

Les enquêtes transversales permettent de décrire l'état de santé d'une population à un moment précis, en recueillant des données sanitaires, des données socio-démographiques et des données sur l'exposition à des facteurs de risque. Ces enquêtes sont réalisées à partir d'un échantillon de population qui doit être le plus représentatif possible de la population étudiée, afin de pouvoir généraliser les résultats obtenus. Pour construire une enquête transversale, il faut choisir une base de sondage, une méthode d'échantillonnage et un mode de recueil des données, et ces choix vont influencer sur la validité des résultats. Les enquêtes transversales auprès d'un échantillon représentatif de la population constituent des outils indispensables capables de mettre en lumière des points faibles du système de santé, et de fournir une aide à la décision pour la mise en place des mesures destinées à améliorer la prévention ou la prise en charge de la santé.

## Comment observer la santé des individus à l'échelle d'une population ?

Nathalie Beltzer<sup>1</sup>, Patrick Zouary<sup>2</sup> – <sup>1</sup>ORS Île-de-France, Paris, France. <sup>2</sup>SI2S, Paris, France

Étudier l'état de santé à l'échelle de toute une population ou dans une fraction de celle-ci peut permettre de mettre en évidence des hétérogénéités dans la distribution des événements de santé dans certains sous-groupes de population, ou dans des espaces géographiques particuliers, source de connaissances utiles pour comprendre leurs origines et éventuellement les actions visant à les corriger.

Pour réaliser ce type de recherches et d'études, il faut avoir accès à des données qui décrivent l'état sanitaire des personnes qui composent la population, et qui renseignent sur l'exposition aux facteurs de risque suspectés. Ce sont des données biologiques, médicales et sanitaires, mais également des informations sur d'éventuels comportements à risque, des données socio-démographiques (âge, sexe,...), voire économiques (catégorie

socio-professionnelle, revenus,...). Comment obtient-on ces données ? On peut parfois les récupérer à partir de sources administratives exhaustives telles que les données de la statistique publique ou des bases de données de type médico-administratif (par exemple les causes médicales de décès). Le plus souvent cela nécessite de réaliser des enquêtes *ad-hoc*, lorsque les sources existantes ne sont pas suffisantes pour rendre compte du phénomène que l'on cherche à étudier.

Les enquêtes ponctuelles réalisées sont appelées « transversales » car elles recueillent des informations instantanées, contrairement aux enquêtes « longitudinales » qui suivent des événements de santé dans la durée (par exemple les études de cohortes) [1]. Les enquêtes transversales permettent

de décrire l'état de santé d'une population à un moment donné. Elles sont parfois répétées dans le temps, et apportent alors une vision dynamique, plus évolutive.

Pour ce type d'enquêtes, il est essentiel que la population étudiée soit représentative de la population générale.

### LES MÉTHODES UTILISÉES

Pour les enquêtes transversales descriptives, plusieurs méthodes de collecte de données en population générale sont envisageables. Elles visent toutes à obtenir des estimations valides de la population étudiée, afin de pouvoir inférer à la population générale les résultats obtenus à partir d'un échantillon de population. Il s'agit d'éviter ce qu'on appelle les

1. Une étude de cohorte concerne un groupe de sujets suivis longitudinalement, à l'échelle individuelle. Elle repose sur la collecte d'informations concernant des caractéristiques et des expositions des sujets suivis qui sont recueillies avant le moment d'apparition des effets analysés, susceptibles de modifier le risque de survenue de phénomènes de santé divers.

Décembre 2011

« biais<sup>2</sup> d'échantillonnage ». Cette recherche de représentativité nécessite de faire trois types de choix techniques.

## 1. Choisir une base de sondage

Une enquête en population générale suppose la sélection (en principe aléatoire, c'est-à-dire par tirage au sort) d'un nombre fini de personnes au sein de la population d'une zone géographique déterminée afin de constituer un échantillon représentatif de la population générale. La source utilisée pour sélectionner ces individus est la base de sondage. Cette base est souvent construite à partir de fichiers de données administratives que l'on souhaite le plus exhaustif possible, c'est-à-dire regroupant le plus largement la population à un moment donné et sur un lieu donné.

