

Pour un véritable accès aux données de santé

Nicolas BELORGEY

Sociologue, CNRS

Dans son numéro 2 de 2014, la revue *Statistique et société* soulève la question de l'accès aux données de santé. Elle souligne notamment qu'en-dehors de cercles restreints à la définition peu claire, cet accès demeure trop limité, notamment dans une perspective scientifique et citoyenne (Bar Hen). Elle rappelle également que « l'Initiative transparence santé » (ITS) provient essentiellement d'acteurs du secteur de l'assurance et des produits de santé, qui ont un intérêt financier à cette ouverture (Briatte et Goeta). Or, cet intérêt n'est pas nécessairement compatible avec la préservation des droits des personnes. En effet, deux types de droits doivent être conciliés en matière d'accès aux données de santé, conciliation qui a fait l'objet de différentes lois, dont celle de 1978, dite « Informatique et libertés » : celui des personnes, qui inclut la protection de leur vie privée et, en matière de santé, la confidentialité de leur état de santé pour éviter tout profilage des assurés dans un but lucratif ; celui de l'accès à ces informations, dans un objectif d'accroissement des connaissances comme bien public.

Nous défendons ici l'idée qu'en dépit de propositions tendant à améliorer l'accès aux données de santé tout en préservant les droits des personnes, le numéro 2 de *Statistique et société* ne prend pas totalement en compte les besoins d'une démarche rigoureuse aujourd'hui en sciences sociales, et esquissons quelques propositions en la matière.

I. Les réflexions de ce numéro ne prennent pas totalement en compte les besoins d'une démarche rigoureuse en sciences sociales.

Ce numéro présente un certain nombre de dispositifs existants dans la statistique publique, qui permettent de concilier protection de l'anonymat des personnes et accès aux informations (Le Gléau). Ils sont utilisés notamment en matière de statistique des entreprises ou de fiscalité. Le plus perfectionné d'entre eux semble résider dans la combinaison de l'accord préalable du Comité du secret statistique (CSS) et du recours à une technique d'accès particulière, le Centre d'accès sécurisé aux données (CASD). Cette dernière permet aux chercheurs de produire des statistiques agrégées à partir des données détaillées des bases, sans pour autant pouvoir individualiser celles-ci dans le compte-rendu qu'ils en font. La proposition faite en matière d'accès aux données de santé sur la réplique de ce dispositif consiste pour les enquêteurs en « un engagement de (...) ne tenter en aucun cas d'identifier un individu précis de la base » (p. 32).

Pour progressiste qu'elle soit par rapport à la situation actuelle, cette proposition semble cependant informée par une conception fortement quantitaviste de la démarche de connaissance en sciences sociales. Or celle-ci doit être appuyée sur des outils non seulement quantitatifs, mais aussi qualitatifs. Bien sûr, l'agrégation des résultats au niveau collectif présente l'avantage de la généralité, de la plus grande extension possible des propositions avancées. Sont ainsi évités les biais de sélection que peuvent présenter les cas individuels. Mais une bonne intelligence des processus à l'œuvre suppose aussi de raisonner sur des cas individuels, comme le font les historiens et les ethnographes (M. Weber 1917; Passeron 1991; F. Weber 2009). On gagne en compréhension ce qu'on perd en extension, et *vice versa* si on passe du qualitatif pur au quantitatif pur. De fait, dans une démarche totalement rigoureuse, on allie les deux outils : raisonnement sur des cas afin d'obtenir une bonne compréhension des phénomènes au niveau le plus fin possible, tests d'hypothèses statistiques afin de vérifier le degré de généralité de ces éléments intermédiaires (Gramain et Weber 2001). En pratique, on effectue souvent des allers-retours entre ces deux niveaux, sans que l'un ait nécessairement la prééminence chronologique, méthodologique ni surtout ontologique (en matière d'interprétation du réel). Cette double démarche semble d'autant plus nécessaire que les données statistiques ont elles aussi leurs biais de construction, puisque fréquemment produites par des administrations dont elles ne sont qu'un sous produit de l'activité, comme la police (Bruno et Didier 2013) ou les hôpitaux (Belorgey 2010). Pour reprendre les termes d'Alain Desrosières (2005), on n'aurait à la limite le choix, si on ne se fiait qu'à l'une des deux méthodes prises isolément, qu'entre « décrire l'État ou explorer la société. »

