

A S S E M B L É E   N A T I O N A L E

X V <sup>e</sup>   L É G I S L A T U R E

# Compte rendu

## Commission des affaires sociales

– Audition, en visioconférence, de M. Dominique Bureau sur  
la statistique et la crise sanitaire ..... 2

Mercredi  
19 mai 2021  
Séance de 16 heures

Compte rendu n° 84

SESSION ORDINAIRE DE 2020-2021

**Présidence de  
Mme Fadila Khattabi,  
Présidente**



## COMMISSION DES AFFAIRES SOCIALES

**Mercredi 19 mai 2021**

*La séance est ouverte à seize heures cinq.*

---

*La commission entend, en visioconférence, M. Dominique Bureau sur la statistique et la crise sanitaire.*

**Mme la présidente Fadila Khattabi.** Depuis le début de la crise sanitaire, les données statistiques ont pris un relief et une importance tout à fait inédits, devenant des indicateurs clés pour la prise de décisions sanitaires, sociales et économiques au fil des semaines. Les statistiques sont entrées très directement dans le débat public, qu'il s'agisse du nombre de décès et de leur cause, du taux d'incidence de la covid, du nombre de tests réalisés, de l'évolution du marché du travail ou de la pauvreté.

Lors d'une crise d'une telle ampleur, les statistiques soulèvent de nombreux enjeux, tels que leur nécessaire réactivité tout en conservant leur niveau de qualité, la coopération entre services statistiques et le renforcement des statistiques sanitaires elles-mêmes, l'utilisation de nouvelles sources de données, par exemple celles des opérateurs de cartes bancaires ou de téléphonie, ou encore la question de l'*open data*.

Pour aborder ces sujets finalement peu explorés, nous recevons aujourd'hui M. Dominique Bureau qui a présidé l'Autorité de la statistique publique (ASP) jusqu'en avril dernier. Je précise que nous avons bien sûr dans un premier temps sollicité la nouvelle présidente de l'Autorité, Mme Mireille Elbaum, qui ne pouvait pas prendre part à l'audition. Celle-ci se tient donc avec son plein accord. Le rapport annuel 2020 de l'Autorité, publié voici quelques semaines, évoque les différents sujets que j'ai cités.

**M. Dominique Bureau.** Le rapport de l'Autorité pour 2020 est centré sur les leçons du covid. Il comprend une partie importante : l'idée d'avoir un projet renouvelé pour la statistique sanitaire et sociale, notamment pour le domaine sanitaire, en lien avec les éléments que vous avez évoqués en introduction.

Je précise que j'interviens bien en plein accord avec la présidente, qui vient de prendre ses fonctions à temps partiel. Notre souci est que nos recommandations fassent l'objet de suites. Je pense que vous aurez l'occasion de dialoguer avec la nouvelle présidente, d'autant plus qu'elle a été directement impliquée dans l'une des missions d'inspection engagées suite à nos demandes.

L'objet de la statistique publique est de mettre à la disposition de tout un chacun une information objective de qualité pour éclairer le débat public, pour éclairer les choix publics et privés ainsi que pour permettre à la recherche de travailler. Nous nous plaçons au niveau de la statistique, c'est-à-dire d'une information construite, et non au niveau des données de gestion, même si les données de gestion sont une source importante pour fabriquer de la statistique, après un certain nombre de retraitements afin de les rendre utilisables pour éclairer les choix et pas seulement pour constater la gestion.

La statistique utilise plusieurs types de sources. Les premières proviennent d'outils dédiés tels que des recensements ou des enquêtes qui ont l'avantage de nous donner la maîtrise intégrale des questions et donc de permettre d'être précis. Ce sont des techniques assez lourdes qui nécessitent des enquêteurs. Nous essayons donc de compléter cette première source par des données existantes et le *big data* offre évidemment beaucoup plus d'ouvertures et une grande multiplicité de sources de données, notamment les sources de gestion. Cela explique que nous disposions en France de services statistiques dans tous les ministères mais cela ne suffit pas puisque certaines données se trouvent aujourd'hui chez les opérateurs.

Les trois sources principales de la statistique publique sont donc les données d'enquêtes approuvées par le Conseil national de l'information statistique (CNIS), les données collectées par les services statistiques ministériels et les données provenant des opérateurs. Ces dernières ne sont intégrées dans la statistique qu'après avoir été labellisées par l'ASP.

