

L'échantillon démographique permanent de l'Insee

 Christine COUET*

Depuis bientôt quarante ans, l'Insee accumule des informations sociodémographiques sur un échantillon d'individus représentatifs de la population de la France métropolitaine. L'échantillon démographique permanent (EDP) est aujourd'hui enrichi chaque année des données en provenance des enquêtes annuelles de recensement et de l'état civil. Les individus de ce panel sont sélectionnés de manière à assurer la représentativité et la permanence de cet échantillon au cours du temps. Actuellement, environ 900 000 histoires individuelles sont ainsi retracées, offrant d'importantes possibilités d'études démographiques et sociales. Un projet est également en cours visant à élargir et enrichir ce panel afin de le rendre plus opérationnel.

À la fin des années soixante, l'Insee a entrepris de garder une trace des informations qu'il accumulait au fil du temps, de nature essentiellement démographique, en créant un échantillon représentatif permanent, autrement dit un panel, de la population française. Le recensement de 1968 a été l'occasion d'initialiser cette collecte. C'est ainsi qu'est né l'EDP : son caractère évolutif en fera une importante base de données individuelles.

Au niveau international, à l'époque, cette initiative n'était pas isolée : son lancement a coïncidé en effet avec la mise en place du « longitudinal study » (LS), l'équivalent outre-manche de l'EDP, qui s'appuie sur les recensements britanniques de 1971, 1981, 1991 et 2001. De nombreux contacts ont été établis au cours de ces périodes avec les équipes de la City University de Londres et du département « Population and Health » de l'Office national de statistiques (ONS) britannique, qui gèrent conjointement la production et l'exploitation du LS.

L'EDP n'a cessé de s'enrichir depuis les années 1960

Actuellement, l'échantillon français est riche de plus de trente années d'informations portant sur près de 900 000 individus, résidant ou ayant



Source : droits réservés

L'EDP retrace 900 000 histoires individuelles

résidé sur le territoire métropolitain. L'introduction de données issues du recensement de la population de 1999 marque la fin d'une étape dans l'histoire de cet échantillon : celle où l'EDP s'appuyait sur des recensements exhaustifs, ceux de 1968, 1975, 1982 et 1990 (et donc, pour finir, celui de 1999), complétés par des données de l'état civil. Depuis 2004, les enquêtes annuelles de recensement visent à recenser en moyenne un individu sur sept. Désormais, l'enrichissement de l'EDP s'effectuera donc de façon partielle et à un rythme annuel.

Comme l'échantillon britannique, l'EDP français suit des individus d'un recensement à l'autre et s'enrichit, le cas échéant, des données en provenance de l'état civil. Son intérêt est de mettre en perspective des informations qui, bien qu'existantes par ailleurs (recensement et état civil), n'étaient pas reliées entre elles. L'EDP, en rapprochant plusieurs sources statistiques, permet aussi de faire des études méthodologiques,

* Christine Couet travaille à la division enquêtes et études démographiques du département de la démographie de l'Insee.

difficiles voire impossibles à réaliser avec une seule source : on peut par exemple observer les erreurs de mémoire des individus, ou encore les distorsions liées aux conditions de collecte de l'information. Autre intérêt de l'EDP : il est actuellement riche en renseignements sur les mouvements géographiques, professionnels et démographiques des individus qui le composent.

Les données individuelles sont anonymes et accessibles aux chercheurs

La base d'études de l'EDP ne contient aucune variable directement identifiante, comme le nom, le jour de naissance, ou encore le numéro d'inscription au répertoire national d'identification des personnes physiques.

L'accès aux données a été favorisé par le travail d'un groupe d'exploitation, dont la vocation était de promouvoir la réalisation d'études utilisant l'EDP. La mise en place de conventions d'accueil a permis à des chercheurs et chargés d'études travaillant en dehors de l'Insee, de disposer des données dans les locaux de l'institut (quatre personnes ont été ainsi accueillies en 2005).

Comment est constitué l'échantillon ?

La mise en place de l'EDP a été motivée notamment par la volonté d'avoir un échantillon qui puisse être enrichi continuellement par les nouvelles informations disponibles. Il était alors nécessaire de simplifier au maximum le mode de tirage de l'échantillon, tout en maintenant sa représentativité transversale (c'est-à-dire à tout moment). Le choix du critère de sélection des individus s'est porté sur le jour de naissance : l'échantillon rassemble tous les actes d'état civil et toutes les informations consignées dans les bulletins de recensement successifs, des personnes nées l'un des quatre jours de référence : les 1^{er}, 2, 3 et 4 octobre de chaque année. Ainsi, à tout moment, la base est proche d'un

échantillon représentatif au 1/100^e (4/365^e) de la population résidant en métropole.

Le renouvellement de l'échantillon est assuré de manière automatique par les naissances ou par la venue en France métropolitaine de nouveaux immigrants signalée à l'occasion d'un recensement ou d'un événement enregistré dans un acte de l'état civil.

À l'inverse, le suivi d'un individu cesse, de fait, en cas de décès ou à la suite d'une émigration hors métropole. Cependant, la trace de cet individu persiste dans l'échantillon avec le détail des événements démographiques qui ont jalonné son parcours sur le territoire métropolitain.

Nature de l'information collectée

Deux sources alimentent le fichier de l'EDP. L'une, le recensement de la population, caractérise les individus à des dates fixées. L'autre, l'état civil, recueille des informations, individuelles elles aussi, à l'occasion d'événements majeurs de la vie familiale.

En premier lieu, le fichier de l'EDP rassemble, pour chaque individu de l'échantillon, les informations consignées dans le bulletin individuel rempli par les intéressés lors des cinq derniers recensements exhaustifs de population de 1968, 1975, 1982, 1990 et 1999. Les informations recueillies sont d'ordre démographique et social : sexe, état matrimonial, date et lieu de naissance, nationalité, département et commune de la résidence actuelle et antérieure, diplôme, catégorie sociale, emploi et activité, structure du ménage et de la famille, etc. Des informations issues de la feuille de logement du recensement viennent aussi compléter ces données.