Les bases individuelles les plus complètes en France, en l'absence de registres de la population, sont le recensement général de la population et sa déclinaison appelée l'échantillon maître de l'INSEE, constitué à partir du recensement général de la population. Il existe de nombreuses autres sources possibles. Les bases de l'assurance-maladie (fichiers des assurés sociaux des différents régimes) sont également souvent mobilisées, comme dans le cas de l'enquête sur la santé et la protection sociale (*Encadré 1*). Le cadastre ou les fichiers EDF peuvent également être exploités. L'accès à ces bases peut cependant représenter un coût financier non négligeable tout en ne fournissant pas toujours l'exhaustivité recherchée. Une alternative fréquemment utilisée pour contourner la difficulté d'accès à ces données est le recours à la base des abonnés téléphoniques. Cependant, elle nécessite de mettre en place des protocoles relativement complexes qui permettent de tenir compte de plusieurs phénomènes : l'augmentation du nombre de personnes qui abandonnent une ligne fixe pour un abonnement exclusif à un téléphone mobile, l'arrivée d'autres opérateurs de fixes que France Telecom et les personnes qui sont sur liste rouge.

## 2. Choisir une méthode d'échantillonnage

La méthode d'échantillonnage est un élément crucial pour la validité de l'inférence à partir de l'échantillon obtenu. C'est

2. Un biais dans une enquête est une démarche ou un procédé qui engendre des erreurs dans les résultats de l'enquête.

## 1 – L'ENQUÊTE SANTÉ ET PROTECTION SOCIALE (ESPS)

L'enquête ESPS, réalisée par l'Institut de recherche en économie de la santé, est l'enquête de référence sur la santé, l'accès aux soins et la couverture maladie en France. Elle recueille depuis 1988 des données sur l'état de santé, la couverture maladie, la situation sociale et le recours aux soins d'un échantillon de 8 000 ménages ordinaires, soit 22 000 personnes environ. La base de sondage est le fichier « assurés » et le fichier « prestations » du Régime général, du Régime des professions indépendantes et du Régime des professions agricoles. Le questionnaire principal est administré par téléphone (4 appels). Pour les personnes sans téléphone ou dont le numéro est inconnu, le questionnaire principal est administré au cours de deux visites d'enquêteur [2].

l'opération par laquelle on sélectionne ou on choisit les individus qui constituent l'échantillon. Dans les enquêtes auprès de la population générale, deux méthodes d'échantillonnage sont en général envisageables : la méthode des quotas et l'échantillon aléatoire.

La méthode des quotas (ou méthode non probabiliste) est souvent utilisée par les instituts de sondage car elle est rapide, peu onéreuse et bien adaptée aux petits échantillons. Elle est basée sur la répartition connue *a priori* d'une population (âge, sexe, situation géographique, catégorie socio-professionnelle...). Une fois fixée la dimension du sondage que l'on souhaite effectuer (selon la précision souhaitée), il suffit de calculer le nombre d'individus à inclure dans l'échantillon pour chaque critère choisi, puis de trouver ces personnes, par exemple en appelant les abonnés au téléphone. Un individu injoignable peut ainsi rapidement être remplacé par un autre dans l'échantillon, dès lors qu'ils partagent les mêmes caractéristiques. Le problème est qu'un individu difficilement joignable du fait d'un mode de vie très mobile peut justement avoir des comportements de santé différents de la personne plus régulièrement présente à son domicile, et ces comportements de santé sont particulièrement intéressants à connaître. Il peut donc être gênant de considérer que des critères socio-démographiques identiques rendent les individus interchangeable. Par ailleurs, cette méthode empêche tout contrôle de la « non-réponse », indicateur souvent utilisé dans les revues internationales pour apprécier la validité scientifique d'une enquête (moins on a de « non-réponses », plus l'enquête est valide).

La méthode aléatoire limite ces inconvénients. Les unités (individus ou ménages) sont tirées au sort de façon aléatoire dans la base de sondage. Le fichier est ainsi constitué d'un nombre fini d'unités à

atteindre. Aucun individu ne pouvant être remplacé par un autre, il est important que tout individu sélectionné et éligible soit interrogé. Cette méthode implique ainsi d'interroger des individus difficiles à joindre et qui ont souvent des profils intéressants en termes de comportements de santé. Elle nécessite la mise en place d'un protocole d'enquête plus long et donc plus coûteux, puisque plusieurs tentatives de contact (plus d'une dizaine) sont envisagées pour joindre l'individu tiré au sort. Contrairement à la méthode des quotas, les informations concernant les tentatives de contact infructueuses peuvent être quantifiées et utilisées comme critères de qualité. Dans la mesure où la non-participation à une enquête peut être une source importante de biais dans l'estimation des indicateurs, ces critères sont particulièrement importants à suivre et donnent lieu à la possibilité d'un « redressement » de l'échantillon. Cette étape consiste à « caler » l'échantillon sur les principales caractéristiques de la population observée afin de s'assurer de la représentativité de l'échantillon en termes de structure d'âge, de catégories socio-professionnelles et de région d'habitation pour les femmes et les hommes.

