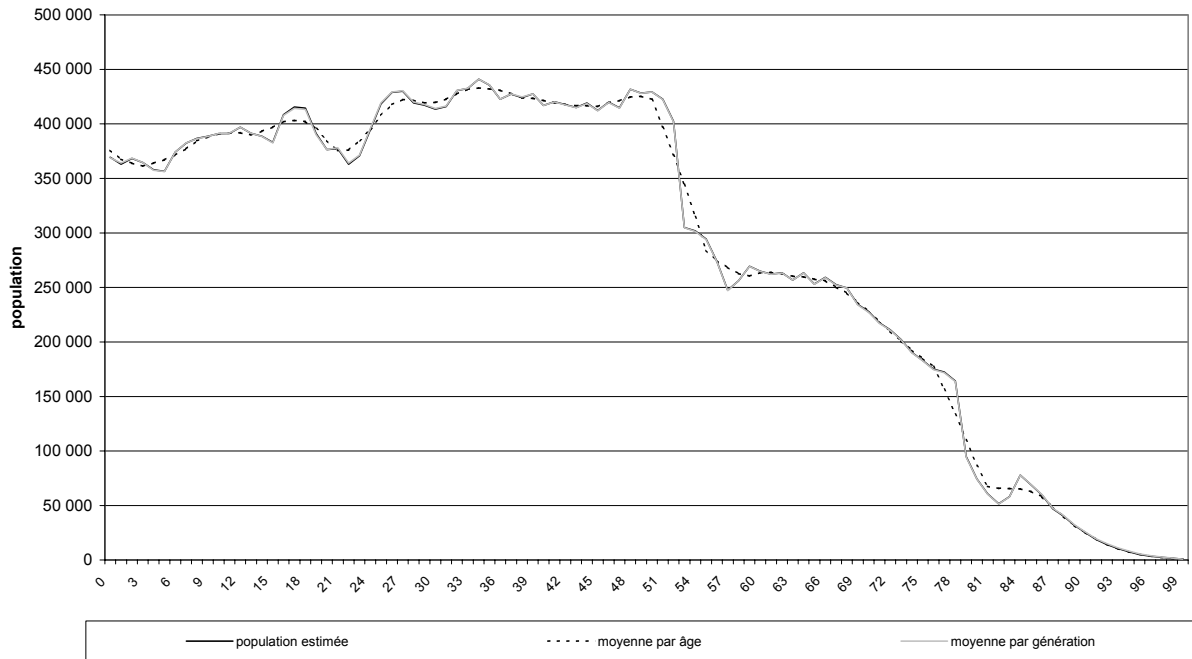
La comparaison des effectifs par âge en niveau relatif montre que les écarts peuvent être importants aux âges correspondant à des changements dans l'histoire démographique : ils peuvent ainsi dépasser 10%, voire 20% à certains âges. Cette comparaison montre au contraire que le cumul par génération permet d'aboutir à une pyramide reflétant beaucoup mieux les aléas démographiques, sauf aux âges élevés, pour lesquels elle surestime de façon croissante avec l'âge les effectifs. 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 si les migrations et les décès affectent relativement peu les effectifs. Or, aux âges élevés, le taux de mortalité devient très important et 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 crée un biais dans les estimations des effectifs des différentes générations qui perturbe ensuite le calcul des indicateurs démographiques en créant ainsi 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 (graphique 2b). L'utilisation de la pyramide de cumul par âge aboutit à une courbe heurtée aux âges pour lesquels l'effet de lissage est important, qui ne reflète pas la réalité de l'évolution de la mortalité avec l'âge. D'une année sur l'autre, ces artefacts se déplaceraient, suivant le vieillissement des générations marquant les « accidents » démographiques. Pour sa part, la pyramide de cumul par génération conduit à une sous estimation des risques de décès aux âges élevés.

¹⁶ Voir *infra* pour le traitement des âges extrêmes pour la pyramide moyenne par génération.

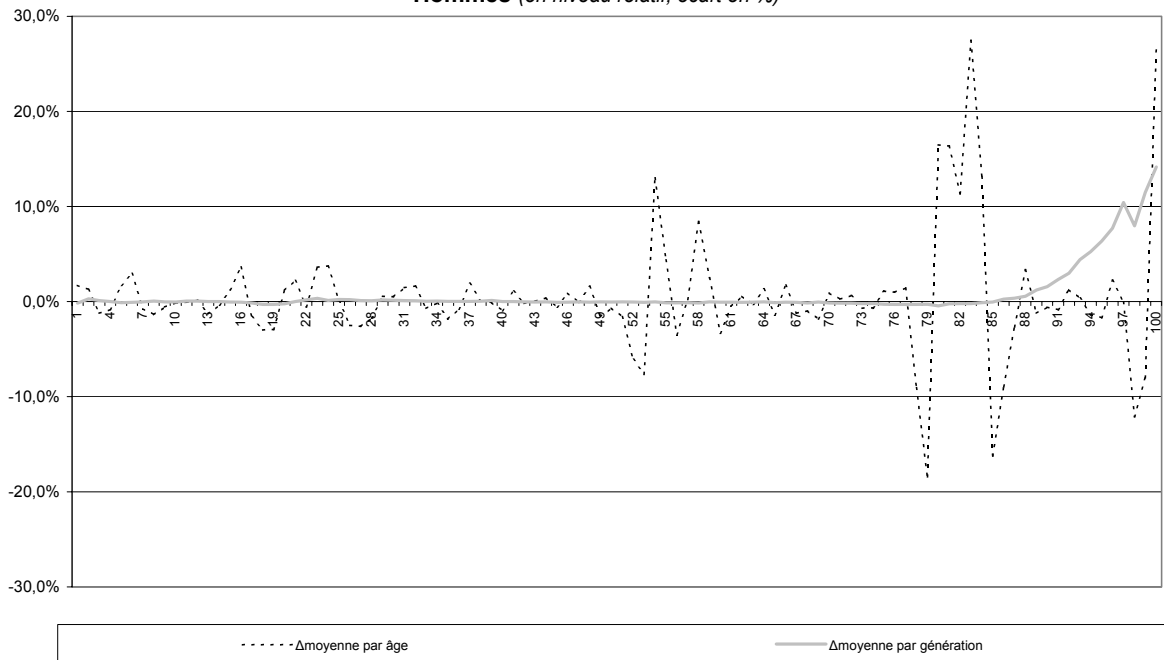
Graphique 1a

Simulation de la pyramide au 1er janvier 1999 - Hommes



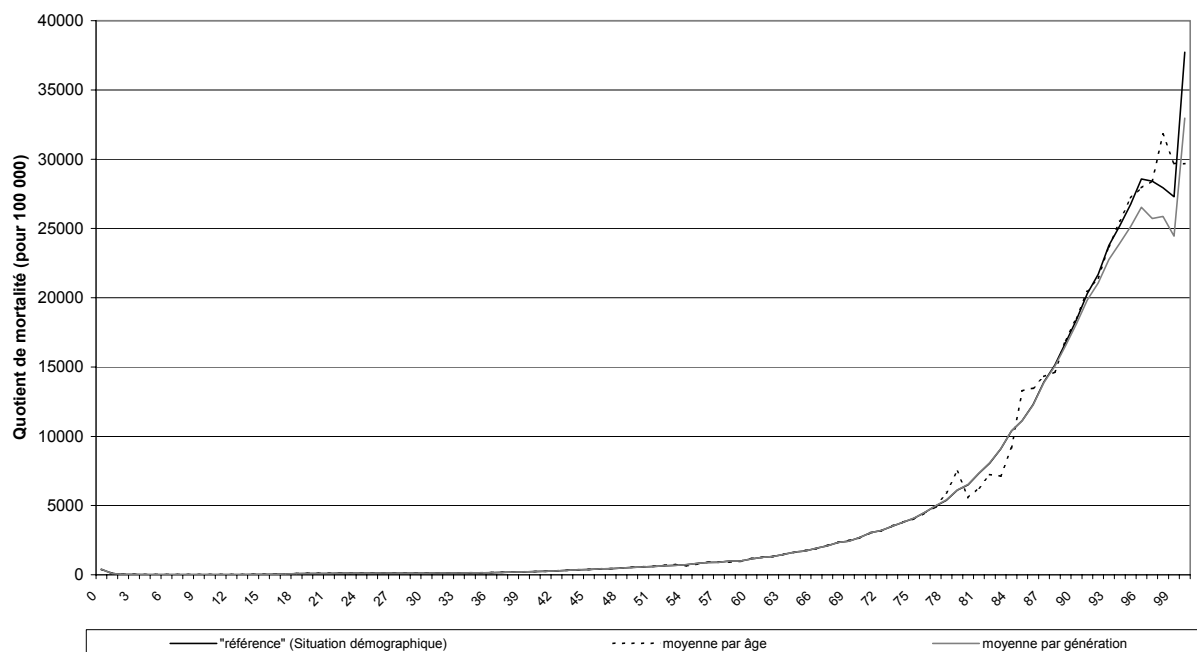
Graphique 1b

Comparaison des pyramides simulées au 1er janvier 1999 avec la pyramide de référence - Hommes (en niveau relatif, écart en %)