Par ailleurs, l'introduction d'événements en provenance de l'état civil contribue également à enrichir l'histoire familiale des individus de l'échantillon. Parmi les événements démographiques majeurs, faisant

d'abord l'objet d'un enregistrement à l'état civil puis d'une mise à jour de l'EDP, on peut citer :

- la naissance d'un individu du champ EDP (qui a lieu, par définition, l'un des quatre jours de référence),
- son éventuelle reconnaissance par un parent,
- le mariage de parents, susceptible de légitimer l'individu EDP,
- son ou ses propres mariages,
- la naissance de ses enfants,
- la reconnaissance de ses enfants,
- son décès.

Pour l'essentiel, ces bulletins relatent les circonstances de l'événement (lieu et date) et fournissent des caractéristiques de la personne (sexe, date et lieu de naissance, commune, département et pays de domicile, état matrimonial, catégorie sociale, nationalité).

L'actuel EDP en chiffres

En janvier 2003, l'EDP rassemble près de 900 000 enregistrements individuels, dont plus de 600 000 personnes présentes au recensement de la population de 1999. Parmi eux, 260 605 individus ont été présents aux cinq derniers recensements (tableau 1). Ce chiffre est suffisamment élevé pour que l'on puisse faire des études, y compris sur des populations spécifiques (chose que ne permettrait pas une enquête ne portant que sur quelques milliers d'individus).

Par ailleurs, 128 493 enregistrements correspondent à des individus déjà décédés, dont on a souhaité néanmoins garder la trace, et 27 752 à des personnes jamais recensées, donc entrées dans l'EDP par un bulletin d'état civil les concernant. Sur l'ensemble de la base, près de 752 000 individus sont nés en France, soit 85 % du total, contre 135 000 nés à l'étranger. La population est décrite par 1 622 variables, dont un nombre non négligeable reprend une même situation à travers des sources diversifiées.

Tableau 1 : Présence des individus EDP aux différents recensements

Effectifs EDP présents au...		et également présents aux recensements précédents de...			
		1990	1982	1975	1968
RP99	617 195	497 699	399 915	327 705	260 605
%	100	81	65	53	42
RP90	588 428		470 808	387 580	309 839
%	100		80	66	53
RP82	551 211			450 965	362 682
%	100			82	66
RP75	526 014				418 025
%	100				80
RP68	485 009				

Lecture : sur les 617 195 individus présents au recensement de la population (RP) de 1999, 497 699 (soit 81 %) étaient déjà présents à celui de 1990, 399 915 (soit 65 %) l'étaient aux recensements de 1990 et de 1982, 327 705 (53 %) à ceux de 1990, 1982 et 1975, et 260 605 (42 %) ont répondu aux cinq recensements.

de recueillir pour tous les individus de l'échantillon EDP – contrairement à ce qui avait été fait en 1982 – les variables brutes et redressées concernant la personne et son environnement.

Malheureusement, les variables brutes traditionnelles du recensement de 1990 récupérées pour l'EDP sont moins nombreuses qu'aux recensements précédents. En revanche, plusieurs variables nouvellement créées, traitant notamment des questions d'emploi, ont été introduites à cette date (indicateur d'emploi, ancienneté de recherche d'emploi, position professionnelle et profession détaillée, temps de travail et conditions d'emploi, etc.).

Avec le recensement de la population de 1999, de nouveaux critères d'étude sur des sujets d'actualité ont été introduits, tels que l'année d'arrivée en métropole pour les personnes nées hors métropole² ou le niveau d'études de l'individu ou encore des renseignements concernant le mode de vie, l'aménagement du lieu d'habitation et le moyen de transport utilisé pour se rendre à son activité professionnelle (tableau 2 page suivante).

Circuits de collecte et supports des données de l'EDP

De 1968 à 1989, la collecte des données de l'EDP a été effectuée de façon entièrement manuelle. À cette époque, l'entrée d'un individu dans l'échantillon entraînait l'ouverture d'un « dossier papier » dans la direction régionale (DR) de l'Insee du lieu de naissance de l'intéressé. L'approche du recensement de 1990 a marqué les débuts de l'automatisation des circuits de mise à jour de l'EDP.

1. En effet, autant les redressements sont indispensables pour améliorer la qualité des données au niveau global, autant, au niveau individuel, ils modifient les données déclarées, empêchant alors des analyses de trajectoires.
2. À partir de 2004, on s'intéresse à l'année d'arrivée en France (et non plus en métropole) des individus nés à l'étranger (et non plus hors de métropole).

Les difficultés rencontrées lors de la collecte des données de l'EDP

Pour suivre convenablement les trajectoires individuelles dans l'EDP, il convient de recueillir les données du recensement sous leur forme « brute », c'est-à-dire avant redressement¹. Mais, pour juger de la qualité de la source, il faut pouvoir comparer le profil de l'individu tel qu'il apparaît dans l'EDP à la codification retenue au recensement. Il importe donc de disposer aussi des variables redressées des non-réponses, marquées comme telles dans l'échantillon. Cette double exigence n'a pas toujours été respectée. Ainsi, les bulletins individuels collectés pour l'EDP à l'occasion des *trois premiers recensements* ont été entièrement recodés à partir des données brutes, après ressaisie des bulletins individuels. Les autres documents du recensement, en revanche, n'ont pas été saisis, en particulier les feuilles de logement, ce qui explique l'absence de données brutes avant 1990, concernant la famille et le logement.

Mais la nature et l'étendue de l'information recueillie sur la situation au moment d'un recensement dépendent aussi de l'éventualité d'un appariement entre le fichier de recensement et l'EDP. Signalons :

– l'absence d'appariement après ressaisie en 1968, sachant que de nombreuses variables de ce recensement n'ont été saisies que par un sondage au quart. De ce fait,

les variables redressées du recensement de 1968 font défaut à tous les niveaux, tant individuel que familial ;

– en 1975, en revanche, le recensement a été exploité de façon exhaustive et l'enregistrement de l'EDP a pu être apparié au recensement, permettant l'introduction de variables redressées des non-réponses, en ce qui concerne l'individu mais aussi son environnement (famille, ménage, immeuble, logement et localisation géographique) ;

– en 1982, le champ de la collecte des données sur l'environnement familial s'est réduit. En effet, les données concernant la famille et le ménage n'ont été traitées que dans l'« exploitation lourde » du recensement. Or, celle-ci n'a été réalisée que sur un sondage au quart de la population. Dès lors, on n'a pu récupérer l'information que pour les individus EDP appartenant à l'échantillon du quart.

