

MYLÈNE CHALEIX

STÉFAN LOLLIVIER

Des panels pour les statistiques sociales

Journal de la société française de statistique, tome 146, n° 3 (2005),
p. 75-81

http://www.numdam.org/item?id=JSFS_2005__146_3_75_0

© Société française de statistique, 2005, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

DES PANELS POUR LES STATISTIQUES SOCIALES

Mylène CHALEIX *, Stéfan LOLLIVIER **

RÉSUMÉ

Depuis plusieurs années, la nécessité de développer des panels dans le domaine des statistiques sociales est débattue au CNIS (Conseil national de l'information statistique). Si l'intérêt de tels outils est reconnu de tous, leur mise en œuvre demande un engagement important initial et dans la durée. C'est dans cet esprit qu'a été menée la mission sur les « outils de suivi des trajectoires des personnes en matière sociale et d'emploi ». Les principales propositions du rapport sont reprises ici.

Mots clés : données administratives, enquêtes, panels, statistiques démographiques et sociales, statistique publique.

ABSTRACT

In recent years, France's National Council for Statistical Information (CNIS) has been debating the need to set up panels for social statistics. While everyone recognizes the value of such tools, their implementation requires a major commitment at the outset and over the long term. These criteria informed the project devoted to "tools for tracking individual trajectories in the social sphere and employment." Our article presents the report's main proposals.

Keywords : administrative data, demographic and social statistics, longitudinal surveys, official statistics.

Les débats menés au CNIS¹ depuis plusieurs années ont mis l'accent sur la nécessité de développer des outils de suivi des trajectoires des personnes. La demande sociale porte en effet de plus en plus sur la compréhension des trajectoires individuelles; les panels deviennent alors indispensables pour comprendre la dynamique des comportements (carrières salariales, insertion des jeunes, immigration ...), les effets du vieillissement, pour analyser les transitions (entrées/sorties du marché du travail, de la pauvreté ...), et pour modéliser l'hétérogénéité individuelle. Ils peuvent aussi s'avérer adaptés à l'évaluation des politiques publiques. Le suivi longitudinal apporte en effet

* Chef du bureau des établissements de santé, DREES, 14 avenue Duquesne, 75350 PARIS 07 SP; courriel : mylene.chaleix@sante.gouv.fr

** Directeur des statistiques démographiques et sociales, INSEE, 18 Bd. A. Pinard, 75675 Paris cedex 14; courriel : stefan.lollivier@insee.fr

1. Conseil National de l'Information Statistique.

une certaine information sur les caractéristiques inobservées des individus², information qu'il convient de prendre en compte lors des exercices d'évaluation. Mais les statisticiens d'enquêtes sont souvent réticents à mettre en place des panels, du fait de leurs coûts et de leur méthodologie de collecte plus complexe (érosion et suivi plus difficile des personnes), et de leur exploitation plus difficile. Certains enfin y sont opposés pour des raisons plus idéologiques : selon eux, le suivi temporel ferait courir des risques à la confidentialité des données, et les analyses causales seraient de nature à mettre en évidence la responsabilité des individus sur leur état actuel.

Suite à une demande du CNIS, le Directeur général de l'Insee a confié une mission sur les outils de suivi des trajectoires des personnes en matière sociale et d'emploi à Mylène Chaleix (alors membre de l'Inspection générale) et Stéfan Lollivier (alors directeur de l'Ensaë). Le rapport³, remis en juin 2004 et discuté au CNIS en septembre 2004, débute par un état des lieux de l'existant et des pratiques au niveau international, puis fait le point sur les données disponibles en France. Il décrit ensuite les besoins exprimés lors des entretiens menés par les rapporteurs. Il présente enfin des pistes possibles et propose un certain nombre de recommandations. Parmi celles-ci figurent des recommandations de court terme, qui peuvent être mises en place rapidement pour un coût modéré. D'autres, plus structurantes, et donc plus longues à mettre en place, vont permettre de renforcer la place des sources administratives dans le dispositif statistique relatif aux individus, alors que cette place est déjà centrale dans le dispositif statistique sur les entreprises. Leur mise en œuvre nécessitera, notamment avec la CNIL⁴, un examen approfondi des procédures afin de garantir les règles de confidentialité et de respect des personnes.

Les propositions faites par les rapporteurs s'articulent autour des trois axes suivants :

- Avant tout, une meilleure utilisation des données administratives ;
- Ensuite, la construction de trois panels de collecte respectivement au sein de la population générale, sur les enfants et sur les personnes âgées ;
- Enfin, une meilleure complémentarité entre enquêtes et données administratives.