En matière de données de santé, cette alliance entre approches d'origines quantitative et qualitative semble particulièrement nécessaire. En effet, les principales bases de données existantes ne sont pas conçues à des fins de recherche, donc d'objectivité scientifique, mais de gestion de services de soin (informatique hospitalière), de comptabilité (idem), ou d'administration de la Sécurité sociale (systèmes d'information de l'assurance-maladie, dont le Sniiram est la partie émergée de l'iceberg). Il y a donc fort à parier que l'écart entre l'image des personnes qui ressort de ces données et ce qu'elles sont et font en réalité, est aussi important que celui qui existe entre un relevé de dépenses de consommation de « loisir » et ce que les personnes retirent de ce type d'activité, ainsi que les déterminants de celles-ci. Les données statistiques de santé sont aujourd'hui produites à des fins thérapeutiques, organisationnelles ou d'administration du système de soin ; elles présentent donc un biais de construction important que seules des enquêtes qualitatives individualisées auprès des personnes concernées, et portant le point de vue, partiel aussi mais nécessaire à la compréhension d'ensemble, de celles-ci, peuvent apporter. Par exemple, la plupart des statistiques de santé, pour précises qu'elles puissent être sur les consommations de soin des personnes, demeurent muettes sur les appartenances sociales de celles-ci : leur métier précis, leur origine sociale, leur environnement familial élargi, etc. Ces données ne sont parfois qu'approchées par certaines variables intermédiaires, comme l'affiliation éventuelle à l'Aide médicale d'État. En leur absence, comment observer les effets de décisions publiques sur les différents groupes de la population ? Comment observer les évolutions des inégalités sociales de santé ? Etc. Pour répondre à de telles questions, il est nécessaire de recouper les données entre les bases et de recourir à des enquêtes directes auprès des personnes. Des démarches comme le projet Monaco¹ ou la cohorte Constances² vont dans le bon sens mais demeurent partielles.

En matière de protection des droits des personnes, la démarche par cas a depuis longtemps élaboré une réflexion et des règles en la matière. Il s'agit, là aussi, de concilier d'une part respect de la vie privée des personnes et d'autre part précision scientifique et caractère de bien public de la connaissance (Beaud et Weber 1997, chap. 9). Cette configuration s'est enrichie

1. <http://www.irdes.fr/recherche/partenariats/monaco-methodes-outils-et-normes-pour-la-mise-en-commun-de-donnees-de-l-assurance-complementaire-et-obligatoire/actualites.html>
2. <http://www.constances.fr>

récemment de la nécessité de protéger les enquêteurs des pressions dont ils peuvent être l'objet, notamment économiques et politiques (Laurens et Neyrat 2010), et les enquêtés d'une forme de spoliation de leur contribution à une œuvre commune (Weber 2011). Mais elle demeure valable : en échange de l'accès à des informations personnelles, les enquêteurs protègent les enquêtés en leur garantissant anonymat et/ou confidentialité (Beliard et Eideliman 2008). L'anonymat consiste en la simple modification des caractéristiques directement identifiantes comme le nom. La confidentialité va plus loin, puisqu'elle fait que les personnes ne puissent être identifiées même dans le milieu d'interconnaissance où elles évoluent, et suppose que des caractéristiques plus spécifiques soient brouillées, comme la présence d'un enfant handicapé dans une configuration familiale originale. Ceci peut être fait sans porter atteinte à la rigueur scientifique du résultat par différentes méthodes, comme la construction de cas fictifs pour la présentation des résultats, présentant toutes les caractéristiques essentielles du cas réel sous-jacent, mais modifié sur toutes ses caractéristiques secondaires.

Une démarche scientifique rigoureuse et protectrice des droits des personnes en matière de santé se doit donc de recouper un maximum d'informations, tant à partir de bases de données statistiques, qu'au niveau individuel, tout en garantissant anonymat voire confidentialité lors de la publication des résultats. Par rapport à la situation actuelle, cela revient à élargir, sous contrôle, la « bulle » de confinement des informations aux chercheurs (Le Gléau, page 29), comme cela a été fait en 1984 au profit des informations sur les entreprises. La logique de cette opération est en fait que ces chercheurs sont précisément les seuls à même de pouvoir recouper toutes ces informations pour faire progresser le niveau de connaissance collective.

II. Propositions pour l'accès aux données de santé

Il est donc nécessaire d'accéder aux données statistiques individuelles, ou granulaires, en amont de toute forme d'agrégation ou d'anonymisation. On peut distinguer deux cas de figure, selon que l'enquête commence par des cas concrets ou par l'accès aux données statistiques.

Dans le premier cas, un accord préalable écrit ne peut être demandé aux intéressés. Cette formule est en effet inadaptée à l'enquête en sciences sociales d'origine qualitative. Bien que codifiée dans les chartes de nombreuses associations anglo-américaines de recherche en sciences de la nature voire en sciences sociales, elle demeure en leur sein contestée, et est en fait née des abus des sciences médicales expérimentales (Bosa 2008). Par exemple, des recherches cliniques avaient été faites sans que le groupe témoin bénéficie des avantages thérapeutiques du groupe test, ce qui s'était traduit par le décès d'une grande partie de ses membres. En sciences sociales, cette visée directement thérapeutique étant par définition absente, le schéma n'est pas le même. De plus, le recueil préalable et formel d'un consentement empêche l'établissement d'une relation d'enquête. Les chercheurs américains soumis à ce type de restrictions sont parfois conduits à se positionner pour accéder aux données dans la catégorie juridique des journalistes, qui sont paradoxalement beaucoup plus libres dans leur relation à leurs enquêtés. L'enquête d'origine qualitative procède donc par des contacts personnels, qui pourraient ensuite être retrouvés dans les bases de données. Dans cette situation, l'accord peut-être beaucoup plus informel.