Élargissement progressif des données du recensement collectées pour l'EDP

À partir du recensement de 1990, l'échantillon démographique permanent a été intégré dans la chaîne de traitement du recensement, afin d'automatiser au maximum l'enrichissement du premier.

De plus, la décision d'inclure dans l'« exploitation lourde » du recensement tous les logements contenant au moins un individu EDP a permis

Tableau 2 : Nature et étendue de l'information collectée aux différents recensements de la population, en vue d'alimenter l'EDP

	Données brutes		Données redressées	
	Niveau individu	Niveau famille, logement	Niveau individu	Niveau famille, ménage, logement, immeuble
RP68	oui	non	non	non
RP75	oui	non	oui	oui
RP82	oui	non	oui	oui sur 1/4
RP90	oui	oui	oui	oui
RP99	oui	oui	oui	oui

Lecture : les données brutes issues du recensement de la population de 1968 existent en ce qui concerne les individus mais pas la famille et le logement.

Le projet de rénovation visait à moderniser le support des dossiers individuels, le support magnétique se substituant progressivement au papier. Ainsi, la création d'une base de données allégeait les contraintes de maintenance et favorisait la réorganisation des circuits. La gestion des bulletins pouvait rester dans la DR de collecte et éviter ainsi tout transfert de bulletins vers la DR du lieu de naissance de l'intéressé.

Ce projet permettait également de greffer la mise à jour automatique de l'EDP à l'extrémité des chaînes de traitement des bulletins, après leur passage à travers les circuits d'exploitation de l'état civil et du RNIPP³. Un tel dispositif assurait à la base EDP une mise à jour en continu, cohérente avec le traitement général de l'état civil.

Grâce à ces progrès dans la gestion de l'EDP⁴, il devenait possible de mettre à disposition des utilisateurs des fichiers d'études, mis à jour chaque année, et donc sans retard excessif par rapport aux événements

3. Répertoire national d'identification des personnes physiques.

4. Actuellement, la responsabilité de la gestion de l'EDP constitue l'une des missions du pôle répertoire et fichiers démographiques de la direction régionale de l'Insee des Pays de la Loire. Les relations entre le pôle et le département de la démographie sont définies par une convention signée en décembre 2001. Les missions confiées au pôle portent, notamment en ce qui concerne l'EDP, sur la gestion de l'application (intégration des données d'état civil), le pilotage de l'intégration des données des recensements de la population, le traitement des cas litigieux et la mise à disposition des fichiers et leur documentation.

5. Numéro d'inscription au RNIPP.

éventuellement intervenus dans la vie des individus de l'échantillon. Enfin, pour répondre aux besoins très différents des gestionnaires et des chercheurs, il a été décidé de créer deux bases de données distinctes, une base de gestion d'une part, et une base d'études utilisable au moyen du logiciel SAS d'autre part. En pratique, la première version de la base d'études a été disponible en 1995. Elle rassemblait les informations sur la période 1968-1990, tirées des quatre recensements de 1968, 1975, 1982 et 1990 et de l'état civil de 1968 à 1989.

Abandon momentané de la mise à jour de l'EDP, pour deux jours de naissance sur quatre

La marche vers l'automatisation s'est faite progressivement, en passant par des périodes de transition au cours desquelles la priorité était donnée à la récupération sur support magnétique des données passées. Ainsi, l'intégration dans la base de l'état civil des années 1982 à 1989 s'est heurtée en 1993 à un manque de moyens des directions régionales. Il a alors été décidé de limiter le champ de la collecte aux individus nés au cours de deux des quatre jours initiaux du champ EDP. C'est ainsi que les données d'état civil de la période 1982-1989 n'ont été intégrées que pour les individus dont l'anniversaire tombait le 1^{er} ou le 4 octobre. Pour les autres, ceux nés le 2 ou le 3 octobre, l'insertion informatique des bulletins a été annulée

et l'histoire familiale de ces individus ne devait connaître de prolongement qu'à travers les recensements de population. Ainsi, le nouveau champ restreint à deux jours anniversaires était également adopté pour les mises à jour ultérieures.

Le manuel du gestionnaire de l'EDP de février 2001 souligne en page 8 qu'« en raison de moyens insuffisants, les bulletins d'état civil des individus nés les 2 et 3 octobre ont été, selon les départements, intégrés ou non sur la période 1982-1989 et qu'aucun n'a été intégré sur la période 1990-1997 ». Mais en pratique, les confrontations périodiques de la base EDP et du RNIPP sur le champ EDP à quatre jours ont permis d'intégrer bon nombre d'événements, généralement sous une forme allégée. Par exemple, en suivant le nombre de décès annuel par jour de naissance, on observe que la majeure partie des décès survenus entre 1982 et 1997 a été intégrée, sans qu'il y ait toutefois mention dans la base des caractéristiques de l'individu au moment du décès.

En 1998, il a été décidé de reprendre le suivi de l'état civil sur la totalité du champ EDP à quatre jours.

L'intégration des données du recensement de 1999

À la suite du recensement de la population de mars 1999, l'objectif prioritaire a été d'enrichir l'EDP au moyen des informations collectées sur les 633 243 individus fraîchement recensés, nés l'un des quatre jours anniversaires du champ de l'EDP. Cette récupération ne pouvait se faire qu'au moyen d'un appariement des deux sources sur la base du NIR⁵. Il fallait en effet pouvoir identifier la personne recensée. Or, les données du recensement ne sont pas suffisantes pour permettre cette identification.

On a donc imaginé un dispositif permettant de résoudre ce problème, mais qui s'est avéré, par la force des choses, relativement complexe : on a eu recours à une base-image réa-

lisée par lecture optique et comportant, sous forme manuscrite, l'état civil de l'individu (nom, prénoms, sexe, date de naissance, commune, département ou pays de naissance). Ces données, une fois ressaisies par vidéo-codage et sélectionnées sur le champ EDP, ont été fusionnées, au moyen d'un numéro d'image, à l'ensemble des données statistiques du recensement (brutes, redressées et codifiées). L'identification de la personne devenait alors possible au moyen du RNIPP avec, comme moteur de la recherche, ses nom et prénom, sa date et son lieu de naissance. La récupération de son NIR permettait ensuite l'insertion des données nouvelles dans la base existante.