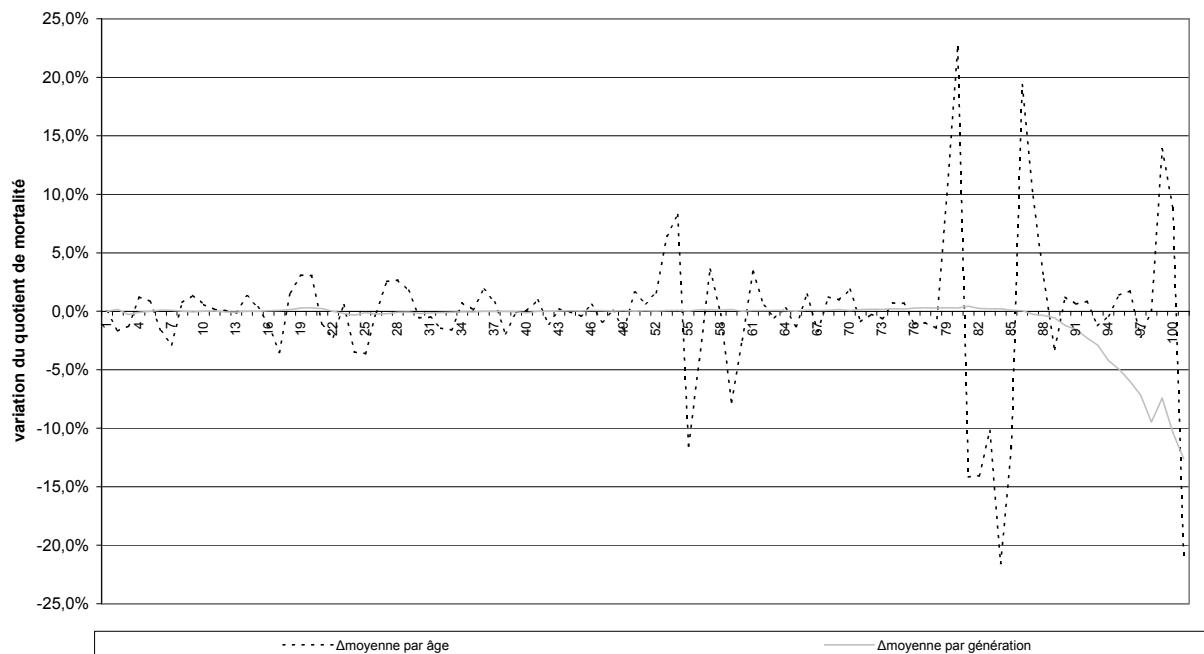
Graphique 2-a

Comparaison des quotients de mortalité des hommes en 1999



Graphique 2b

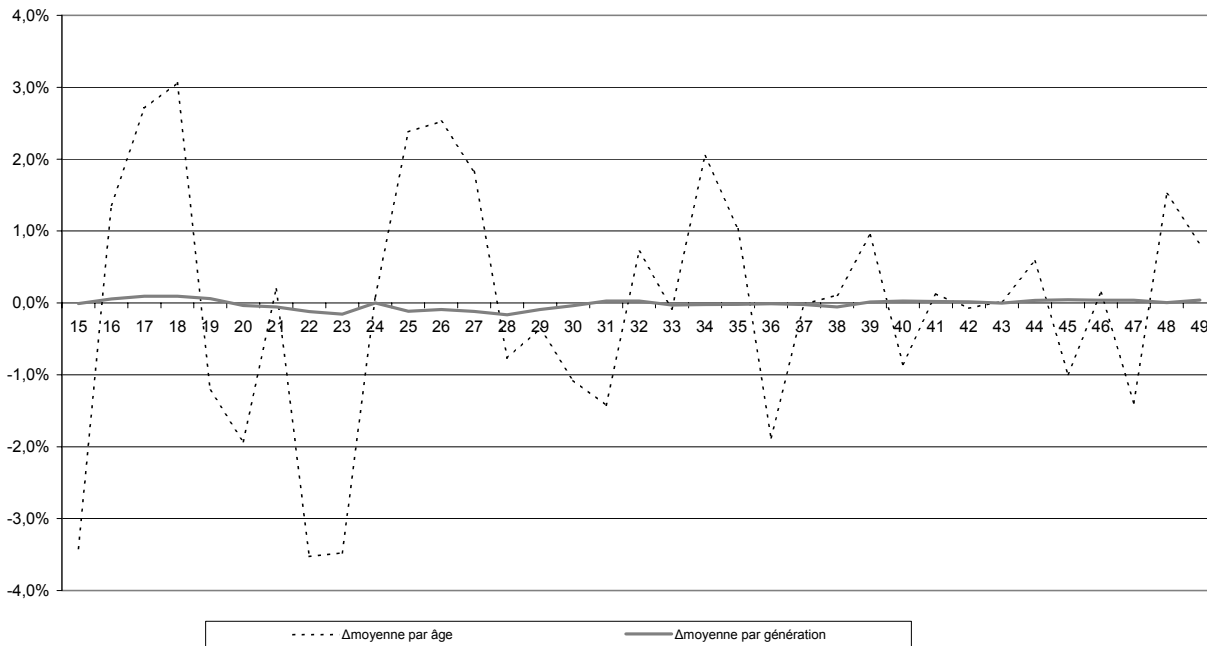
Ecart des quotients de mortalité par âge des hommes calculés avec les pyramides simulées, par rapport à la référence (en %)



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

Graphique 3

Ecart des taux de fécondité par âge avec les pyramides simulées rapport à la référence
(en %)



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 née l'année g , présente au 1er janvier de l'année n (population d'âge x , en âge atteint, avec $x=n-g$) ;
 $P_r(g,n)$, la population née l'année g recensée dans l'enquête annuelle de l'année n ;
 $D(g,n)$, les décès dans l'année n de personnes nées l'année g ;
 $N(g)$, les naissances dans l'année g ;
 $SM(g,n)$, le solde migratoire dans l'année n de personnes nées l'année g ;
 $e(g,n) = P(g,n) - P_r(g,n)$, l'écart entre la population née l'année g recensée au 1er janvier de l'année n dans l'enquête annuelle de l'année et l'effectif réel, inconnu¹⁷.

Dans la suite, l'année n est 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 .

¹⁷ Cet écart résulte à la fois des conditions de collecte sur le terrain, des conditions de retraitements des informations et de 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, permettent de mieux contrôler les conditions de collecte et de retraitement, donc de réduire les écarts qui leurs sont imputables.

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

Quand la pyramide est obtenue par le cumul des informations collectées sur cinq ans selon l'âge l'année du recensement, l'effectif à chaque âge $x=n-g$, est estimée, dans notre cas simplifié, par $P^*(x)$:

$$\begin{aligned}
 P^*(x) &= 1/5. [P_r(g-2,n-2) + P_r(g-1,n-1) + P_r(g,n) + P_r(g+1,n+1) + P_r(g+2,n+2)] \\
 &= 1/5. [P(g-2,n-2) + P(g-1,n-1) + P(g,n) + P(g+1,n+1) + P(g+2,n+2)] \\
 &\quad - 1/5. [e(g-2,n-2) + e(g-1,n-1) + e(g,n) + e(g+1,n+1) + e(g+2,n+2)] \\
 &= 1/5. [P(g-2,n) + D(g-2,n-2) + D(g-2,n-1) - SM(g-2,n-2) - SM(g-2,n-1) \\
 &\quad + P(g-1,n) + D(g-1,n-1) - SM(g-1,n-1) \\
 &\quad + P(g,n) \\
 &\quad + P(g+1,n) - D(g+1,n) + SM(g+1,n) \\
 &\quad + P(g+2,n) - D(g+2,n) - D(g+2,n+1) + SM(g+2,n) + SM(g+2,n+1)] \\
 &\quad - 1/5. [e(g-2,n-2) + e(g-1,n-1) + e(g,n) + e(g+1,n+1) + e(g+2,n+2)] \\
 &= 1/5. [P(g-2,n) + P(g-1,n) + P(g,n) + P(g+1,n) + P(g+2,n)] \\
 &\quad + 1/5. [D(g-2,n-2) + D(g-2,n-1) + D(g-1,n-1) - D(g+1,n) - D(g+2,n) - D(g+2,n+1)] \\
 &\quad - 1/5. [SM(g-2,n-2) + SM(g-2,n-1) + SM(g-1,n-1) - SM(g+1,n) - SM(g+2,n) - SM(g+2,n+1)] \\
 &\quad - 1/5. [e(g-2,n-2) + e(g-1,n-1) + e(g,n) + e(g+1,n+1) + e(g+2,n+2)]
 \end{aligned}$$

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. Les écarts entre les données de collecte et la réalité, également. Dans ces conditions, l'effectif d'un âge donné obtenu par un cumul sur cinq ans des informations par âge est proche du premier terme, donc d'une moyenne des effectifs de cinq âges successifs. Il correspond donc à un lissage de la vraie pyramide.

2.1.2. Cas de la pyramide obtenue par cumul des informations selon la génération

Quand la pyramide est obtenue par le cumul des informations collectées sur cinq ans selon la génération, l'effectif à chaque âge $x=n-g$, correspond à l'effectif de la génération g , estimée, dans notre cas simplifié, par $P^*(g)$:

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

$$\begin{aligned}
& - 1/5. [e(g,n-2) + e(g-1,n-1) + e(g,n) + e(g+1,n+1) + e(g+2,n+2)] \\
& = P(g,n) \\
& + 1/5. [D(g,n-2) + 2.D(g,n-1) - 2.D(g,n) - D(g,n+1)] \\
& - 1/5. [SM(g,n-2) + 2.SM(g,n-1) - 2.SM(g,n) - SM(g,n+1)] \\
& - 1/5. [e(g,n-2) + e(g,n-1) + e(g,n) + e(g,n+1) + e(g,n+2)]
\end{aligned}$$

Sous les mêmes conditions que précédemment, l'effectif d'une génération obtenu par un cumul sur cinq ans des informations par génération approche le premier terme, qui correspond aux effectifs exacts de la génération.

Par ailleurs, dans le cas de la pyramide obtenue par le cumul des informations selon la génération, l'estimation des effectifs des âges extrêmes se heurte à un effet de bord :

- pour la génération née l'année $g=n-2$, l'effectif recensé l'année $n-2$, est nul¹⁸. Par contre, cette génération est recensée les quatre années suivantes. Dans la mesure où la mortalité et les migrations aux bas âges sont négligeables, 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.
- 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.

2.1.3. Ampleur des biais selon l'âge

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, à deux sources d'écart près : l'une liée 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. Pour les estimations de population, le premier facteur se décompose en deux : le mouvement naturel et le mouvement migratoire.