Le second cas de figure fonctionne en sens inverse : quand il est souhaitable de compléter les données statistiques sur certaines personnes par une enquête qualitative auprès d'elles. Comme c'est le cas aujourd'hui pour ce type de démarche, ces personnes pourraient être contactées par courrier pour recueillir leur accord préalable et écrit à toute investigation plus poussée. Un refus ou une absence de réponse de leur part vaudrait bien sûr abandon de tout approfondissement et ne serait consigné nulle part. Si elles sont d'accord, un volet qualitatif pourrait se déployer.

Dans les deux cas de figure, que l'on commence par l'étape qualitative ou quantitative, tous les recoupements d'informations seraient autorisés, et les droits des personnes enquêtées seraient protégés par le passage à l'anonymat voire à la confidentialité avant toute publication ou communication à des tiers (comme les organismes producteurs des données statistiques). Ainsi, en résumé, l'accès aux données individuelles brutes se ferait sous les conditions suivantes :

- que les cas individuels soient uniquement utilisés à des fins de compréhension des mécanismes à l'œuvre ;
- que les cas individuels finalement publiés soient anonymisés et rendus confidentiels, ce qui garantirait le respect des droits des personnes ;
- que cette présentation ne distorde pas les résultats issus du recoupement des données statistiques et d'enquêtes qualitatives ;
- que les personnes ne puissent faire l'objet d'investigations plus poussées à partir de la base de données sans leur accord ;
- que toute information facilitant une forme de profilage des risques permettant d'individualiser davantage la couverture maladie soit exclue de la publication ;
- que le tout s'opère sous le contrôle d'un comité analogue au CSS, représentant les différents intérêts en présence : personnes enquêtées, organismes produisant des statistiques, professionnels de santé, chercheurs, État. Ce comité aura pour charge de veiller au respect de chacun des points précédents.

L'ensemble de la démarche vise donc à permettre un véritable accès aux données de santé, tout en préservant les droits et les intérêts des personnes. L'éthique médicale et celle du chercheur en sciences sociales se retrouvent au moins sur un point – qu'on pourrait qualifier tout simplement de moral dans un sens kantien : avant tout, ne pas nuire. En respectant ce principe, la marge de progression est large en faveur d'un droit à une meilleure connaissance des processus sociaux qui encadrent la santé publique.

Références

- Beaud, Stéphane, et Florence Weber. 1997. *Guide de l'enquête de terrain: produire et analyser des données ethnographiques*. 2010 éd. Guides Repères. Paris: La Découverte.
- Beliard, Aude, et Jean-Sébastien Eideliman. 2008. « Au-delà de la déontologie. Anonymat et confidentialité dans le travail ethnographique ». In *Les politiques de l'enquête. Epreuves ethnographiques*, édité par Alban Bensa et Didier Fassin, 123-42. Recherches - Bibliothèque de l'Iris. Paris: La Découverte.
- Belorgey, Nicolas. 2010. *L'hôpital sous pression: enquête sur le « nouveau management public »*. Textes à l'appui - Enquêtes de terrain. Paris: La Découverte.
- Bosa, Bastien. 2008. « A l'épreuve des comités d'éthique. Des codes aux pratiques ». In *Les politiques de l'enquête. Epreuves ethnographiques*, édité par Didier Fassin et Alban Bensa, 205-27. Recherches - Bibliothèque de l'Iris. Paris: La Découverte.
- Bruno, Isabelle, et Emmanuel Didier. 2013. *Benchmarking: l'État sous pression statistique*. 1 vol. Paris: Zones.
- Desrosières, Alain. 2005. « Décrire l'Etat ou explorer la société: les deux sources de la statistique publique ». *Génèses* 58: 4-27.
- Didier, Emmanuel. « Mettre en responsabilité. "Compstat" et le nouveau management statistique de la Préfecture de police de Paris ». Document de travail.
- Gramain, Agnès, et Florence Weber. 2001. « Ethnographie et économétrie: pour une coopération empirique ». *Génèses* 44: 127-44.
- Laurens, S. et Neyrat, F. éd., 2010. *Enquêter: de quel droit? : Menaces sur l'enquête en sciences sociales*, Bellecombe-en-Bauges: Editions du Croquant.
- Passeron, Jean-Claude. 1991. *Le raisonnement sociologique*. 2006 éd. Paris: Nathan.
- Weber, Florence. 2009. *Manuel de l'ethnologue*. Paris: PUF.
- Weber, Florence., 2011. La Déontologie Ethnographique À L'Épreuve du Documentaire. *Revue de Synthèse*, 132(3), p.325-349.
- Weber, Max. 1917. *Essais sur la théorie de la science*. 1965 éd. Paris: Plon.