Le programme d'identification automatique a parfois entraîné des rejets. Le nombre total de litiges ainsi rencontrés s'est élevé à près de 75 000, soit un taux de litige d'environ 12 %. Un traitement manuel a pris le relais, conduisant à l'intégration d'un grand nombre de ces dossiers et à l'abandon de 4 700 d'entre eux. En final, le taux d'échec est donc très faible, inférieur à 1 %.

À l'issue de ces opérations, un premier fichier d'études a été mis à disposition au centre national informatique d'Orléans de l'Insee, début 2003. Il rassemble les données des recensements de 1968 à 1999, les naissances de 1967 à 1999, les mariages et les décès de 1968 à 1997.

Un second fichier d'études a été constitué courant 2003. Il comprend, en plus des informations existant dans l'échantillon de janvier de la même année, la récupération de renseignements à propos des décès manquants, des données complémentaires sur l'état civil de 1998 et 1999 (reconnaisances et décès) et plusieurs indicateurs de zonage (typologie socioéconomique des quartiers et communes de N. Tabard, aires urbaines, zones urbaines sensibles,...), afin d'inclure les spécificités locales au domaine d'études. La version actuelle du fichier « études »

est complétée par l'ensemble des données d'état civil jusqu'en 2003.

Parallèlement, pour faciliter l'exploitation du fichier, un nouveau manuel de l'utilisateur de l'EDP a été réalisé à des fins de statistiques et d'études.

Jour de naissance : effet d'attraction du premier jour du mois

L'effectif des individus de l'échantillon varie légèrement selon le jour de naissance. Quel que soit le recensement considéré, il se trouve toujours un peu plus d'individus nés le 1^{er} octobre que les 2, 3 ou 4 du même mois (graphique 1). Ceci pourrait traduire l'attraction qu'exerce le premier jour du mois chez les individus qui ignorent leur date de naissance. C'est en effet parmi les individus nés le 1^{er} octobre que l'on trouve le pourcentage le plus élevé de personnes nées à l'étranger, notamment dans les pays dont on sait que l'état civil n'a pas la même qualité qu'en France. Ainsi, les écarts d'effectifs selon le jour de naissance sont révélateurs de petites imperfections dans la réalisation de l'échantillon.

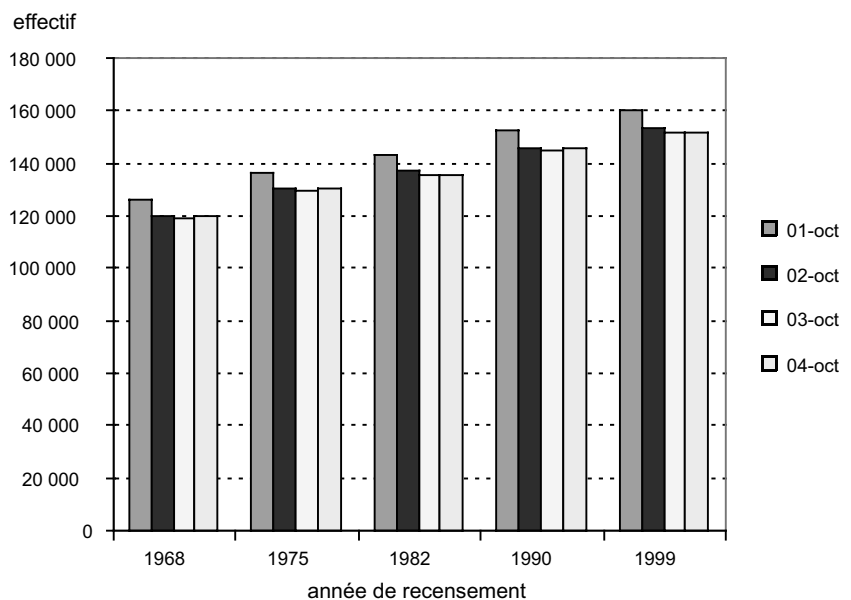
Le taux de couverture de l'EDP s'améliore lentement

En mars 1999, l'échantillon représentait 1,05 % de la population de la France métropolitaine, avec 617 195 individus EDP présents au RP99 sur 58 513 700 personnes recensées en métropole.

Le taux de couverture de l'EDP augmente un peu à chaque nouveau recensement et tend à croître au fil des générations (graphique 2 page suivante). Ceci suggère des progrès dans l'identification des individus EDP au cours du temps et, de la part des intéressés les plus jeunes, nouvellement entrés dans l'échantillon, moins d'erreurs dans la déclaration de leur date de naissance ou d'ignorance vis à vis de celle-ci (problème essentiellement rencontré chez des personnes nées à l'étranger).

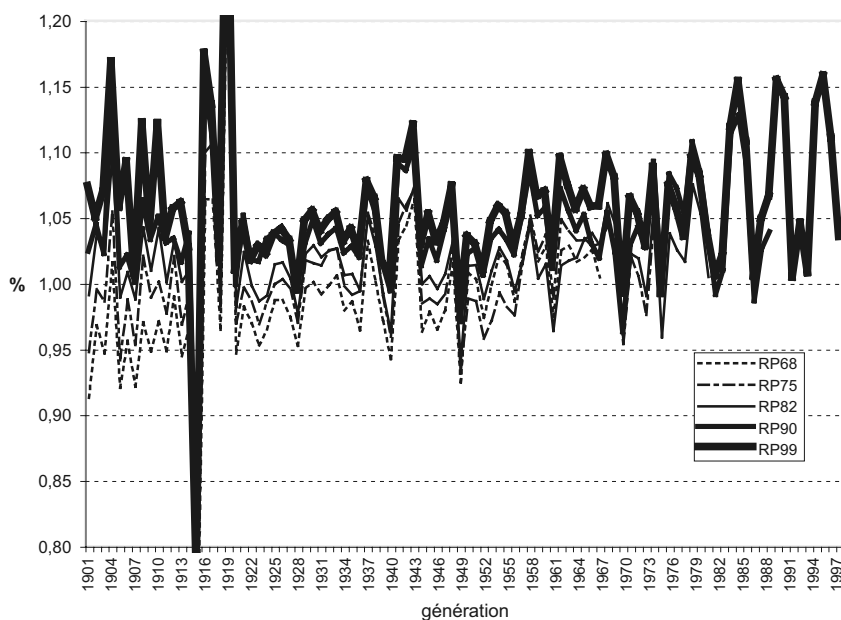
Ce taux de couverture par génération correspond, on l'a vu, à la part que représente le nombre de naissances ayant eu lieu au cours des 4 jours de référence parmi les 365 ou 366 jours de l'année. Si on néglige la saisonnalité des naissances (le fait qu'il y ait par exemple plus de naissances en