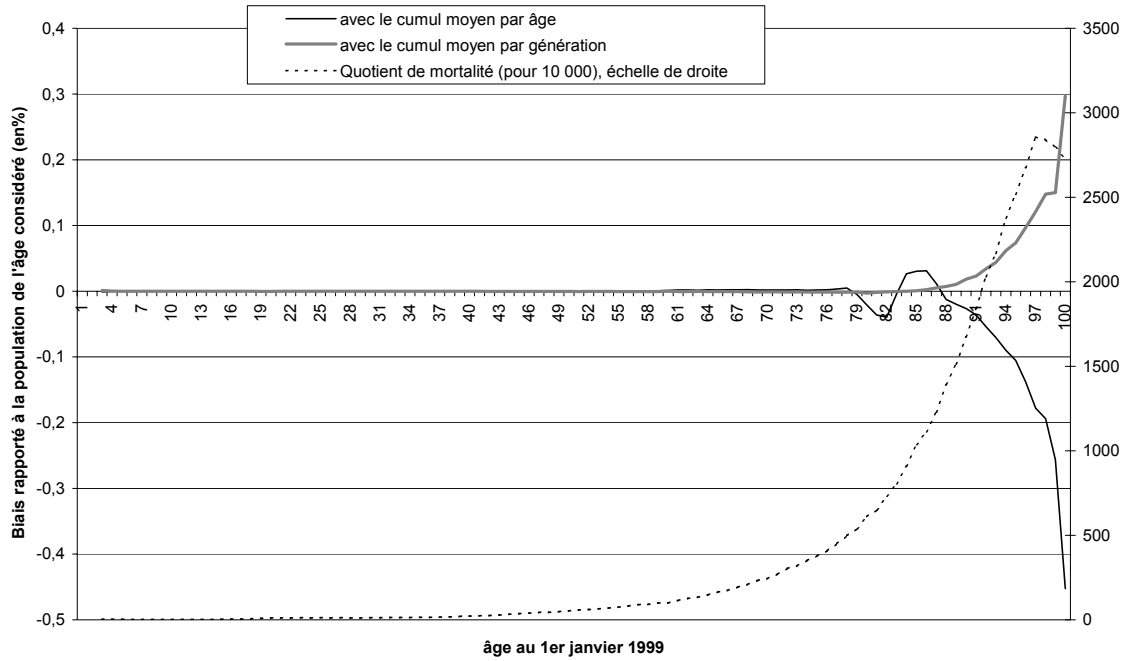
Il est possible d'estimer l'ampleur du biais lié à la non prise en compte des décès par âge à partir des statistiques d'état civil. Le graphique 4 montre le niveau relatif du biais, celui-ci étant rapporté à la population réelle de l'âge considérée. Le biais est négligeable jusqu'aux alentours de 80 ans dans les deux approches. Il progresse ensuite aux âges élevés, lorsque les quotients de mortalité deviennent élevés.

Au niveau national, il est aussi possible d'estimer l'ampleur du biais lié à la non prise en compte du solde migratoire à partir des estimations du solde migratoire par sexe et âge. Le graphique 5 montre le niveau relatif de ce biais, celui-ci étant rapporté à la population réelle de l'âge considérée. Le biais apparaît négligeable à tout âge. Il s'élève en valeur absolue 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. Si les décès par âge sont bien connus grâce aux statistiques d'état civil, les estimations de solde migratoire par âge repose sur des informations limitées et sont donc moins robustes.

¹⁸ 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.

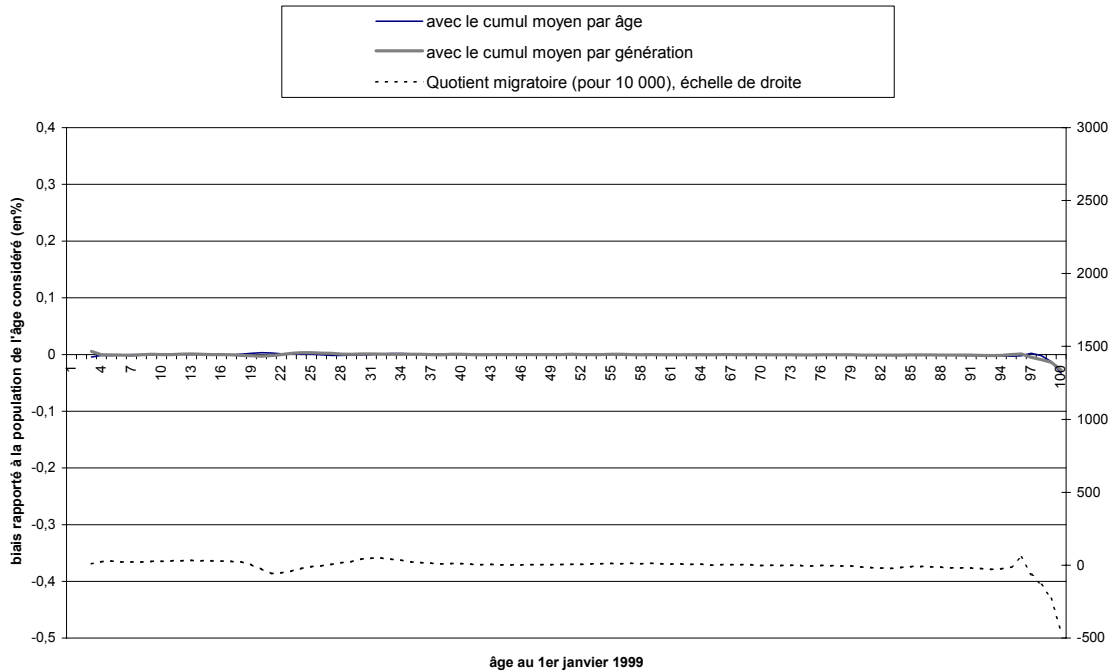
Graphique 4

Biais relatif lié aux sorties par décès (cas des hommes)



Graphique 5

Biais relatif lié aux migrations (cas des hommes)



2.2. Une structure moyenne légèrement déformée

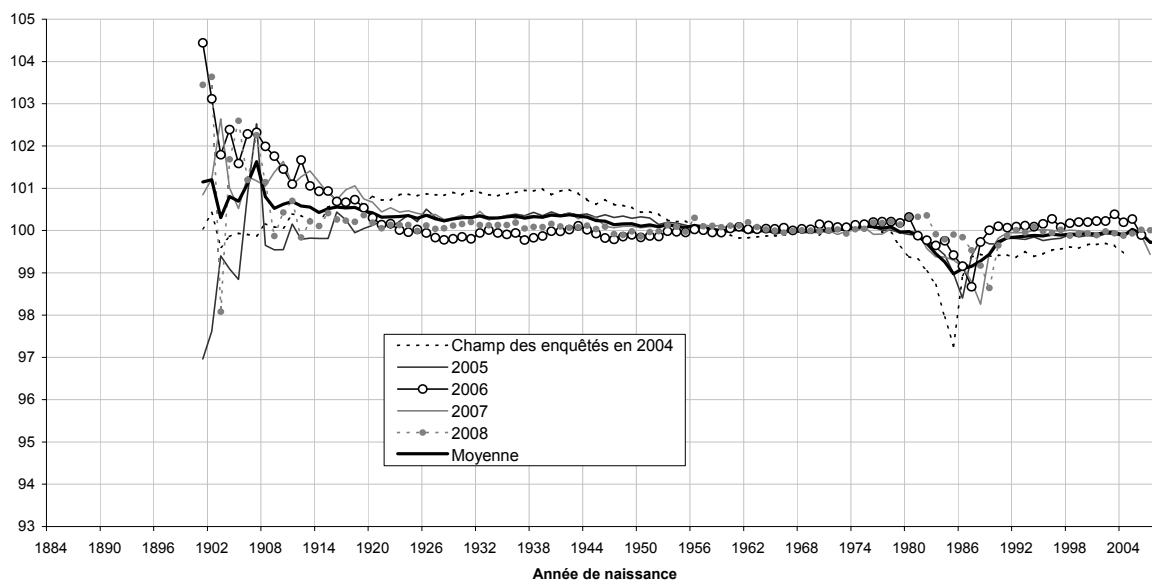
En pratique, les résultats du recensement ne sont pas obtenus par simple moyenne des informations collectées chaque année. Pour fournir une population légale pour chaque commune à la même date, une modification des pondérations est effectuée, par extrapolation et interpolation dans les ménages des petites communes et les communautés, par recalage sur le répertoire d'immeubles localisés « médian », pour les ménages des grandes communes.

Comme le coefficient de redressement appliqué au niveau de population recensé n'est pas identique d'une petite commune à l'autre et d'un IRIS 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 où l'on regarderait la structure par âge de l'ensemble des petites communes recensées une même année donnée.

Le graphique 6 illustre les écarts de la structure par âge et sexe selon que l'on utilise les résultats d'une enquête annuelle, avec la pondération « annuelle », ou les résultats du cumul 2004-2008. A chaque fois, l'exercice est réalisé sur le champ de seules personnes recensées lors de l'enquête annuelle considérée.

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)



Aux âges les plus avancés, le poids d'une génération d'un sexe donné dans la population, donc ses effectifs, peut être très différent entre celui obtenu avec la pondération de cumul (celle des populations légales) et celui obtenu avec la pondération annuelle, mais les effectifs sont très faibles.

En dehors de ces cas, l'écart du poids, 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. Il dépasse 1% pour les jeunes majeurs et aux grands âges. Pour disposer des données de cadrage dont ont besoin la plupart des utilisateurs, ces écarts restent toutefois minimes et négligeables. La répartition entre hommes et femmes est elle-aussi concernée mais de façon encore plus négligeable : 48,37% d'hommes dans la population avec la pondération annuelle, 48,38% avec la pondération de cumul.

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

données détail du recensement, deux poids sont attribués à chaque individu : le poids annuel qui permet de disposer de résultats sur la collecte d'une année donnée, et le poids de cumul, utilisé pour produire la population légale à partir des cinq enquêtes annuelles consécutives.

Soit : - $P_r(n)$, la population totale à l'enquête annuelle collectée l'année n (il s'agit ici de l'effectif total obtenu en utilisant le poids annuel).
 - $P_r(g,n)$, la population totale de la génération g à l'enquête annuelle collectée l'année n (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 $\alpha(g,n)$, le poids de la génération g dans la population totale, dans l'enquête annuelle n (poids annuel). $\alpha(g,n) = P_r(g,n) / P_r(n)$.

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

On appelle $k(g,n) = \alpha^*(g,n) / \alpha(g,n)$: 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 n (dans le graphique 6, les courbes de parts relatives correspondent à des courbes de $k(g, \text{année de collecte})$ en indice base 100).

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

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

Pour simplifier l'écriture dans l'expression suivante, on appelle : $\alpha_1(g,n) = k(g,n-2) \cdot (P^*(n-2) / P_r(n-2))$, la valeur du produit de ces deux coefficients pour la génération g , collecté 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 :

$$\begin{aligned}
 P^*(g) &= P^*(g,n-2) + P^*(g,n-1) + P^*(g,n) + P^*(g,n+1) + P^*(g,n+2) \\
 &= (\alpha_1(g,n) + \alpha_2(g,n) + \alpha_3(g,n) + \alpha_4(g,n) + \alpha_5(g,n)) \cdot P(g,n) \\
 &\quad + \alpha_1(g,n) \cdot D(g,n-2) + (\alpha_1(g,n) + \alpha_2(g,n)) \cdot D(g,n-1) - (\alpha_4(g,n) + \alpha_5(g,n)) \cdot D(g,n) - \alpha_5(g,n) \cdot D(g,n+1) \\
 &\quad - \alpha_1(g,n) \cdot SM(g,n-2) + (\alpha_1(g,n) + \alpha_2(g,n)) \cdot SM(g,n-1) - (\alpha_4(g,n) + \alpha_5(g,n)) \cdot SM(g,n) - \alpha_5(g,n) \cdot SM(g,n+1) \\
 &\quad - \alpha_1(g,n) \cdot e(g,n-2) + \alpha_2(g,n) \cdot e(g,n-1) + \alpha_3(g,n) \cdot e(g,n) + \alpha_4(g,n) \cdot e(g,n+1) + \alpha_5(g,n) \cdot e(g,n+2)
 \end{aligned}$$

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 sur cinq ans sans prise en compte de cette déformation conduirait à perturber le profil du solde migratoire apparent par âge.