Pourquoi effectuons-nous cette labellisation ? Pour que ces données soient utiles au débat public, il est absolument essentiel qu'elles soient de qualité et que le public ait confiance en la manière dont elles ont été construites, qu'elles n'aient pas été biaisées pour une raison quelconque. Assurer ce contrôle est le rôle de l'ASP, suite à la loi de 2008 qui l'a créée avec le double souci de garantir son indépendance professionnelle – sans indépendance professionnelle, la suspicion est immédiate – et de s'assurer que la statistique publique en France soit de qualité, objective et au meilleur niveau.

Nos recommandations sur la statistique sanitaire et sociale interviennent surtout dans ce sens, en faveur d'un projet renouvelé appliquant un code de bonnes pratiques avec des règles assez précises sur ce qui doit être vérifié. Nous appliquons essentiellement deux principes pour formuler nos recommandations : que la statistique réponde aux attentes de l'utilisateur et que les moyens affectés à la statistique soient en ligne avec ce qui est nécessaire pour répondre à ces besoins. Ces deux aspects sont présents dans notre diagnostic mais celui-ci n'est pas uniquement un diagnostic sur les moyens. Notre principal souci reste, fondamentalement, d'avoir une statistique publique qui réponde mieux aux attentes de l'utilisateur.

Je précise, pour éviter un malentendu, que notre recommandation de projet renouvelé sur la statistique sanitaire et sociale ne reflète pas un diagnostic selon lequel quelqu'un aurait commis une faute quelque part. Au contraire, nous constatons que le service de la statistique publique s'est mobilisé de manière générale pour suivre l'activité économique, pour suivre le marché du travail, pour suivre la situation sociale ainsi que dans le domaine sanitaire.

Le domaine sanitaire illustre d'ailleurs la façon de réagir du système statistique, lequel a souvent utilisé des outils existants ou en cours de développement. Typiquement, dans le domaine sanitaire, le service de statistique publique fait depuis toujours des analyses de surmortalité à partir des données de l'état civil collectées par l'Institut national de la statistique et des études économiques (INSEE) ; il a évidemment adapté ces évaluations de surmortalité au contexte du covid et à la demande nouvelle de disposer d'un retour local, à l'échelon des départements voire des communes. De même, en ce qui concerne la prise en charge des patients, la statistique publique suit des panels de généralistes et a obtenu des informations sur la prise en charge durant le covid au travers de ces panels.

En dépit des conditions plus ou moins complexes pour mener des enquêtes, la statistique publique a développé des instruments nouveaux. Ainsi, l'enquête EpiCov

développée avec, d'une part Santé publique France, d'autre part l'Institut national de la santé et de la recherche médicale (INSERM), vise non seulement à connaître l'état de santé de la population en termes de prévalence de l'épidémie mais aussi à recueillir des informations sur les conditions sociales dans le contexte du covid. Par ailleurs, la direction de la recherche, des études, de l'évaluation et des statistiques (DREES) s'est attachée à appuyer le ministère, notamment pour améliorer et consolider les remontées des tests de réaction en chaîne par polymérase (PCR) lorsque la question s'est posée. Le service statistique public s'est donc incontestablement mobilisé mais nous avons constaté que la crise révélait un certain nombre de faiblesses.

Le premier point, constaté dès la fin du printemps 2020, est le fait que, contrairement à d'autres pays, nous n'avons pas de statistiques sur les causes de mortalité en temps réel en France. Les données sur les causes de mortalité sont consolidées par une unité de l'INSERM, le centre d'épidémiologie sur les causes médicales de décès (CépiDc). Nous nous sommes rendu compte qu'il se posait un double problème de conception et de mise en œuvre. Le problème de conception est que, jusqu'à présent et contrairement aux données sur la mortalité qui remontent directement par l'état civil, ces données sur les causes de mortalité étaient vues comme des données dites « structurelles », ceci signifiant de façon un peu caricaturale qu'il n'était pas considéré comme grave qu'elles arrivent deux ans plus tard puisque l'évolution des maladies cardiovasculaires ou du pourcentage de cancers évoluait de façon lente. La crise a révélé que cette situation n'est pas tenable. Il est nécessaire de disposer d'une information consolidée et non fragmentée sur les causes de décès, qui doit être construite à partir de l'ensemble des informations disponibles sur les causes de décès. Par ailleurs, les consolidations de données par la statistique publique étaient particulièrement tardives, souvent en retard par rapport aux exigences européennes de transmission. Certes, les processus de consolidation et de gestion dématérialisée sont compliqués mais le CépiDc était confronté voilà quelques années à un retard qu'il n'a pas rattrapé. Cette situation ne paraissait pas tenable, ce qui a conduit à une mission d'inspection. Elle est en cours et précisera les faits.