Graphique 1 : Effectifs d'individus EDP présents aux différents recensements de la population selon leur jour de naissance



Lecture : parmi les individus présents au recensement de la population de 1999, environ 160 000 sont nés le 1^{er} octobre, contre un peu plus de 150 000 les 2, 3 ou 4 octobre.

Graphique 2 : Taux de couverture dans l'EDP selon le recensement considéré et l'année de naissance de l'individu



Lecture : Les individus de l'échantillon EDP, recensés en 1999, représentent 1,15 % de l'effectif de leur génération recensé en 1999 s'ils sont nés en 1990, et seulement 1 % s'ils sont nés en 1992. Cet écart s'explique par le fait qu'en 1992, contrairement à 1990, on compte un dimanche parmi les quatre jours de référence de l'échantillon.

[Le creux de 1915 (le pic de 1919) s'explique par un fort déséquilibre vers le bas (vers le haut) du poids relatif de l'effectif des individus EDP nés cette année-là sur l'ensemble des naissances de l'année considérée, perturbant ainsi fortement la couverture de l'échantillon.

En effet, la conception de ces enfants EDP, qui remonte à neuf mois plus tôt, est intervenue au moment où les hommes se sont retrouvés brutalement moins nombreux (plus nombreux) à la maison en raison de départs massifs à (de retours massifs de) la guerre.]

mars qu'en octobre), l'EDP devrait, en moyenne, concerner 1,1 % de la population (4/365), soit une couverture légèrement supérieure à ce qui est observé. Ceci semble accréditer l'idée d'un taux d'omission dans l'échantillon.

Au cours des vingt dernières années, les fluctuations du taux de couverture de l'EDP par génération suivent assez bien celles du rapport du nombre de naissances observées sur le territoire métropolitain entre les quatre jours de référence et l'ensemble de l'année. Ces variations s'expliquent en grande partie par la présence ou non d'un dimanche parmi les quatre jours de référence d'une année donnée : la présence d'un dimanche se traduit en effet par une réduction sensible du nombre journalier de naissances, depuis que le système de soins français est organisé pour éviter les naissances ces jours-là. Les fluctuations s'am-

plifient chez les générations récentes du fait du rôle de plus en plus perturbateur des dimanches sur le niveau des naissances.

Le suivi longitudinal des individus au travers des recensements successifs

L'un des intérêts de l'EDP est de pouvoir expliquer les comportements d'individus appartenant à des générations voisines – par exemple en matière de migrations, de mobilité professionnelle, de fécondité... – au moyen d'informations recueillies aux recensements et de se livrer à des comparaisons intergénérationnelles.

L'échantillon permet de retracer l'histoire de nombreuses générations allant de 1870 à nos jours. Mais ces générations sont inégalement représentées dans l'EDP en terme d'effectifs, et diversement concernées par les cinq recensements.

Parce que l'échantillon a commencé par une photographie prise en 1968, seuls les survivants à cette date définissent les effectifs des générations les plus anciennes.

Cependant, à partir de la génération 1900, la part de personnes décédées devient suffisamment faible pour que le nombre d'individus EDP par génération soit proche de celui des naissances de l'état civil survenues les quatre jours de référence, corrigé des flux migratoires.

Ensuite, depuis 1968, chaque période intercensitaire voit l'arrivée massive des générations nouvelles et l'entrée plus modeste d'immigrés de tous âges.

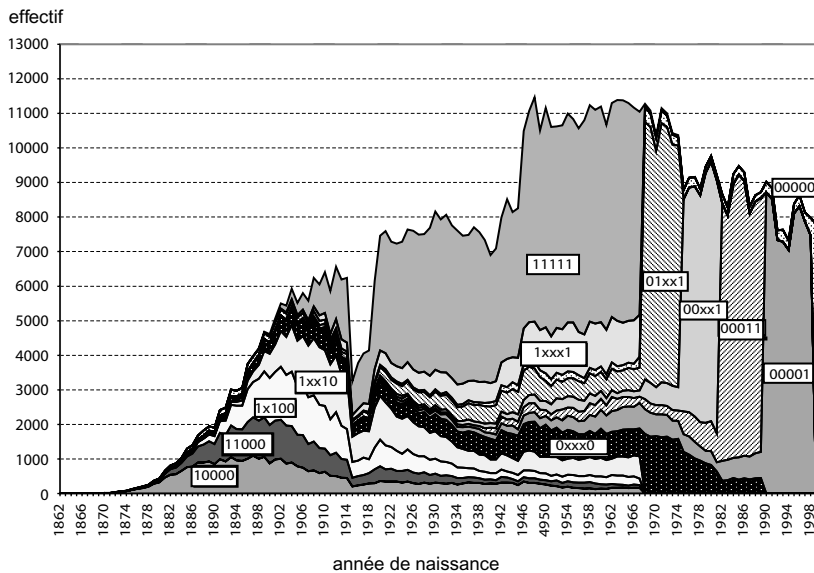
Le graphique 3 ci-après permet de visualiser les effectifs présents et absents aux différents recensements selon les générations.

Il est construit autour d'un indicateur de présence aux recensements. Partant des 32 combinaisons possibles (2^5 cas selon que l'individu est présent ou absent à chacun des cinq recensements), on a regroupé ces combinaisons en 12 situations plus générales et relativement homogènes.