En France, depuis le début des années 1990, la plupart des enquêtes de santé en population générale sont réalisées à partir d'échantillons aléatoires, facilitant la comparaison avec les enquêtes étrangères, dont la grande majorité repose sur cette méthodologie. Cette méthode est la seule qui permette d'appliquer avec rigueur les principes de l'inférence statistique. En ce sens, elle est même meilleure qu'un sondage « exhaustif » pour lequel on ne contrôle pas les données manquantes.

## 3. Choisir un mode de recueil des données

Longtemps le recueil des données s'est fait essentiellement lors d'entretiens en

face-à-face, plus rarement par voie postale, ce mode de collecte offrant en France des taux de réponse beaucoup trop faibles. Le mode de collecte en face-à-face repose principalement sur le déplacement d'enquêteurs auprès des personnes à interroger, permettant donc théoriquement à toute personne de pouvoir participer à l'enquête. Les réponses, qui étaient d'abord recueillies sur un support papier, sont aujourd'hui directement saisies sur un ordinateur. Les entretiens intègrent parfois une partie auto-administrée (la personne enquêtée remplit elle-même le questionnaire), ce qui favorise la réponse à des questions plus sensibles ou personnelles.

Avec l'extension de la couverture téléphonique en France s'est posée la question au début des années 1990 d'un recueil des données par téléphone. Une comparaison des deux modes d'entretien a été réalisée lors de l'enquête sur l'analyse des comportements sexuels en France (Encadré 2). Si les réponses correspondant aux attentes sociales jugées « normales » semblent davantage exprimées par téléphone qu'en face-à-face, on observait aussi une plus grande cohérence dans les réponses et des taux de refus similaires, rendant difficile de conclure sur l'avantage de l'un ou l'autre de ces modes de passation du questionnaire. D'autres critères ont déterminé le choix du téléphone : l'homogénéité plus grande des entretiens (grâce à un système de collecte de données assistée par téléphone et informatique) et surtout un coût moindre, ce qui permet d'augmenter la taille de l'échantillon et donc la robustesse de l'enquête.

Depuis le début des années 1990, la majorité des grandes enquêtes en santé réalisées auprès de la population adulte en France repose sur un protocole d'échantillonnage aléatoire à partir de fichiers de numéros de téléphone (annuaire téléphonique) et des entretiens téléphoniques. Toutefois, devant l'évolution de la couverture téléphonique en France évoquée plus haut, cette méthode de sondage ne permet de joindre qu'une population de plus en plus restreinte. Parallèlement, l'absence d'annuaire commun à l'ensemble des opérateurs réduit l'exhaustivité du plan de sondage initial. Il a donc été nécessaire de faire évoluer la méthodologie des enquêtes en santé en population générale.

La sixième enquête téléphonique KABP sur les connaissances, attitudes, croyances et

comportements face au VIH/Sida de la population habitant en France métropolitaine (Encadré 3) a été l'occasion de tester les nouvelles modalités de sélection d'un échantillon capable de couvrir le mieux possible la population résidant en France. En plus de l'échantillon habituel des personnes interrogées à partir de leur téléphone fixe, un large échantillon a été sélectionné *via* le téléphone mobile, que ces individus possèdent ou non une ligne de téléphone fixe. Et pour la première fois, en l'absence d'annuaires téléphoniques regroupant l'ensemble des opérateurs présents sur le marché français, les numéros de téléphone fixe comme mobile ont été générés totalement aléatoirement. Les premiers résultats de cette enquête montrent qu'il est possible aujourd'hui de faire passer un questionnaire de près de 40 minutes sur téléphone mobile, sans que cela ne conduise à des taux d'interruption plus élevés que sur un fixe.

Il reste toutefois difficile aujourd'hui de présager de la méthodologie des futures enquêtes en population générale. La diffusion des technologies de l'information, comme internet, et donc de la téléphonie par internet, risque de modifier encore le paysage de la téléphonie en France, rendant nécessaire l'adaptation méthodologique des enquêtes. En 2010, 71 % de la population dispose d'un accès à internet à domicile [4], proportion qui devrait augmenter dans les prochaines années. Aussi les enquêtes réalisées par internet représentent-elles une alternative possible. Moins coûteuses que celles en face-à-face et même que celles réalisées par téléphone, elles consistent à proposer aux enquêtés sélectionnés le remplissage d'un questionnaire « *on line* » selon leur disponibilité (large plage horaire et reprise possible si interruption). Mais on se trouve là

encore confronté à différentes limites. Ni l'éligibilité, ni l'identité de la personne sélectionnée ne peuvent être vérifiées. L'absence d'enquêteurs limite la cohérence et l'exhaustivité des réponses. Enfin, en l'absence d'une base de sondage permettant de sélectionner les personnes disposant d'une connexion internet, il est difficile d'obtenir un échantillonnage précis et d'estimer le taux de participation. Une généralisation de ces enquêtes par internet n'est pour le moment pas encore envisageable dans la mesure où elles sont essentiellement réalisées à partir d'échantillons de personnes volontaires, et par là même non représentatives de la population générale (on parle de biais d'auto-sélection). Différentes expériences menées en France et à l'étranger devraient fournir des éléments complémentaires sur la faisabilité de telles enquêtes en population générale.