L'autre point révélé par la crise,, que nous avons déjà senti poindre voici quelques années en auditionnant le CépiDc, est que les causes de mortalité sont de plus en plus complexes, pas forcément univoques. Nous avons besoin d'informations sur les expositions, sur la morbidité et donc de faire le lien entre morbidité et mortalité. La question des comorbidités est au cœur des sujets liés au covid. Nous disposons de manière un peu cloisonnée des causes de mortalité d'un côté et, de l'autre, de la morbidité qui ne rentrait pas vraiment dans le champ de la statistique publique. La statistique publique se caractérise en effet par le fait de respecter un calendrier avec des arrivées régulières de données tandis que les données de morbidité sont souvent fragmentées. Elles proviennent généralement d'études plutôt que de statistiques et nous aurions sans doute besoin d'un cadre qui permette de les faire rentrer dans le débat. Les débats sur la santé et son périmètre n'ont pas manqué. Au-delà des données démographiques, nous avons donc aussi besoin d'un meilleur suivi de la morbidité dans la statistique publique.

Le responsable de la statistique publique à la DREES a changé et, à chaque changement de titulaire, nous auditionnons le service pour savoir s'il a un projet. Nous avons ainsi constaté que ses moyens avaient de manière tendancielle décréu régulièrement, même s'ils ont un peu réaugmenté durant cette dernière période. Ses missions se sont pourtant élargies avec la création de différents observatoires, dont l'observatoire du suicide, et le développement d'outils pour le modèle de micro-simulation qui permet d'évaluer les réformes de la protection sociale. Il faut préciser le diagnostic entre ces moyens en décroissance et

l'augmentation de la demande mais nous n'avons pas de doute en ce qui concerne deux marqueurs.

Le premier est le fait que des enquêtes « non nécessaires », au sens que leur fréquence n'est pas imposée, peuvent être différées d'un ou deux ans mais, tendanciellement, le calendrier de toutes les enquêtes dérape par manque de moyens. Ainsi, l'enquête « Santé scolaire » a été reportée alors que nous avons bien vu au début du covid que la santé scolaire était un sujet important. Ne pas disposer de données de référence est dommage.

L'autre marqueur consiste en ce que ce service ministériel s'avère uniquement capable de répondre aux obligations internationales alors que le débat public français nécessite plus d'éclairages que ce qui est demandé par les autorités bruxelloises. Nous sommes pourtant quasiment à l'étiage et à nos limites.

Nous avons donc besoin d'un projet renouvelé, construit, coordonné car la bonne statistique nécessite une coordination, pas simplement de piocher des données et de les éclairer en posant sous le réverbère ce que nous pouvons éclairer. Cette coordination est normalement le rôle du service statistique ministériel, qui est producteur et peut être coproducteur, comme dans le cas des données de gestion sur le chômage dont nous disposons en fin de mois.

Il n'est pas forcément grave que des décalages se produisent dans les données de Pôle emploi, par exemple lorsqu'un chômeur a un rappel parce qu'il n'a pas été enregistré à temps mais, pour faire du diagnostic conjoncturel, il est important que les gens soient bien affectés aux mois où ils étaient chômeurs et non à ceux où ils ont touché leurs indemnités de chômage. Un travail est donc nécessaire pour s'assurer que les données de gestion puissent être utilisées en données statistiques. L'un des problèmes, en particulier, est que les statistiques changent lorsque les règles de gestion changent. Il est important que les séries soient homogènes dans la durée de manière à ce que les évolutions puissent être interprétées comme la mesure du phénomène étudié et non comme un artefact ou un changement institutionnel créant une rupture de série. Le rôle du service statistique ministériel est donc important pour coordonner et passer à des données statistiques.

Il ne s'agit pas de faire un projet seulement pour la DREES. Il faut développer un projet d'ensemble pour le système statistique sanitaire et social puisqu'il faut enrichir les données et, notamment, mobiliser plus de données. Sur la morbidité, par exemple, que j'ai déjà évoquée, nous ne disposons pas de données de Santé publique France labellisées dans la statistique publique. Il nous semble que c'est qualitativement une anomalie et je n'ai aucun doute qu'un enrichissement est possible si la DREES et Santé publique France se mettent collectivement au travail de façon à ce que toutes les données de Santé publique France puissent être utilisées ainsi que les données des autres opérateurs.