L'indicateur est alors un nombre composé de cinq chiffres, dont chacun des chiffres successifs concerne un des cinq recensements successifs : en se déplaçant de la gauche vers la droite de l'indicateur, on a donc une information qui va du recensement de 1968 à celui de 1999, information symbolisée soit par le chiffre 1 en cas de présence de l'individu au recensement en question, soit par le chiffre 0 en cas d'absence. Pour signifier que l'individu est « présent ou absent » au recensement en question (autrement dit, la réunion des deux événements possibles « présent » ou « absent »), on a choisi pour code la lettre x.

Ainsi, pour tous les individus nés après 1968, individus qui, par définition, ne pouvaient être présents au recensement de cette année-là, l'indicateur a pour premier chiffre un 0 ; de la même façon, chez les per-

Graphique 3 : Effectifs des individus appartenant à l'échantillon selon leur génération et leur présence aux divers recensements (y compris les individus décédés depuis leur intégration)



Lecture :

11111 désigne les individus présents aux cinq recensements (remarque : pratiquement aucune personne née avant 1900 n'est présente en 1999).

1xxx1 désigne les individus présents aux deux recensements de 1968 et 1999, et dont la présence n'est pas systématique aux 3 autres recensements.

1x100 désigne les individus présents au recensement de la population de 1968 et de 1982, pas forcément présents à celui de 1975 et absents en 1990 et 1999.

01xx1 désigne les individus absents au recensement de 1968, présents à ceux de 1975 et 1999, et présents ou pas à ceux de 1982 et 1990.

00000 désigne les individus entrés dans l'EDP via l'état civil.

Enfin, l'EDP sert parfois de base préliminaire pour tirer d'autres échantillons. Son avantage est de mettre à disposition des chercheurs des informations issues des recensements qui sont susceptibles d'expliquer toutes sortes de comportements. Par exemple, le rapprochement du fichier électoral et de l'échantillon démographique permanent sert de base pour sélectionner régulièrement les individus qui vont contribuer à l'étude de la participation électorale.

Finalement, grâce à sa composition, l'EDP permet d'étudier à grande échelle (pour 900 000 individus) de multiples histoires familiales et sociales, et ce, à des conditions plutôt avantageuses. En effet, la réalisation initiale de cet échantillon, et depuis, sa mise à jour, qui utilisent des données déjà collectées, se font à un coût bien moindre que s'il fallait réaliser une enquête. De plus, la collecte de l'information n'oblige pas à solliciter continuellement les personnes concernées, puisqu'elle ne nécessite aucune interview supplémentaire. Enfin, les données recueillies sont de meilleure qualité car le mode de collecte ne fait pas appel à la mémoire, bien souvent trompeuse.

Contraintes spécifiques à prendre en compte lors de l'utilisation de l'EDP

Malgré l'insertion régulière dans l'EDP d'événements en provenance de l'état civil, l'histoire individuelle des « personnes EDP » ne se déroule pas totalement en continu. Elle se fait plutôt par étapes, avec parfois des sauts de plusieurs années dans le temps.

Par exemple, un changement de lieu de résidence ou de profession n'est pas forcément lié à la survenue d'un événement démographique. Si un recensement permet *in fine* d'entériner ce changement, il ne permet pas de le dater et il n'écarte pas le risque d'occulter d'autres mutations inter-censitaires. De même, les faits survenus à l'extérieur du territoire métropolitain sont ignorés.

sonnes nées après 1975, ce sont les deux premiers chiffres de l'indicateur qui sont égaux à 0, etc.

Le graphique 3 permet donc de comprendre les relations existant entre le suivi longitudinal (l'appartenance à l'effectif d'une génération) et la collecte d'informations transversales (la répartition de cet effectif selon une présence éventuelle aux recensements). Cette information est intéressante pour l'étude des comparaisons entre générations.

Les types d'études que l'EDP permet de réaliser

La vocation principale de l'EDP est de permettre de reconstituer des trajectoires individuelles, d'analyser les changements dans la situation des individus entre deux recensements (le plus souvent en matière de mobilité géographique ou socioprofessionnelle), d'expliquer le devenir d'individus en fonction de circons-

tances particulières (la précarité, l'immigration) ou de dégager les déterminants d'une situation donnée (participation électorale, mobilité résidentielle...). L'échantillon permet également d'étudier la fécondité (voir dans ce même numéro du *Courrier des statistiques*, l'étude de la fécondité réalisée à partir de l'EDP par Isabelle Robert-Bobée).

L'un des atouts de l'EDP est également d'aider à apprécier la qualité des sources statistiques. Par exemple, l'identification des individus EDP recensés peut conduire à une évaluation des doubles comptes aux recensements de la population.

De même, le suivi longitudinal des individus dans l'échantillon donne une estimation des omissions aux recensements, principalement parmi les populations difficiles à appréhender : les personnes très âgées et les jeunes adultes. Il permet également de juger de la cohérence des réponses et donc de leur qualité.



Source : Sircom, Minefi

L'EDP est régulièrement enrichi de données issues de l'état civil

Un suivi historique n'est facilité que si le descriptif des variables reste stable. Or, les trente dernières années ont vu, par exemple, une refonte de la nomenclature des secteurs d'activité en 1975 et des professions en 1982. Toutefois, la catégorie sociale est également disponible sur l'ensemble des cinq recensements concernés dans sa définition la plus récente, celle de 1982, car les bulletins de 1968 et de 1975 n'ont été saisis et codés que dans les années quatre-vingt, lorsque cette nouvelle nomenclature était disponible.

Enfin, pour surmonter la gêne causée par la fréquente révision de leurs contours géographiques, les communes dans l'EDP sont codées, pour ce qui est des informations sur la résidence actuelle et antérieure issues des recensements, à la fois selon la définition du moment mais aussi selon un découpage actualisé, qui tient compte des créations et des fusions de communes.

6. Commission nationale de l'informatique et des libertés. Aujourd'hui, cette obligation n'existe plus pour les chercheurs ayant établi une convention-type leur permettant d'accéder à ces données à partir des locaux de l'Insee.