2.3. Les changements de concepts par rapport au recensement de 1999

2.3.1. L'âge révolu à la date du recensement

Pour les comparaisons entre les différents recensements, par exemple pour les rétopolations, et pour l'établissement de la pyramide des âges de référence, 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, il est possible de calculer deux âges :

- l'âge atteint dans l'année, obtenu en faisant la différence de millésime entre l'année de l'enquête de recensement et l'année de naissance ;
- l'âge en année révolue à la date du recensement, c'est-à-dire, l'âge atteint au dernier anniversaire, avant l'enquête de recensement.

Dans les résultats diffusées 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 1999, la date de référence utilisée est le 8 mars 1999. Pour les enquêtes annuelles de recensement, il s'agit :

- 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.

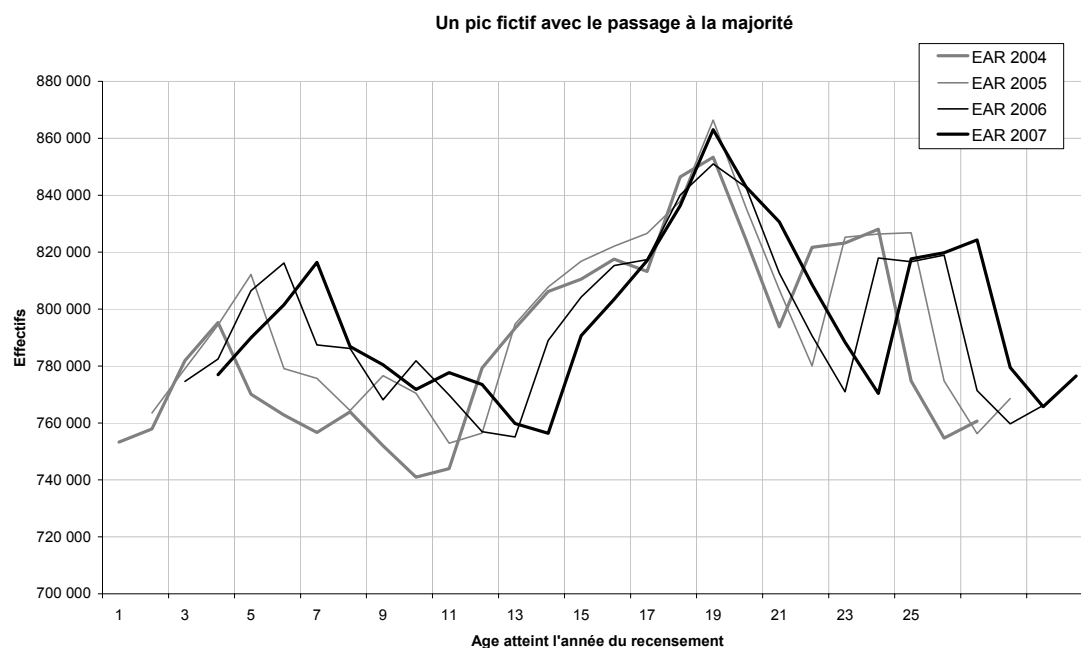
Pour les calculs d'une pyramide des âges par génération, c'est l'année de naissance qui est prise en compte. Par souci d'homogénéité dans les travaux présentés, la pyramide des âges est donc construite selon l'âge atteint dans l'année.

2.3.2. 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) 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



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 recensés 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, ce qui supprimera tout risque de ce point de vue.

2.3.3. Retour sur la date de référence

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 la pyramide issue du recensement au 1er janvier. 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 vivants 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 (cf. supra). Seules les personnes nées avant la date de référence sont prises en compte.

Comme avec le recensement général, la collecte décrit une situation moyenne légèrement postérieure 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 population 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.

In fine, compte tenu de la complexité des calculs à réaliser et des hypothèses sous-jacentes à faire pour cela, il n'est pas certain que la correction marginale qui serait apportée pour tout rapporter effectivement au 1er janvier ne déformerait pas autant (à la marge) la pyramide des âges que l'absence de correction.

3. La nouvelle pyramide des âges de référence pour le calcul des indicateurs démographiques

Pour éviter des artefact 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. Les deux options envisageables sont dès lors :

- soit l'utilisation d'une pyramide des âges par génération à partir des cinq années de collecte en la corrigeant des décès ;
- soit l'utilisation d'une pyramide des âges telle qu'elle résulte d'une seule enquête annuelle de recensement¹⁹.

¹⁹ Age révolu à la date de référence du recensement (mi-janvier)

3.1. 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 :

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

L'application du coefficient multiplicateur $(P^*/P')^{20}$ permet de caler la somme des effectifs par génération ainsi obtenue sur les effectifs du recensement, c'est-à-dire la population légale au 1er janvier de l'année médiane.

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 :

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

Cette deuxième relation n'est pas privilégiée d'emblée dans la mesure où les estimations du solde migratoire par âge sont pour le moment fragiles compte tenu du peu de sources d'informations disponibles sur les flux d'entrées et de sorties (hormis les entrées légales d'étrangers ressortissant de pays non membres de l'Union européenne).

Par ailleurs, cette solution s'appuie sur l'utilisation des enquêtes annuelles (donc des pondérations des enquêtes annuelles) plutôt que sur une tabulation directe des résultats du cumul des cinq années par année de naissance (c'est-à-dire la pondération du cumul), ce qui permet de s'affranchir du calcul des coefficients de déformations 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 modifie de toute manière la pyramide des âges par rapport à celle obtenue avec les seuls résultats du recensement.

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

- l'utilisation de la même démarche générale, en se limitant aux seules enquêtes annuelles pertinentes (utilisation de 4 enquêtes annuelles pour la génération née l'année $n-2$ et de trois enquêtes annuelles pour celle née l'année $n-1$), soit :

$$P'(n-2) = \left(\frac{1}{4} \cdot [P_r(n-2, n-1) + P_r(n-2, n) + P_r(n-2, n+1) + P_r(n-2, n+2)] \right. \\ \left. - \frac{1}{4} \cdot [D(n-2, n-1) - 2 \cdot D(n-2, n) - D(n-2, n+1)] \right) \cdot (P^*/P')$$

$$P'(n-1) = \left(\frac{1}{3} \cdot [P_r(n-1, n) + P_r(n-1, n+1) + P_r(n-1, n+2)] \right. \\ \left. - \frac{1}{3} \cdot [-2 \cdot D(n-1, n) - D(n-1, n+1)] \right) \cdot (P^*/P')$$

- 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 conduit à un effectif plus élevé que la

²⁰ 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 génèrent du solde migratoire apparent, par comparaison de deux pyramides des âges.

première puisque les enquêtes annuelles, comme les autres recensements, conduisent à une estimation de population d'âge 0 plus faible que le nombre de naissances corrigé des décès.

Pour illustrer les impacts de ce choix sur les indicateurs démographiques, la simulation effectuée avec les données publiées dans la *Situation démographique* à partir des pyramides des âges au 1er des années 1997 à 2002, peut être poursuivie.

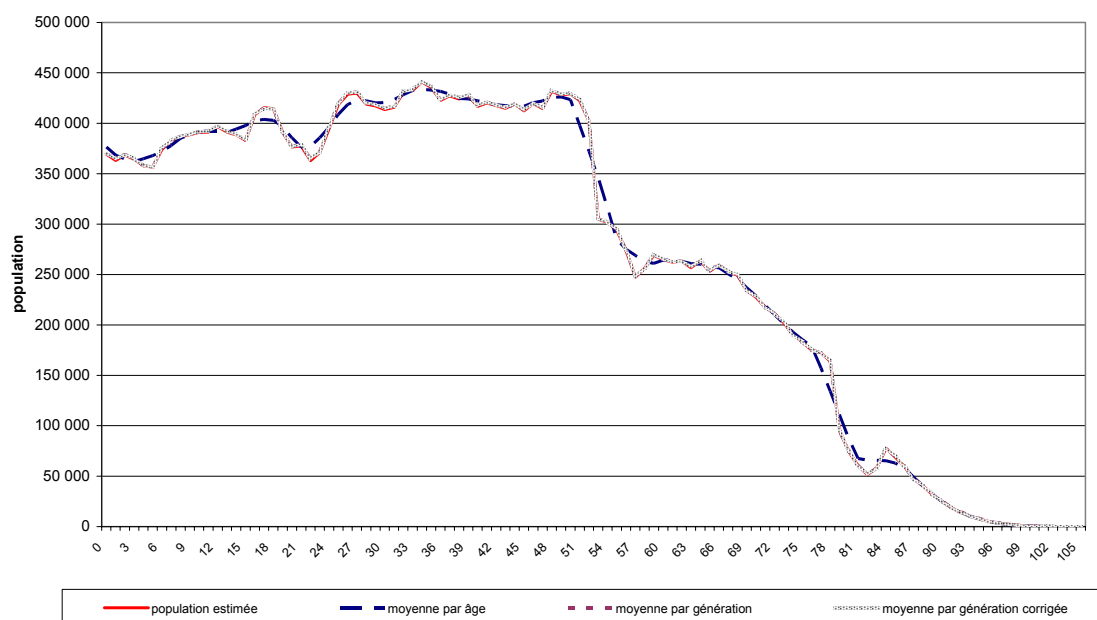
Quatre pyramides sont ensuite utilisées²¹ :

- la pyramide au 1er janvier 1999 diffusée (« population estimée ») qui représente la pyramide réelle dans cette simulation ;
- une pyramide moyenne « par âge », correspondant à ce que pourrait donner un cumul sur cinq ans fait à partir d'une variable « âge » (utilisation de la moyenne simple des effectifs des cinq années),
- une pyramide moyenne « par génération », correspondant à ce que pourrait donner un cumul sur cinq ans fait à partir d'une variable « année de naissance » (utilisation de la moyenne simple des effectifs des cinq années),
- une pyramide moyenne par génération corrigée, qu'on pourrait obtenir à partir d'une variable « année de naissance » sur un cumul de cinq ans et des corrections faites à partir des statistiques d'état civil sur les décès (les décès ayant eu lieu en 1997 et 1998 pour une génération donnée sont retirés des estimations de population 1997 pour les caler sur 1999 ; de la même façon, les décès ayant eu lieu en 1998 sont déduits des estimations 1998 ; inversement, les décès ayant eu lieu en 1999 sont ajoutés aux estimations 2000, etc.).