Certaines données d'opérateurs sont déjà labellisées, par exemple celles de la Caisse nationale de l'assurance maladie (CNAM), mais cela reste très limité. Il nous semble donc nécessaire d'avoir un projet vraiment construit, qui tienne compte des données facilement mobilisables mais prenne aussi soin de ne pas laisser d'angle mort. Il faut mobiliser toutes les données existantes pour avoir une statistique de qualité répondant aux demandes de l'utilisateur. Nous suggérons par conséquent que le service statistique public puisse se mobiliser avec des moyens adaptés.

Ceci donne actuellement lieu à deux missions. L'une porte spécifiquement sur les causes de mortalité et l'autre est une cartographie destinée à recenser l'existant et à l'utiliser au mieux. Mon intuition est qu'un architecte est nécessaire pour réaliser un saut qualitatif, que nous avons trop procédé par « bricolage » en utilisant des données hétérogènes. Les enjeux à traiter dans le domaine sanitaire et social sont trop lourds pour s'en contenter et la crise du covid a montré que la somme de coups de projecteur ne remplace pas une véritable statistique construite. Notre conclusion est qu'il faut construire ce projet en le concevant vraiment comme un projet, c'est-à-dire en répondant aux attentes du public sans chercher à savoir qui a fauté.

**Mme Valérie Six.** La crise sanitaire a bouleversé les organisations de l'ensemble de la société et la statistique publique a été confrontée à ce défi. Je vous remercie de venir échanger avec nous car, si nous connaissons bien l'INSEE, je vous avoue que le rôle et les missions de l'ASP m'étaient moins familiers.

Dans le rapport d'activité 2020, vous évoquez le système statistique européen face à la crise et les mesures de partage de l'information entre les pays. Cette coordination a notamment permis la mise en place par Eurostat d'un tableau de bord en ligne, spécifique à la crise sanitaire. Cependant, vous indiquez dans votre rapport que la comparabilité des statistiques entre pays n'est pas encore acquise, chaque pays ayant dû s'adapter et les délais de remontée des informations ou des politiques publiques étant différents. Il me semble important d'encourager la consolidation des statistiques au niveau européen. Quelles pistes avez-vous identifiées pour parvenir à cet objectif ?

Par ailleurs, pouvez-vous développer le sujet du report de certaines collectes de données, notamment l'enquête annuelle de recensement de la population qui est une des missions essentielles de l'INSEE ? Quelles conséquences ce report a-t-il eues sur les informations transmises aux pouvoirs publics ? Les pistes d'accélération de la numérisation sont-elles en cours avec les collectivités chargées de l'état civil pour améliorer le partage des données ?

Enfin, le rapport d'activité indique que la crise sanitaire a généré un besoin de questionnements nouveaux par rapport aux enquêtes habituelles. J'ai relevé des sujets portant sur la consommation des ménages pendant les confinements, la complexité administrative et les difficultés d'accès au droit ou la participation à la vie associative. Qu'en est-il du développement du télétravail et de ses impacts sur l'état de santé des personnes, sur les modes de transport ou le logement ?

**Mme Bénédicte Pételle.** Face à cette crise sanitaire aux conséquences exceptionnelles et au besoin de connaître la situation de façon plus précise qui se fait plus que jamais ressentir, nous ne pouvons que nous réjouir de la réactivité du service public de la statistique. Sur toutes les questions importantes de santé, de pauvreté ou d'éducation devenues encore plus sensibles dans ce contexte, il a su contribuer à apporter un éclairage décisif aux décideurs et nous permettre de prendre les meilleures décisions. Nous souhaitons aussi saluer l'indépendance de ce service, indépendance à laquelle veille l'ASP.

Alors que nous avons tous ressenti ce besoin d'enrichir la statistique publique dans des moments où la situation devenait moins claire, où l'imprévu venait bouleverser nos politiques, nous devons noter la baisse de 12 % en dix ans des effectifs et des moyens des différents services ministériels de statistique comme vous le dites dans votre rapport. Ainsi, même si la DREES relevant du ministère des solidarités et de la santé a vu une stabilisation récente de ses

moyens, son activité statistique a été réduite, notamment en ce qui concerne les statistiques sur la prévention en santé scolaire. Particulièrement en première ligne durant cette crise, il faut noter le renforcement de son appui auprès des différents ministères sociaux avec la mobilisation de quinze à seize agents pour la gestion de crise et son renforcement en tant que centre de référence pour assurer la qualité et la cohérence des données.