7. Institut national de la santé et de la recherche médicale.

Un avenir prometteur

La richesse des informations contenues dans l'EDP et le fait qu'il s'agisse d'un panel d'individus, ont favorisé le développement de nombreuses études à partir de cette source, même s'il semble que celle-ci soit encore sous-exploitée.

La liste des travaux, ayant donné lieu à des publications, réalisés à partir de l'EDP (annexe pages suivantes) montre l'importance de la place prise par les thèmes concernant la mobilité professionnelle ou géographique des individus ; en revanche, relativement peu d'études ont été axées sur des sujets démographiques à proprement parler.

Bien que l'échantillon ait contribué à des études démographiques concernant par exemple les différences de mortalité selon tel ou tel critère, l'EDP a probablement subi la concurrence d'autres sources – principalement « l'enquête famille » et « les échantillons de mortalité » –, mieux adaptées aux études démographiques très ciblées, ou plus faciles d'accès (pour accéder aux données de l'EDP, les chercheurs « hors Insee » devaient déposer des demandes à la CNIL⁶, alors que l'accès à d'autres fichiers est moins contraint).

Ces dernières sources n'offrent pourtant pas la richesse d'indicateurs sociodémographiques à différentes dates que l'on trouve dans l'EDP. L'un des objectifs du groupe d'exploitation de l'EDP est donc de favoriser l'accès des chercheurs à ses données, pour compléter avantageusement les études réalisées sur ces sujets, comme c'est déjà le cas pour les travaux sur la mortalité.

Par ailleurs, un projet de refonte de l'EDP est à l'étude. Il vise à élargir et à enrichir l'échantillon. Sa taille devrait être multipliée par quatre, passant ainsi de 1 % à 4 % de la population. L'échantillon devrait reposer sur la sélection des individus nés chaque année durant 16 jours contre 4 jours actuellement. Ceci compensera partiellement le problème lié à la réduction des effectifs d'individus recensés par les enquêtes annuelles de recensement (en moyenne un individu sur sept chaque année).

De plus, le champ de l'EDP sera étendu aux départements d'Outre-mer, jusque-là exclus de l'échantillon. Il pourrait également être élargi à de nouvelles sources. Dans un premier temps, il s'agira du fichier électoral et du fichier des causes de décès géré par l'Inserm⁷. Dans un horizon plus lointain, les possibilités d'appariement de l'EDP aux déclarations annuelles de données sociales (DADS) et aux fichiers de paie des agents de la fonction publique seront étudiées (faisabilité technique et légale).

Enfin, la rénovation de l'EDP s'accompagnera d'une réorganisation de la structure des fichiers qui devrait être mieux adaptée aux travaux d'études. L'échantillon comprendra des modules articulés, chacun regroupant les données par thème.

La refonte de l'EDP devrait voir le jour en 2008, conjointement à la mise en place de la nouvelle application de la base des répertoires des personnes physiques (BRPP). ■

Annexe : liste chronologique des études publiées ayant utilisé l'EDP

1988

- MORIN J. – Participation électorale, in *Économie et statistique* n° 152 et 165.
- SAUTORY O. – Près de la moitié de la population a changé au moins une fois de commune en 20 ans, in *Économie et statistique* n° 209, avril 1988.

1989

- ISNARD M. – Mortalité différentielle, communication à la 5^e réunion du réseau CICRED, octobre 1989 (Paris).

1990

- ISNARD M. – La mobilité en France entre 1962 et 1982, communication au congrès de l'ISI, septembre 1990 (Le Caire).

1991

- DESPLANQUES G. et ISNARD M. – Mobilité géographique d'après l'EDP, communication au congrès européen de démographie, octobre 1991 (Paris).

1992

- CHENU A. – Itinéraires professionnels et mobilité sectorielle in *La Documentation Française*, Cheminements professionnels et mobilités sociales, octobre 1992.

1993

- CRIBIER F. et KYCH A. – A comparaison of retirement migration from Paris and London, in *Environnement and Planning* Février 1993, vol. 25.
- MOLINIÉ A.F. – Des secteurs et des âges, in *Population* Novembre-décembre 1993, vol. 6.
- RICHARD J.L. – Étude sur l'entrée dans la vie adulte des jeunes d'origine immigrée. Un exemple d'utilisation de l'échantillon démographique permanent de l'Insee – mémoire de DEA (dir. Desplanques G.).

1995

- HÉRAN F. et ROUAULT D. – La double élection de 1995 : exclusion sociale et stratégie, Insee Première n° 414, novembre 1995.

1996

- COLLIN C. – Pour vivre vieux, vivons mariés, Regards sur l'Île-de-France n° 32, juin 1996.
- GALLAND O. et ROUAULT D. – Des études supérieures inégalement rentables selon les milieux sociaux, Insee première n° 469, juillet 1996.
- LAROCHE N. – Femme échange moins volontiers travail contre bébé Regards sur l'Île-de-France, n° 32, juin 1996.
- PÉRÈZ A. – L'échantillon démographique permanent, un trésor pour les démographes, Insee, Regards sur l'Île-de-France, juin 1996.
- ROUAULT D. – Les immigrés dans l'EDP : représentativité, omissions de collecte systématiques ou accidentelles, note interne, février 1996.
- ROUSSEL D. – Partir, revenir, in *Économie lorraine* n° 156, septembre 1996.

1997

- DEVOS C. – Des parcours professionnels mouvementés de 1968 à 1980, in *Profils Nord – Pas-de-Calais* n° 2, février 1997.
- ESPINASSE L. et RICCI C. – La participation électorale en Bourgogne, *Dimensions Bourgogne* n° 49, décembre 1997.
- HÉRAN F. (1997) – Les intermittences du vote : un bilan de la participation de 1995 à 1997, Insee Première n° 546, septembre 1997.
- LAROCHE N. – Mariage et mobilité : bague au doigt, changement de toit, Regards sur l'Île-de-France, n° 36, juin 1997.
- LAROCHE N. – Changer de profession et changer de région, Regards sur l'Île-de-France, n° 38, décembre 1997.
- RICHARD J.L. – Unemployment among young people of foreign origin in France : Ways of measuring discrimination, p. 1001-135, in *Swiss Federal Statistical office (ed.), on the Way to a Multicultural Society ?* Bern, BFS/OFS, 1997.
- ROUAULT D. et THAVE S. – L'EDP comme base d'études longitudinales sur les immigrés, « L'estimation du nombre d'immigrés » in *Insee Méthodes* n° 66, avril 1997.