Sur la pyramide des âges

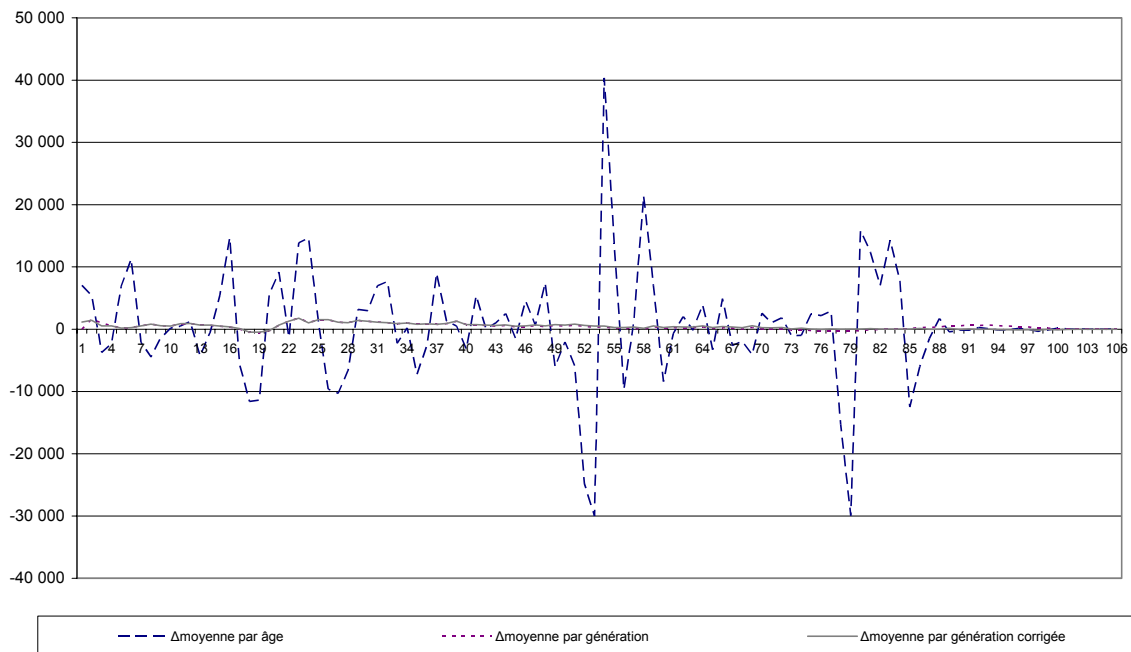
Graphiques 8a et 8b

Pyramides - homme 1999



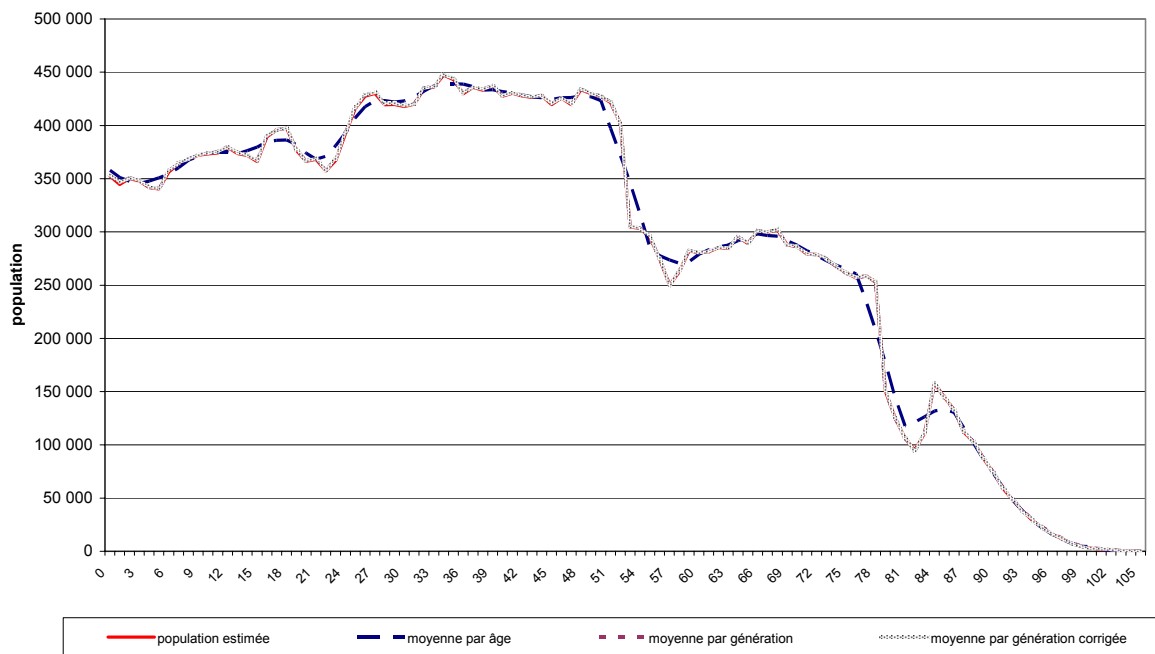
²¹ De 1997 à 1998, les estimations de populations par âge sont réalisées sur des populations allant de 0 à 100 ans et plus ; à partir de 1999, jusqu'à 110 ans et plus. Pour plus de cohérence et parce qu'il y avait des répercussions pour les estimations de populations aux âges élevés et par conséquent sur les indicateurs démographiques, pour les années 1997 et 1998, la population de 100 ans et plus a été répartie entre 100 et 110 au prorata de la population de 1999.

Comparaison des méthodes d'estimations des pyramides avec celle publiée dans le bilan - hommes 1999

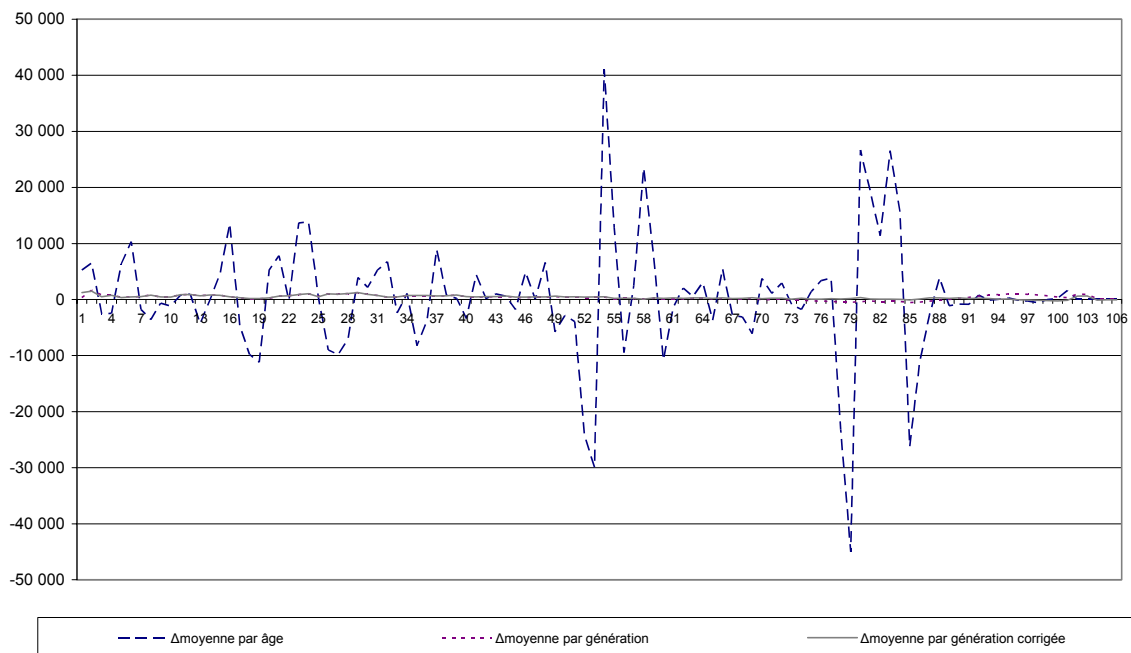


Graphiques 9a et 9b

Pyramides - femme 1999



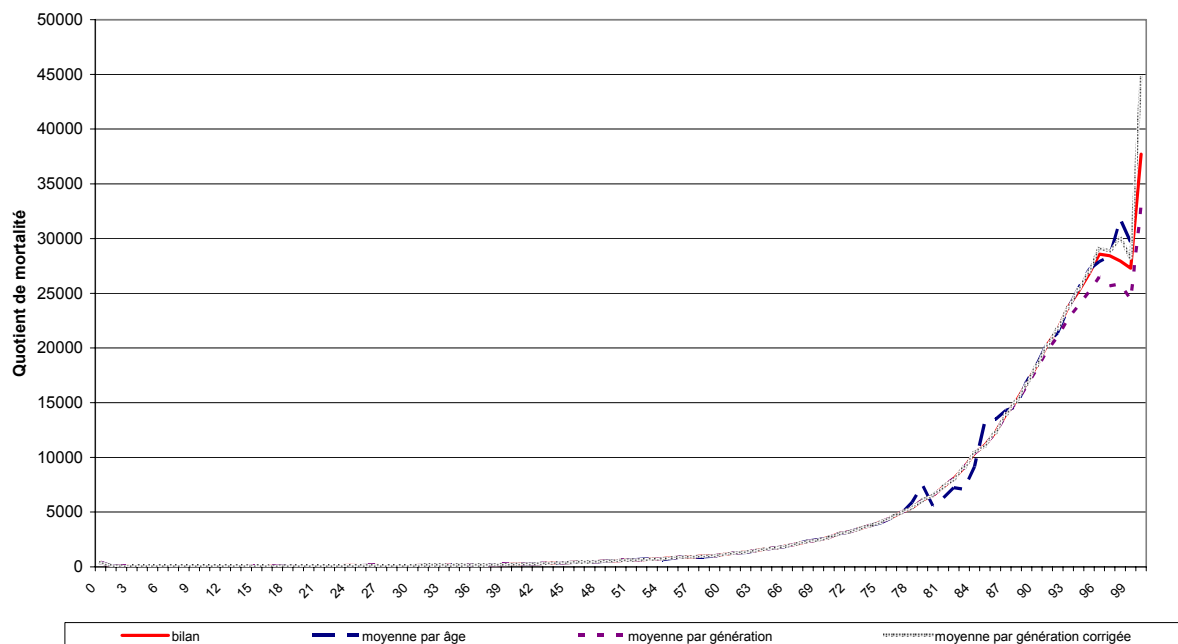
Comparaison des méthodes d'estimations des pyramides avec celle publiée dans le bilan - femmes 1999