Ces orientations sont complémentaires : seules les données obtenues par entretien sont suffisamment riches pour répondre aux interrogations de la statistique sociale, et notamment fournir les indicateurs communautaires issus du processus de Lisbonne⁵. Ainsi en est-il des indicateurs de privation. Au contraire, la qualité des variables de revenus n'est véritablement satisfaisante

2. Par exemple, faire apparaître des effets fixes individuels dans les estimations permet de prendre en compte toutes les variables individuelles invariantes au cours du temps qui ne seraient pas disponibles dans les fichiers.

3. Le rapport est disponible sur le site du CNIS, <http://www.cnis.fr/ind-doc.htm>, rubrique rapports présentés au CNIS.

4. CNIL : Commission nationale de l'informatique et des libertés.

5. La stratégie de Lisbonne a fixé aux États membres de l'Union toute une série d'objectifs à atteindre d'ici 2010. Certains d'entre eux sont précis et concrets, par exemple le fait de

que si l'on mobilise la source fiscale, ce qui justifie les appariements d'enquêtes avec les données administratives. Par ailleurs, au moins en Europe, il est admis que la durée d'interrogation par entretien des individus ne saurait excéder huit ou neuf ans. Au-delà, les risques liés à l'attrition⁶ sont jugés trop élevés. Aussi, le suivi des trajectoires de longue durée n'est-il véritablement possible qu'au travers des sources administratives, avec un corpus de données de fait limité.

Un panel alimenté par des données administratives

Le système statistique public portant sur les individus dispose, en France, d'un ensemble très riche de données administratives. Pourtant, chacun de ces éléments pris séparément souffre de défauts (faible nombre de variables, champ souvent restreint, coût de la mise en forme de l'information administrative de base parfois élevé, accessibilité faible). Le rapport recommande de construire, à partir de l'Échantillon démographique permanent (EDP, voir annexe 2), un panel d'individus à la fois élargi en taille et enrichi de sources administratives fiscales et sociales. Le dispositif nécessaire à la construction d'un tel outil, baptisé Centre d'Appariement Sécurisé (CAPS, voir annexe 1), est coûteux, mais porteur d'enjeux importants. Il devra donc être mutualisé au sein du système statistique public et garantir à la fois la sécurisation des traitements et une mise à disposition des données respectueuse des règles de confidentialité.

Des panels obtenus par collecte auprès des ménages

Le rapport recommande de compléter le dispositif administratif avec des panels obtenus par collecte, en s'appuyant sur les initiatives en cours. Il s'agit, en premier lieu, de pérenniser le panel en population générale SILC⁷ porté par Eurostat, comme successeur du panel européen. Deux sous-populations nécessitent la mise en place d'outils plus spécifiques portant respectivement sur les enfants et sur les personnes âgées. L'initiative européenne SHARE⁸ devrait doter la France d'un panel de personnes âgées et permettre d'éclairer la réflexion sur les conséquences économiques et sociales du vieillissement. Enfin, le projet de cohorte nationale d'enfants suivi par l'INED⁹ compléterait utilement le dispositif en fournissant des statistiques sur le devenir des enfants et tout particulièrement sur les enfants pauvres.

Tout en favorisant la complémentarité entre données d'enquêtes et données administratives

Cette complémentarité ne devrait pas soulever de problème technique tant est présente dans le système statistique public la culture d'enrichissement mutuel entre données administratives et données d'enquêtes. En revanche, il s'agit

parvenir à un taux d'emploi de 70 %, dont 60 % pour les femmes et 50 % pour les travailleurs les plus âgés (55 à 64 ans).

6. L'attrition est le fait que l'échantillon diminue avec le temps, si bien que la non-réponse augmente et le biais des réponses aussi [NDLR].

7. SILC : Statistics on income and living conditions.

8. SHARE : Survey of health, ageing and retirement in Europe.

9. INED : Institut national d'études démographiques.

d'un domaine où la CNIL est très vigilante, notamment depuis la loi du 6 août 2004, et les finalités devront être très soigneusement établies¹⁰.

Plusieurs exemples sont emblématiques de telles opérations : l'enquête sur les revenus fiscaux résulte, depuis 1956 (soit deux ans après le recensement de 1954), de l'appariement entre les données fiscales et le recensement puis l'enquête emploi et permet des études sur l'inégalité et la pauvreté; plus récemment, l'enquête « trajectoire des demandeurs d'emploi » suit une cohorte de nouveaux inscrits à l'ANPE, tirés dans les fichiers administratifs et interrogés tous les ans par entretien. Enfin, l'ESPS (Enquête sur la santé et la protection sociale) associe une enquête principalement téléphonique de l'IRDES¹¹ et les données administratives de l'Échantillon permanent d'assurés sociaux de l'assurance maladie, l'EPAS, qui lui sert par ailleurs de base de sondage.