## LES LIMITES DE CES ENQUÊTES

Quels que soient la base de sondage, la méthode d'échantillonnage et le mode de passation du questionnaire, les données produites par les enquêtes transversales sont en général utilisées comme des indicateurs descriptifs d'un état de santé à un moment donné. Or leur production dépend bien souvent du contexte de l'enquête (le sujet ou thème, le contexte politique et social), de la formulation des questions, voire de leur place dans le questionnaire. Il semble donc important de suivre l'évolution de ces indicateurs à travers la répétition dans le temps d'une même enquête, et de limiter leur utilisation comme reflet d'une réalité à un moment donné.

### 2 – L'ENQUÊTE « ANALYSE DES COMPORTEMENTS SEXUELS EN FRANCE » (ACSF)

**L'enquête ACSF est la première enquête en population générale réalisée par téléphone en France dans le domaine de la santé. Elle a été réalisée en 1992 au plus fort de l'épidémie de sida auprès de 20 055 personnes (tirées au sort parmi les abonnés au téléphone) qui ont accepté de répondre à un questionnaire sur leurs comportements, en particulier sexuels et d'usages de drogues, potentiellement à risque d'infection par le VIH. Cette enquête a permis d'orienter les stratégies nationales de prévention du sida, en particulier parmi les adolescents, les homosexuels masculins, les personnes en situation de vulnérabilité sociale, en France métropolitaine et dans les départements d'Outre-Mer.**

**Elle a été suivie en 2006 par l'enquête sur le « Contexte de la Sexualité en France » (enquête CSF) réalisée sur un échantillon aléatoire de la population générale des deux sexes, de 18 à 69 ans [3].**



Une critique souvent faite aux enquêtes transversales en population générale est d'ailleurs celle de produire des données déclaratives qui ne sont donc pas des mesures réelles de l'état de santé ou des comportements. Pour autant, l'interprétation de données déclaratives est souvent enrichissante puisque celles-ci peuvent traduire la représentation que se font les individus de l'objet étudié (par exemple, la perception de leur état de santé). Elles peuvent par exemple mettre en évidence des écarts entre une prévalence<sup>3</sup> déclarée et une prévalence diagnostiquée à travers le système de soins, ce qui peut refléter des problèmes d'accès au système de soins et révéler une méconnaissance de certaines pathologies.

Une autre interrogation sur la validité des résultats obtenus par ce type d'enquêtes

3. Nombre de personnes atteintes d'une certaine maladie à un moment donné dans une population donnée

est la non-participation de plus en plus importante de la population aux enquêtes. En effet, les taux de refus sont importants (souvent de plusieurs dizaines de pourcent). Ceci traduit sans doute une certaine lassitude des individus à répondre, trop sollicités par des démarchages téléphoniques, peut-être aussi une défiance vis-à-vis des sondages. Aussi une rémunération des participants est-elle parfois envisagée aujourd'hui. Certains font l'hypothèse que le biais introduit par une incitation financière à la participation (sélection d'un certain type d'individus) est moins important que celui engendré par une faible participation.

Enfin, ce type d'enquêtes ne permet pas d'obtenir des informations sur des comportements très minoritaires ou sur des groupes d'individus particuliers. L'obtention de telles données nécessite de recourir à d'autres méthodes, comme les

enquêtes analytiques ou des entretiens qualitatifs (ces derniers permettant de recueillir des informations en interrogeant d'une manière moins directive un petit nombre d'individus).

## CONCLUSION

Les enquêtes auprès d'un échantillon représentatif de la population permettent d'observer des événements de santé impossibles à appréhender dans les sources de données existantes (grandes enquêtes de la statistique publique par exemple), de décrire des évolutions concernant des pathologies ou des facteurs de risque, si l'enquête est répétée dans le temps, et de fournir ainsi des données utiles pour l'élaboration de modèles prévisionnels. Ces observations peuvent mettre en lumière des points faibles du système de santé, et fournir une aide à la décision pour la mise en place des mesures destinées à améliorer la prévention ou la prise en charge de la santé.