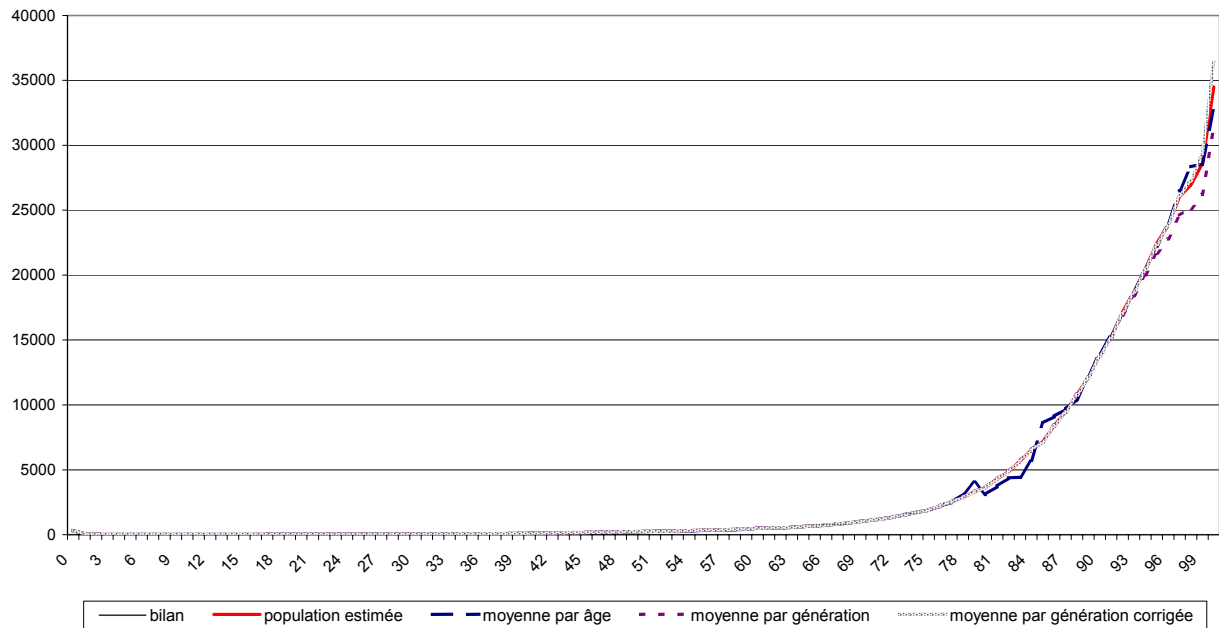
Sur les quotients de mortalité

Graphiques 10a et 10b

Comparaison des quotients de mortalité des hommes



Comparaison des quotients de mortalité des femmes



Ces courbes montrent à nouveau que les pyramides « par génération » donnent des indicateurs moins heurtés que la pyramide « par âge » et illustrent le fait que la correction du biais des décès dans la pyramide « par génération » permet de d'éviter la dérive observée aux âges élevés de la courbe non corrigée des décès par rapport à la courbe de référence.

Aux âges les plus élevés, des écarts apparaissent mais les effectifs sont très faibles, rendant les indicateurs moins stables et moins robustes, y compris lorsqu'ils sont calculés à partir de la pyramide de référence.

A partir du nombre de survivants à l'âge exacte, on en déduit l'espérance de vie à la naissance :

	Hommes	Femmes
Publiées dans la <i>Situation démographique</i>	75	82,52
Calculées à partir de la pyramide 1999	75	82,52
À partir d'une moyenne par âge des données de 1997 à 2002	75,05	82,57
À partir d'une moyenne par génération des données de 1997 à 2002	75,03	82,55
À partir d'une moyenne par génération des données de 1997 à 2002, corrigée des naissances et des décès	75	82,53

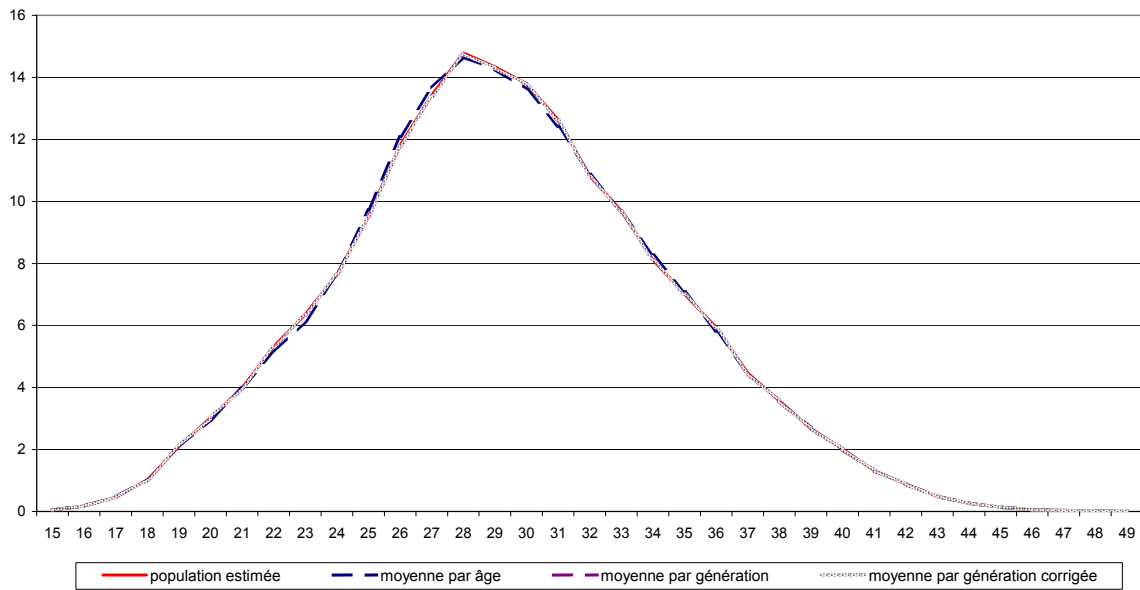
L'utilisation de la pyramide « par génération corrigée » donne les espérances de vie les plus proches de celles de référence. Toutefois, les indicateurs restent très proches quelle que soit la pyramide utilisée. Ils ne sont pas de nature à modifier les analyses en niveau et en évolution.

Sur les taux de fécondité

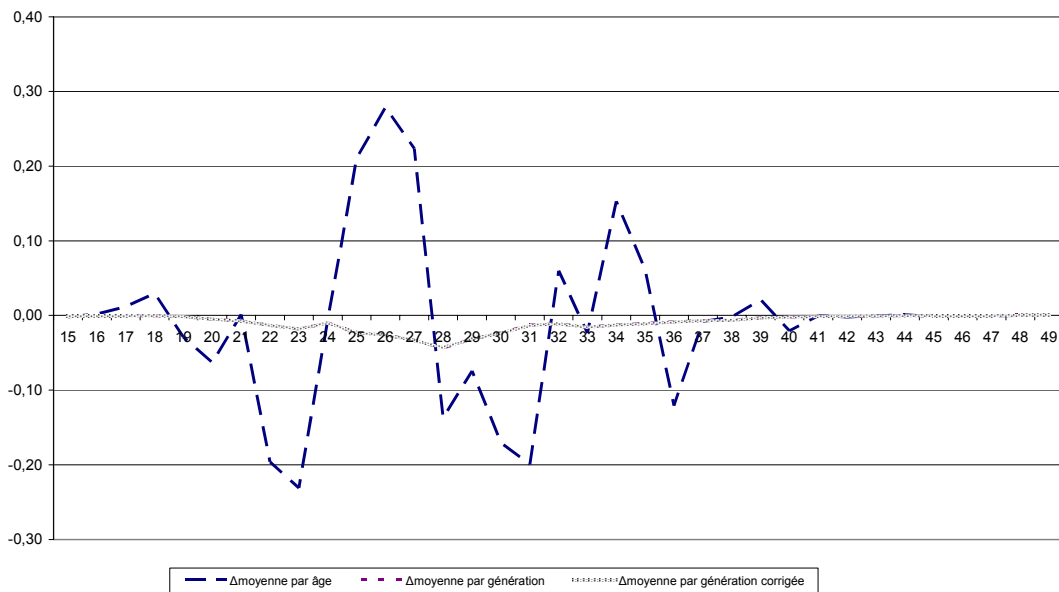
	Indicateur conjoncturel de fécondité
Publiés dans la <i>Situation démographique</i>	1,791
Calculées à partir de la pyramide 1999	1,78
À partir d'une moyenne par âge des données de 1997 à 2002	1,778
À partir d'une moyenne par génération des données de 1997 à 2002	1,777
À partir d'une moyenne par génération des données de 1997 à 2002, corrigée des naissances et des décès	1,777

Graphiques 11a et 11b

Taux de fécondité



Variation du taux de fécondité par rapport à l'estimation 1999



Là encore, l'indicateur synthétique global qu'est l'indicateur conjoncturel de fécondité est peu sensible à la méthode retenue et il existe quelques écarts pour les taux de fécondité par âge, l'utilisation des pyramides « par génération » aboutissant à des résultats plus réguliers. Compte tenu de la tranche d'âges prise en compte, le fait de corriger ou non le biais lié aux décès ne modifie guère les résultats.

3.2. 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 aux effectifs totaux issus du cumul de cinq enquêtes annuelles.