Accessibilité et coordination

Le rapport insiste également sur la nécessité de rendre accessibles les données en les documentant et en les mettant largement à disposition. Il insiste sur un besoin d'évaluation de la qualité des dispositifs statistiques existants, et sur une nécessaire mise à disposition de travaux méthodologiques sur la collecte des panels, mais aussi sur les techniques de traitements. Toutes ces mesures devraient aider à promouvoir auprès des statisticiens publics et des chercheurs nationaux l'utilisation des panels d'individus, et le développement d'une méthodologie des enquêtes longitudinales.

La mise en œuvre des propositions du rapport exigera un investissement conséquent. Aussi les producteurs de sources statistiques démographiques et sociales devront-ils se coordonner et élaborer une stratégie commune au sein du système statistique français.

La création d'un tel dispositif suppose acquis conjointement deux préalables : une étude volumétrique de la durée et de la taille d'échantillon définissant les coûts optimaux, et l'explicitation des garanties de respect de la confidentialité et de la collecte loyale¹². Ces deux aspects doivent être traités ensemble pour permettre d'élaborer des solutions adéquates et qui ne soient pas démesurés, financièrement et techniquement, par rapport aux objectifs.

Depuis la présentation du rapport...

Dans son assemblée plénière de février 2005, le CNIS a largement approuvé les orientations générales du rapport, et émis l'avis suivant :

« Le Conseil a pris connaissance avec intérêt du rapport sur les trajectoires des personnes en matière sociale et d'emploi : il encourage le système statistique public à mettre en place des opérations statistiques permettant des analyses longitudinales et à développer les collaborations avec les chercheurs. Dans un souci d'efficacité collective, il recommande, d'une part, d'articuler au

10. Voir l'article de Michel Isnard dans ce même dossier.

11. Institut de recherche et de documentation en économie de la santé.

12. La collecte loyale renvoie à l'appariement de données d'enquête et de données administratives, éventuellement répété dans le temps, et donc à la transparence auprès des enquêtés.

mieux les sources administratives et les enquêtes, d'autre part, d'améliorer la coordination entre les services producteurs. Lors de la mise en place de ces opérations longitudinales, une attention particulière devra être apportée au respect des règles de confidentialité et de garantie de l'anonymat des individus. Sous ces conditions, le Conseil encourage la mise en place par le système statistique public d'un panel destiné au suivi des enfants et la constitution d'un fichier longitudinal articulé autour de l'échantillon démographique permanent, enrichi de différentes sources administratives. Le Conseil demande à être informé des avancées de ce dernier projet, en particulier en termes de garantie de la confidentialité et d'étude des coûts pouvant induire des arbitrages aux dépens d'autres opérations».

Dans cet avis, le Conseil a donné un mandat clair au système statistique public, mais il s'est également montré vigilant sur les questions de confidentialité et de coûts. D'un point de vue opérationnel, la deuxième collecte de SILC a eu lieu, la première collecte de SHARE a d'ores et déjà donné des résultats jugés très prometteurs au niveau international ; le panel d'enfants, plus complexe à mettre en place, a fortement mobilisé, autour de l'INED, le système statistique public, l'INSERM¹³ et l'InVS¹⁴. On peut raisonnablement espérer que cette opération démarre avant 2010. Enfin, le projet de valorisation des sources administratives longitudinales autour de l'échantillon Démographique Permanent se développe à l'Insee en collaboration avec le système statistique public. Sous réserve d'une approbation définitive par le CNIS et d'un avis favorable de la CNIL, ce projet pourrait également déboucher avant la fin de la décennie.

Annexe 1 : CAPS (Centre d'Appariement Sécurisé)

Il s'agit de mettre en place un centre de ressources permettant l'appariement de données administratives et de données d'enquêtes en généralisant la méthode du « double aveugle » utilisée pour l'enquête « revenus fiscaux ». L'idée est d'empêcher la possibilité de faire un lien direct entre les identifiants des deux sources. La sécurisation de ce centre doit porter à la fois sur les processus et les données pour garantir l'anonymat et le respect des règles de confidentialité. Un tel outil répond à un besoin de centralisation du système, afin de compenser par la multiplicité des sources l'insuffisance de chacune d'elles (champ partiel, population spécifique,...). Une telle proposition est réaliste car elle reprend en l'élargissant la démarche entreprise par la DREES dans la confection de l'échantillon national inter régimes de minima sociaux. Elle permet de mutualiser les coûts de développement d'instruments issus de différents services du système statistique public. Le statut juridique de ce centre devra être élaboré en accord avec la CNIL. Ce centre permettrait d'abord l'enrichissement de l'échantillon démographique permanent par des sources

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

14. Institut de veille sanitaire.

administratives fiscales ou sociales. Puis deux axes seraient développés : l'extraction de données pour enrichir des enquêtes réalisées par ailleurs, soit grâce à des variables supplémentaires, soit en apportant un complément sous forme de « population témoin » ; dans un processus inverse, la fourniture de bases de sondage.