## RÉFÉRENCES

- [1] Enquêtes longitudinales et rétrospectives (chapitre 5). In : Tremblay ME, Lavallée P, El Hajjirami M, eds. *Pratiques et méthodes de sondage*. Paris : Dumod, 2011.
- [2] Allonier C, Dougnon P, Rochereau T. *Enquête sur la santé et la protection sociale* 2008. Rapport IRDES, juin 2010.
- [3] ACSF. Analysis of sexual behaviour in France (ACSF). A comparison between two modes of investigation : telephone survey and face-to-face survey. *AIDS* 1992 ; 6 : 315-23.
- [4] Bigot R, Croutte P. *La diffusion des technologies de l'information et de la communication dans la société française*. Rapport du CREDOC, n°269, décembre 2010.
- [5] Beltzer N, Saboni L, Sauvage C, Sommen C, et l'équipe KABP. *Les connaissances, attitudes, croyances et comportements face au VIH/sida en Île-de-France en 2010*. Rapport de l'ORS Île-de-France, décembre 2011.

### 3 – L'ENQUÊTE KABP (KNOWLEDGE, ATTITUDES, BELIEFS AND PRACTICES)

Réalisées à intervalle régulier depuis près de 20 ans (1992, 1994, 1998, 2001, 2004 et 2010), les enquêtes KABP permettent de suivre l'évolution des connaissances, attitudes, croyances et comportements face au virus du sida des adultes vivant en France. Les enquêtes KABP de 1987 et 1990 reposaient sur un échantillon interrogé en face à face et constitué selon la méthode des quotas. Afin de tenir compte de l'évolution de la couverture téléphonique en France (téléphonie mobile et multiplicité des opérateurs), deux échantillons ont été générés aléatoirement, l'un à partir de numéros filaires, l'autre à partir de numéros portables. L'enquête 2010 a été réalisée sous la responsabilité de l'ORS Île-de-France avec la collaboration de l'INPES, l'InVS et de l'Inserm. Les premiers résultats seront publiés en décembre 2011 [5] et disponibles sur le site : [www.ors-idf.org](http://www.ors-idf.org)

## PRÉSENTATION DE L'INSTITUT DE RECHERCHE EN SANTÉ PUBLIQUE

L'Institut de Recherche en Santé Publique (IReSP) est un groupement d'intérêt scientifique créé en 2007 par une convention entre 23 partenaires, acteurs de la recherche en Santé Publique (voir ci-dessous). Son objectif général est de constituer une communauté scientifique de taille internationale capable de répondre au développement souhaité de la recherche en Santé Publique et de contribuer aux nouveaux dispositifs mis en place par la loi du 9 août 2004 relative à la politique de Santé Publique. Pour atteindre cet objectif, le GIS-IReSP s'appuie sur une mutualisation des compétences et des moyens de ses partenaires. Le GIS-IReSP est dirigé par Alfred Spira, médecin, épidémiologiste, professeur de Santé Publique et d'épidémiologie.

Les domaines de recherche soutenus sont les suivants :

- Fonctionnement du système de santé

- Politiques publiques et santé
- Interaction entre les déterminants de la santé

Les modalités d'actions du GIS sont :

- Lancement d'appels à projets ciblés
- Aide à l'émergence d'équipes de recherche
- Mutualisation d'outils pour la recherche en Santé Publique
- Constitution de groupes de travail sur des sujets émergents
- Aide à la mise en place et à l'exploitation de grandes enquêtes et de grandes bases de données
- Valorisation et communication

Afin de pallier le manque de visibilité des résultats de la recherche en Santé Publique en France, l'IReSP a décidé de créer ce bulletin trimestriel à large diffusion

intitulé *Questions de Santé Publique*. Chaque trimestre, un sujet de recherche en Santé Publique intéressant le grand public est traité par un chercheur.

### LES PARTENAIRES DE L'IReSP

Ministères (Ministère de la Santé [DGS et DREES], Ministère délégué à la Recherche), Opérateurs de la recherche en Santé Publique (CNRS, Inserm, IRD, INED, EHESP, UDESA, CPU, Institut Pasteur, CNAM, Sciences Po), Agences et opérateurs de la Santé Publique (InVS, HAS, AFSSAPS, ANSES, EFS, ABM, INPES, InCa), Organismes de protection sociale (CNAMTS, RSI, CNSA).

Site internet : [www.iresp.net](http://www.iresp.net)

Cet article ainsi que les précédents numéros de *Questions de Santé Publique* sont téléchargeables sur le site internet de l'IReSP : [www.iresp.net](http://www.iresp.net)