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 les formules déjà utilisées, cela serait le cas si :

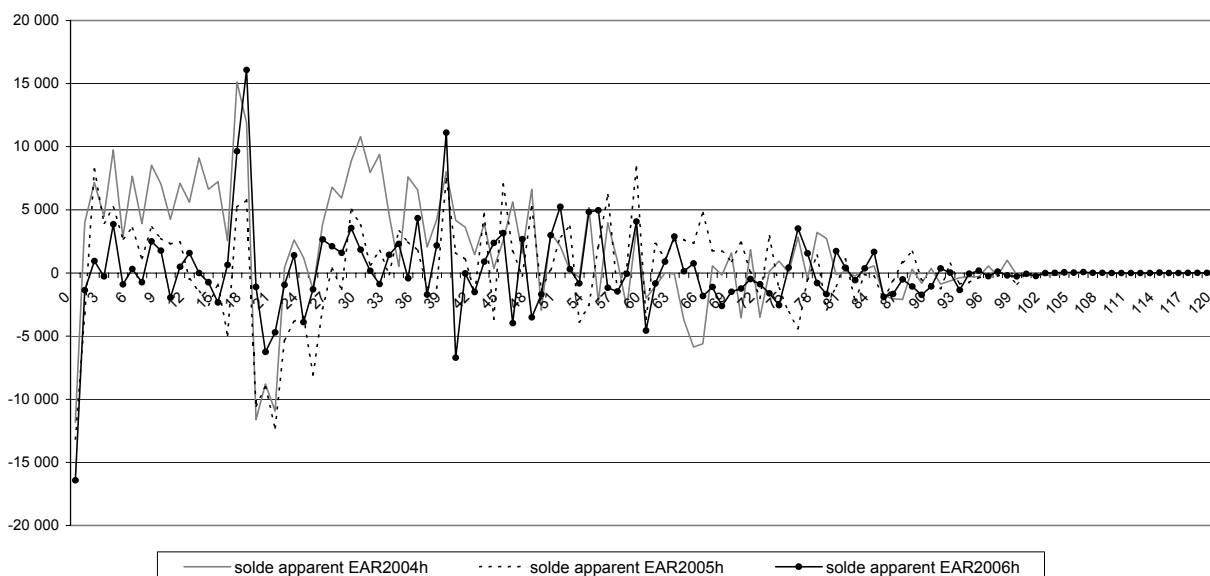
$$e(g,n) > 1/5. [e(g,n-2) + e(g,n-1) + e(g,n) + e(g,n+1) + e(g,n+2)]$$

Par ailleurs, un aléas faible 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 14), et beaucoup plus heurtés que ceux implicitement publiés actuellement (graphique 15).

Graphique 12

Soldes migratoires apparents des hommes

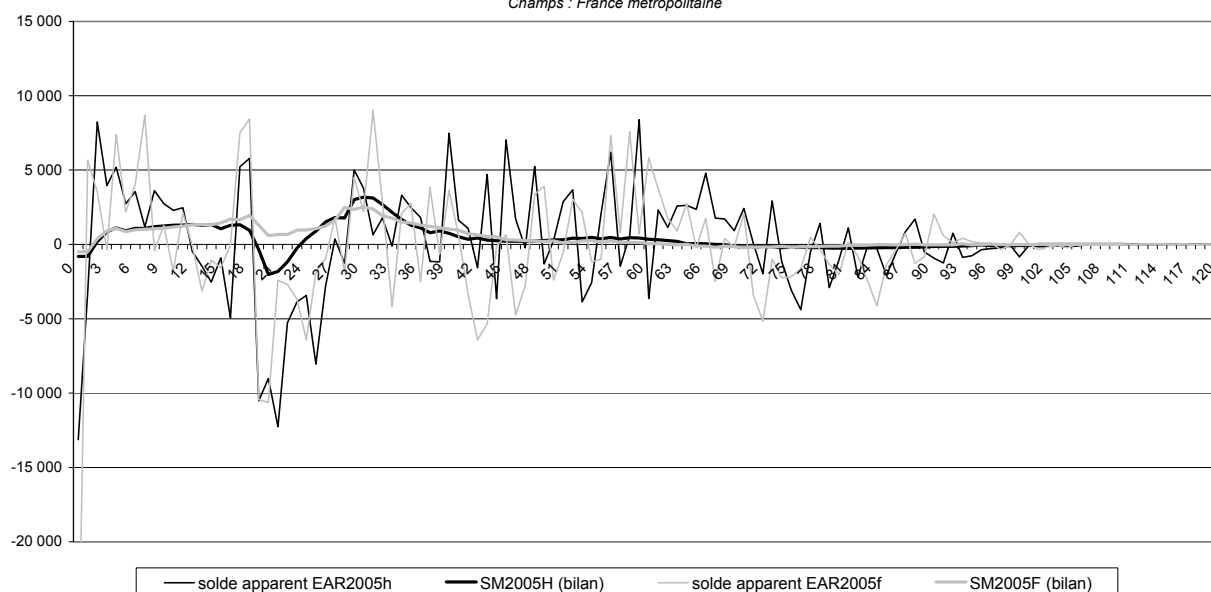
- source EAR - champs : France métropolitaine -



Graphique 12b

Comparaison des soldes migratoires estimés par les EAR
avec ceux du Bilan démographique en 2005

Champs : France métropolitaine



3.3. Le pic des 18 ans

En l'absence d'informations pour estimer l'ampleur de l'artefact sur les effectifs des personnes de 18 révolus au recensement, il n'est pas possible de le corriger. Une pyramide par génération obtenue par cumul de cinq enquêtes annuelles lisse toutefois par construction cet artefact.

3.4. Faut-il décliner la même méthode à tous les échelons géographiques ?

L'utilisation d'une pyramide des âges par cumul moyen 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 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 :

$$\begin{aligned}
 & -\alpha_1(g,n).SM(g,n-2) + (\alpha_1(g,n) + \alpha_2(g,n)).SM(g,n-1) - (\alpha_4(g,n) + \alpha_5(g,n)).SM(g,n) - \alpha_5(g,n).SM(g,n+1) \\
 & - \alpha_1(g,n).e(g,n-2) + \alpha_2(g,n).e(g,n-1) + \alpha_3(g,n).e(g,n) + \alpha_4(g,n).e(g,n+1) + \alpha_5(g,n).e(g,n+2)
 \end{aligned}$$

D'autant plus que, si des estimations nationales permettent d'envisager, quand cela deviendra nécessaire, de corriger des flux migratoires, ce n'est pas envisageable au niveau local, faute d'estimations disponibles sur les soldes migratoires infra nationaux.

Dans ce contexte, l'utilisation d'une pyramide lissée telle que celle issue d'un cumul des enquêtes annuelles selon l'âge peut être suffisante pour la plupart des utilisateurs.

Il convient ici de souligner que, dans ces utilisations infra nationales, le fait d'avoir expliciter les biais possibles entre la pyramide des âges issue du nouveau recensement et la pyramide « réelle » inconnue, supprime sans doute le confort d'utilisation d'un recensement exhaustif ponctuel mais ce confort ne signifiait pas une absence de biais.

D'une part, comme déjà indiqué plus haut, il existait déjà des aléas de collecte. 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.

3.5. S'intéresser à un comportement sociodémographique lié à l'âge, selon l'âge

La pyramide des âges peut être détaillée pour construire, par exemple, 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. Elle peut aussi servir à construire une pyramide des personnes vivant en ménage ordinaire pour caler une enquête auprès des ménages.

Plusieurs approches sont alors possibles. Nous l'illustrons ensuite avec le taux de personnes vivant en ménages ordinaires. En partant de la pyramide des âges par génération présentée ci-dessus, il « suffit » d'appliquer des taux de personnes vivant dans un « ménages ordinaires » par sexe et âge. Il reste alors à choisir la façon de calculer ces taux. On peut en effet:

- 1) Retenir les taux par sexe et âge produits par le recensement à partir du cumul des informations des cinq enquêtes annuelles et de la pondération de cumul, utilisée pour le calcul des populations légales (l'âge est alors l'âge révolu à la date du recensement).
- 2) Calculer des taux par sexe et âge à partir d'une tabulation des résultats du recensement par année de naissance, en utilisant la pondération de cumul et calculant l'âge par différence de millésimes entre l'année médiane de la période de collecte et l'année de naissance.
- 3) Calculer des taux par sexe et âge à partir des résultats distincts des cinq enquêtes annuelles de recensement, en utilisant les pondérations annuelles, et en faire la moyenne, la moyenne pouvant se faire par âge (comme dans l'option 1 avec la pondération de cumul) ou par année de naissance (comme dans l'option 2, avec la pondération de cumul).
- 4) Utiliser directement les résultats des enquêtes annuelles de recensement : les taux appliqués en 2004 sont alors ceux estimés avec l'enquête annuelle de 2004 (pondération annuelle), etc.

Avec l'option 4, l'année 2008 est alors la dernière année connue de façon définitive. Dans les trois autres cas, les taux sont estimés à partir des informations collectées de 2004 à 2008 et datés de 2006 ; la dernière année connue de façon définitive est donc 2006.

L'option 3, avec une tabulation selon l'année de naissance, correspond à celle retenue pour l'estimation de la pyramide des âges des estimations de population.

L'examen des résultats des cinq enquêtes annuelles de recensement (voir graphiques en annexe), montre que :

- l'utilisation d'informations tabulées selon l'année de naissance n'est pas pertinente, les phénomènes de passage dans des établissements étant très liés à l'âge (notamment les cités universitaires, les maisons de retraites, les établissements de santé) ;
- l'utilisation des résultats annuels des recensement entraînerait une certaine instabilité des taux d'une année sur l'autre, qui pourrait s'expliquer, en partie au moins, par le plan d'échantillonnage²².

L'option 1 apparaît donc préférable, d'autant qu'elle a le mérite de s'obtenir facilement à partir de données publiées par l'INSEE : les pyramides des âges des estimations de population d'une part ; la part de la population vivant en ménage ordinaire dans le recensement, à âge et sexe donné.

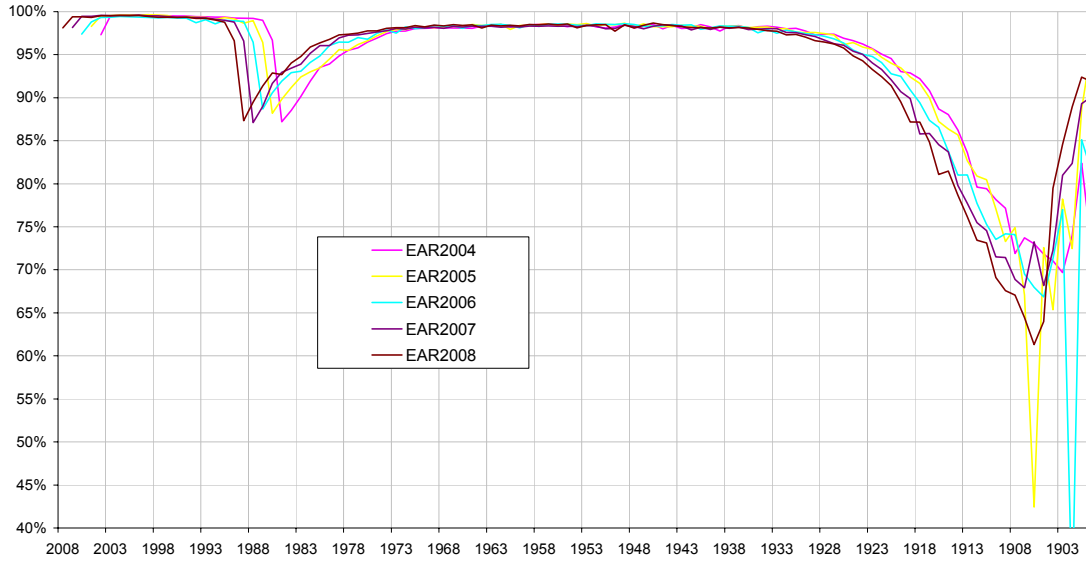
Bibliographie

- [1] 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.
- [2] 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.
- [3] 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.
- [4] « Pour comprendre le recensement de la population », *Insee Méthodes*, hors série, mai 2005.
- [5] 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.
- [6] « 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
- [7] Desplanques G., « Avantages et incertitudes des enquêtes annuelles de recensement », *Population*, volume 63, n°3, 2008, INED.