1998

- BELLESOT J. et THOMAS J.-N. – Rester, partir ou revenir en Limousin in *Insee Limousin* n° 3, juillet 1998.
- Dossier échantillon démographique permanent, Six études des mobilités sociales et professionnelles in *Économie et statistique* n° 316-317, juin-juillet 1998. 1) La mobilité des personnes seules en France et en Grande Bretagne (P.E. OGDEN et R. HALL). 2) Devenir cadre dès trente ans (O. GALLAND et D. ROUAULT). 3) Le renouvellement de la main-d'œuvre industrielle (A.F. MOLINIÉ). 4) De recensement en recensement, le devenir professionnel des ouvriers et employés (A. CHENU). 5) Rester en France, devenir français, voter : trois étapes de l'intégration des enfants d'immigrés (J.L. RICHARD). 6) Reconstruire des trajectoires de mobilité résidentielles (D. COURGEAU, E. LELIÈVRE et O. WOLBER).

1999

- MESRINE A. – Les différences de mortalité par milieu social restent fortes, in Insee Données sociales.
- THAVE S. – Les vagues d'entrée et de départ des immigrés, Regards sur l'immigration depuis 1945 in *Synthèses* n° 30, octobre 1999.

2000

- BRUTEL C., JEGOU M. et RIEU C. – La mobilité géographique et la promotion professionnelle des salariés : une analyse par aire urbaine in *Économie et statistique* n° 336, juin 2000.
- MESRINE A. – La surmortalité des chômeurs : un effet catalyseur du chômage ?, in Insee Économie et Statistiques n° 334.

2002

- EKERT-JAFFRE O., JOSHI H., LYNCH K., MOUGIN R. et RENDALL M. (2002) – Fécondité, calendrier des naissances et milieu social en France et en Grande-Bretagne : politiques sociales et polarisation socioprofessionnelle, Ined, Population n° 3, mai-juin 2002.
- RICHARD J.L. – Scolarité, insertion sur le marché du travail et mobilité sociale, l'apport d'une analyse des parcours individuels à partir de données longitudinales 1975-1990 Présentation au groupe « Les obstacles à l'intégration et les moyens d'y remédier » du conseil d'analyse économique, avril 2002.
- RICHARD J.L., MOYSAN-LOUAZEL A. (2002) – De l'immigration étrangère parentale à la mobilité sociale des jeunes adultes : lignée familiale et dynamiques professionnelles individuelles au début des années 1990, in documents Cereq n° 164, 2002, p. 307-320.

2003

- CLANCHÉ F. – La participation électorale au printemps 2002, de plus en plus de votants intermittents, Insee Première n° 877, janvier 2003.
- COUET C. – La couverture de l'échantillon démographique permanent (EDP) par rapport au recensement de la population de 1999, Insee note n° 094/F170 du 21 novembre 2003.
- COUET C. – Le degré d'exhaustivité du recensement de la population de 1999 apprécié à travers l'échantillon démographique permanent, Insee note n° 095/F170 du 21 novembre 2003.
- Insee – Échantillon démographique permanent, manuel de l'utilisateur en statistiques et études, septembre 2003.
- RENDALL M. – « How important are intergenerational cycles of teenage motherhood in England and Wales ? A comparison with France », ONS, Population Trends, printemps 2003.
- RICHARD J.L. – Scolarité, insertion sur le marché du travail et mobilité sociale : l'apport d'une analyse des parcours individuels à partir de données longitudinales, in les rapports du CAE n° 45, 2003, p. 225-254.

2004

- CAMBOIS E – Carees and mortality in France : Evidence on how far occupational mobility predicts differential risks, in Social Science and Medicine n° 58.
- CAMBOIS E. – Occupational and Educational Differentials in mortality in French Elderly People : Magnitude and Trends over recent Decades, in Demographic Research, Avril 2004.
- RICHARD J.L. – Partir ou rester ? Destinées des jeunes issus de l'immigration, in Le lien social, PUF.
- ROBERT-BOBÉE I. – Fécondité et EDP : quelques éléments sur la qualité des données de l'EDP, Insee note n° 100/F170 du 27 novembre 2003.

2005

- LECLERC A. – Les inégalités en matière de santé, in Les Cahiers français n° 324, février 2005 – La documentation française.
- MENVIELLE G., LUCE D., GEOFFROY-PEREZ B., CHASTANG J.F., LECLERC A. for the EDISC group. – Social inequalities and cancer mortality in France. 1975-1990. Cancer Causes&Control, à paraître.
- MONTEIL C. et ROBERT-BOBÉE I. – « Les différences sociales de mortalité : en augmentation chez les hommes, stables chez les femmes », Insee première n° 1025.
- ROBERT-BOBÉE I. et MONTEIL C. – « Quelles évolutions des différentiels sociaux de mortalité pour les hommes et les femmes ? tables de mortalité par catégorie sociale en 1975, 1982 et 1990 et indicateurs standardisés de mortalité en 1975, 1982, 1990 et 1999 », Insee, document de travail n°F0506.
- RENDALL M., ROBERT-BOBÉE I., LAPPEGARD T., ROSEN M., SMALLWOOD S., COUET C. « First births by age and education in 1950s and 1960s birth cohort in Britain, France and Norway », ONS, Population Trends n° 121, automne 2005.

2006

- COUET C. – « Mobilité résidentielle des jeunes », Insee, Données Sociales 2006.
- FOUGÈRE D., SAFI M. – « L'acquisition de la nationalité française : quels effets sur l'accès à l'emploi des immigrés ? », Insee, France Portrait Social 2005-2006.
- MIKOL F., TAVAN C. – « La mobilité professionnelle des employés et ouvriers immigrés », Insee, Données Sociales 2006.
- ROBERT-BOBÉE I., RENDALL M., LAPPEGARD T., ROSEN M., SMALLWOOD S., COUET C. – « Âge au premier enfant et niveau d'études : une analyse comparée entre la France, la Grande-Bretagne et la Norvège », Insee, Données Sociales 2006.