Annexe 2 : Les modèles

Les panels de collecte

Panel en population générale : la plupart des grands pays disposent d'un panel généraliste construit sur le modèle du Panel Study of Income Dynamics (PSID) américain. Ce dernier, lancé en 1968, assure le suivi longitudinal d'un échantillon représentatif d'individus vivant aux États-Unis, ainsi que des familles auxquelles ils appartiennent. Il s'intéresse en priorité à la dynamique des comportements économiques et démographiques, même s'il aborde aussi des questions sociologiques ou psychologiques. À l'issue de la collecte de 2003, le PSID avait recueilli une information auprès de 65 000 individus pendant 36 années de cycle de vie. L'enquête est réalisée par le Survey Research Center, Institute for Social Research, de l'Université du Michigan. De nombreuses enquêtes spécifiques ont été collectées sur la base de l'échantillon du PSID.

En France, après les expériences du panel lorrain et du panel européen (1994-2001), le projet SILC au sein de l'Union européenne a pris le relais : la deuxième collecte a lieu en 2005.

Panel d'enfants : les pays anglo-saxons ont depuis de nombreuses années organisé de vastes cohortes constituées d'enfants suivis depuis leur naissance (ou depuis la grossesse de leur mère) jusqu'à l'âge adulte, sur la base d'échantillons représentatifs au plan national. Les buts de ces suivis sont divers : ils incluent presque toujours l'étude du « développement de l'enfant », appréhendé dans un contexte familial, social, économique, à travers son histoire familiale et son parcours éducatif. Un questionnaire sur la santé est presque toujours présent. La plus ancienne de ces cohortes a été constituée en Grande-Bretagne en 1946 (suivie des cohortes 1958 et 1970). Tout récemment, les Anglais ont lancé en 2000 la « Millenium Cohort Study » (MCS) auprès de plus de 21 000 enfants nés en 2000-2001, dont une partie des enfants nés de parents de la cohorte 1970.

En France, le projet « Enfants » est porté par l'INED, il est soutenu par le CNIS, les administrations en charge de la santé, et la Statistique publique ; la première collecte pourrait avoir lieu d'ici 3 à 5 ans.

Panel de personnes âgées : aux États-Unis, l'Université du Michigan conduit depuis 1994 le Health and Retirement Study (HRS) en interrogeant tous les deux ans plus de 22 000 américains de plus de 50 ans. Financée par le National Institute on Ageing, l'enquête s'intéresse aux effets du vieillissement sur la santé physique et mentale, ainsi qu'à des aspects économiques tels la couverture sociale, les aspects financiers, l'état sur le marché du travail

et la préparation de la retraite. En Grande Bretagne, a débuté l'English Longitudinal Survey of Ageing (ELSA), inspirée du HRS.

Il sert de modèle au projet européen SHARE au sein de onze pays, auquel participe la France (la maîtrise d'œuvre est assurée par l'Insee). Après le succès de l'opération pilote de 2004, la décision de la panélisation de l'échantillon est à l'étude.

Panel administratif : le Canada a mis en place la banque de Données administratives longitudinales (DAL), outil de recherche sur le revenu et la démographie. Ce panel d'individus contient de nombreuses variables annuelles, portant aussi bien sur des questions démographiques que sur le revenu des individus et de leur ménage. En France, l'Échantillon démographique permanent (EDP) a été mis en place après le recensement de la population de 1968. Il répond à la volonté de tirer parti d'informations sur les mêmes individus collectées au fil du temps dans les sources démographiques, recensements et état civil. L'échantillon est constitué par les personnes nées quatre jours donnés, quelle que soit l'année. Il se renouvelle à partir des naissances, ou par la venue d'immigrés révélée par les recensements ou encore par un événement enregistré dans un acte de l'état civil. À tout moment, la base est proche d'un échantillon au 1/100 (4/365) de la population résidant en métropole.

Conformément aux préconisations du rapport, le projet d'élargissement de l'EDP se met en place avec le soutien du CNIS, tant d'un point de vue de la taille de l'échantillon que des premières sources administratives en sus des données de l'état civil et des recensements de la population. Une attention toute particulière sera portée au coût de l'opération et à la confidentialité des données.