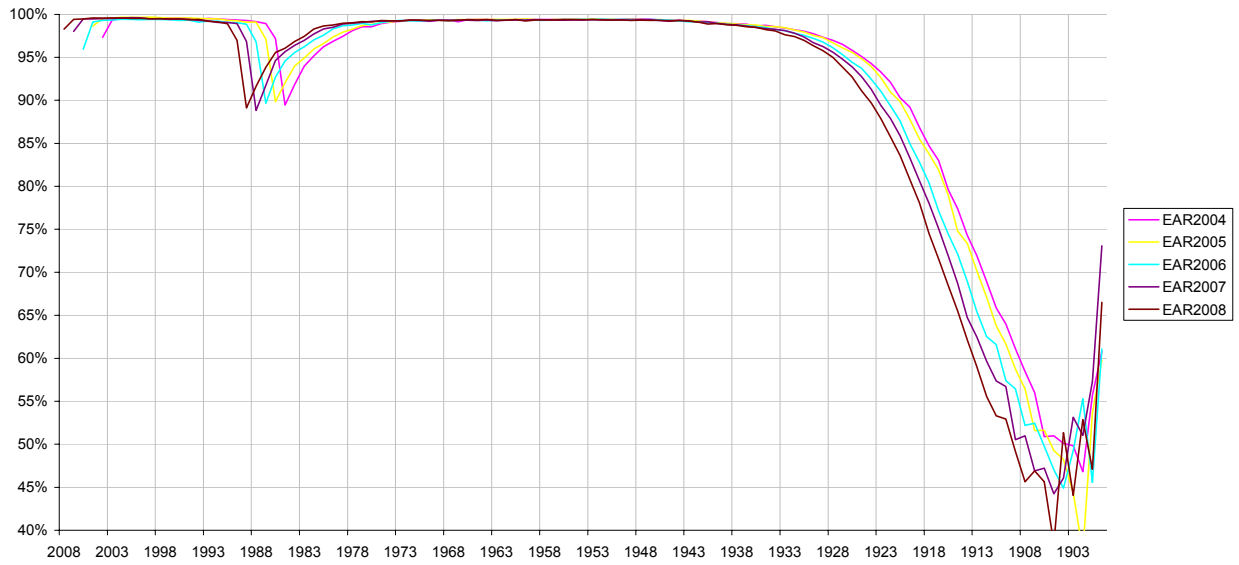
²² Le plan de sondage n'a pas été construit avec l'idée de produire à termes des résultats annuels sur les seules populations hors ménages.

ANNEXE

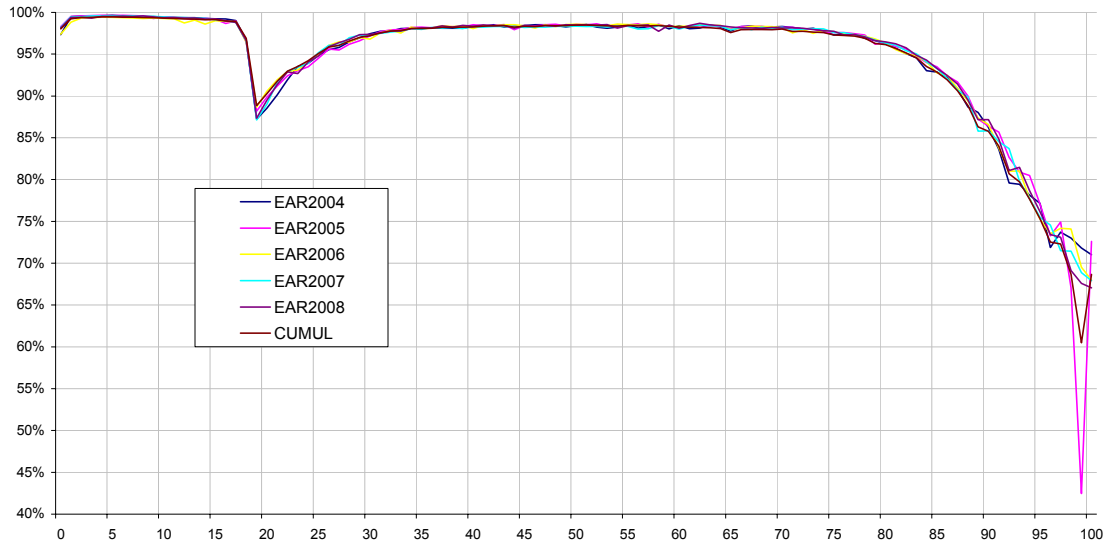
Part des hommes vivant en ménages, selon leur année de naissance
Résultats par enquête annuelle, pondération annuelle, France métropolitaine



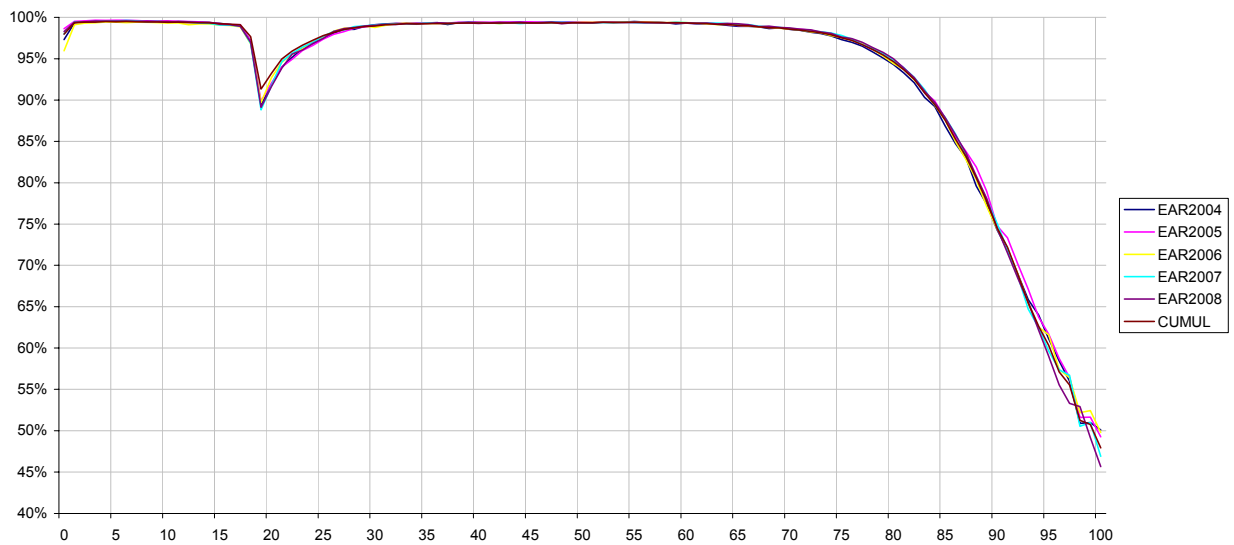
Part des Femmes vivant en ménages, selon leur année de naissance
Résultats par enquête annuelle, pondération annuelle, France métropolitaine



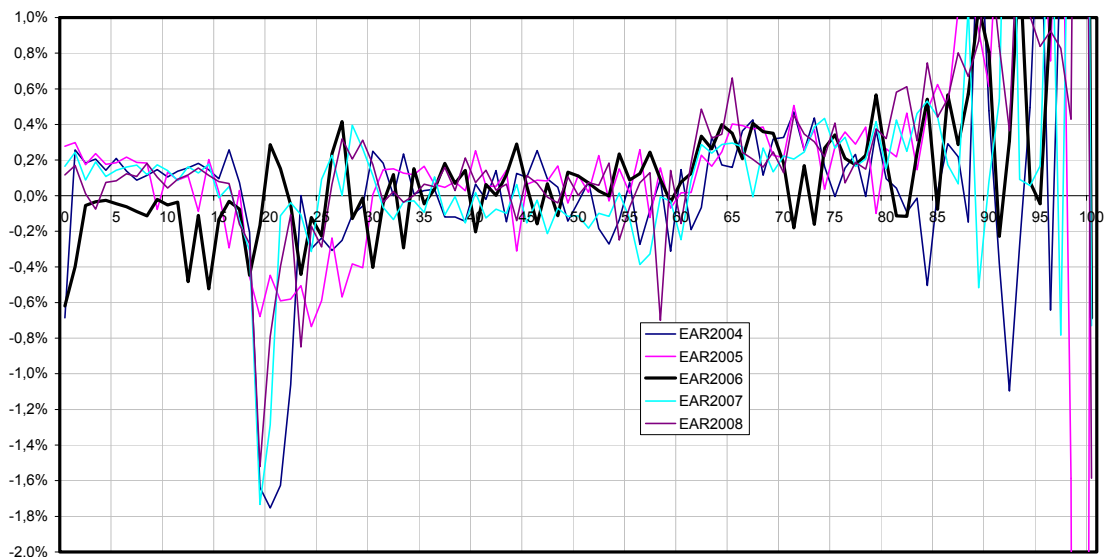
Part des hommes vivant en ménages, selon leur âge au recensement
 Résultats des enquêtes annuelles avec pondération annuelle, le cumul avec pondération de cumul,
 France métropolitaine



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



**Hommes, écart entre le % en ménage dans l'EAR (poids EAR, âge révolu au 1er janvier)
et le % en ménages dans cumul (poids cumul, âge révolu au RP)**
France métropolitaine



**Hommes, écart entre le % en ménage dans l'EAR (poids EAR, âge révolu au 1er janvier)
et le % en ménages dans cumul (poids cumul, âge révolu au RP)**
France métropolitaine