Cette crise a obligé les différents services de la statistique à s'adapter. Vous avez parlé dans votre rapport des nouvelles conditions de travail avec le télétravail et les études à distance, sans rendez-vous personnel possible. Vous évoquez aussi le sous-équipement informatique dans certains services ministériels. Nous aurons besoin d'investissements à l'avenir pour empêcher que les décideurs deviennent aveugles du fait de l'arrêt pur et simple d'études statistiques dans leurs ministères.

Dans votre rapport concernant l'année 2020, vous préconisez de repérer les angles morts. J'attire votre attention à ce sujet sur le manque de données concernant la protection de l'enfance comme j'ai pu le constater lors de mes missions sur l'aide sociale à l'enfance. Ces statistiques seraient un point d'appui pour faire évoluer l'aide sociale à l'enfance.

Durant cette crise, la jeunesse en général a été pénalisée. Le sujet du mal-être des jeunes a été régulièrement commenté dans l'opinion publique. Pouvez-vous nous dire quels travaux ont été menés sur ce sujet par les différents services de la statistique publique ? Qu'en est-il ressorti ?

**M. Jean-Pierre Door.** En tant que médecin, j'ai toujours été attaché aux statistiques médicales et aux données de santé, dans le cadre de l'activité médicale et des connaissances scientifiques. Nous avons chaque année des suivis statistiques et nous pouvions, année après année, avoir des tableaux de la mortalité en général.

En 2020, nous avons constaté une augmentation de 9 % environ de la mortalité, toutes causes confondues. Cette surmortalité est évaluée à un peu plus de 50 000 décès. Disposez-vous des tableaux – qui sont toujours à peu près les mêmes chaque année – avec 23 à 25 % de décès liés aux affections cardiovasculaires, 24 à 25 % liés aux maladies cancéreuses ? Les détails du reste sont-ils déjà évalués et pouvons-nous les analyser ?

Vous avez par ailleurs évoqué certaines faiblesses dans l'évaluation des causes de la mortalité et de la surmortalité. Y incluez-vous ce que j'appelle la « surmortalité invisible », c'est-à-dire celle dont nous ne connaissons que très peu les causes exactes ? Cela permet de savoir si les certificats de décès sont réellement opérationnels. N'existe-t-il pas des défauts généraux dans l'analyse des certificats médicaux de décès ? J'ai eu connaissance de personnes qui nous disent que, parfois, la cause covid a été imputée au décès de personnes de leur famille qui n'étaient en réalité pas atteintes du covid. Le terme « covid » avait pourtant été noté sur le certificat médical de décès. Il existe donc parfois des défauts dans le cadre des certificats de décès.

Comment pourrions-nous avoir une plus grande précision dans la définition des causes de décès dans les certificats médicaux ? Ils relèvent à la fois des hôpitaux ainsi que des médecins de médecine générale et de toutes les spécialités médicales et sont déposés en général en mairie. C'est au niveau des mairies également que nous pouvons avoir le nombre de décès. Toute cette analyse est assez difficile et complexe mais je souhaite savoir ce que vous en pensez.

**Mme Agnès Firmin Le Bodo.** Je tiens d'abord à saluer la qualité du travail que les services statistiques ministériels conduisent depuis le début de la crise sanitaire sous le contrôle de l'ASP. Je crois pouvoir dire que tous nos concitoyens sont maintenant très attachés à ces chiffres régulièrement donnés et y prêtent une attention toute particulière.

La capacité à produire une information pertinente, objective et quotidiennement mise à jour s'est révélée particulièrement précieuse pour mieux apprécier la situation et guider les décisions que le Gouvernement et le législateur ont été amenés à prendre. Les statistiques de santé ont rarement fait l'objet d'autant d'intérêt, de commentaires et d'interprétations.

À l'heure où la parole officielle est sans cesse relativisée, nous pouvons nous réjouir de disposer d'un juge de paix impartial pour trancher la bataille des chiffres. Transparence, indépendance et respect de la déontologie sont des principes cardinaux pour assurer la confiance de nos concitoyens dans la statistique publique. L'Autorité y veille et sa mission est plus que jamais nécessaire.

Cette situation de crise sanitaire a aussi mis en lumière les enjeux spécifiques à la production de statistiques de santé. Comme pour les indicateurs économiques, un besoin d'une comparaison internationale des indicateurs sanitaires s'est rapidement fait jour. Je pense par exemple aux comparaisons du nombre de personnes décédées, du nombre de tests réalisés ou du nombre d'injections de vaccins. Pouvez-vous nous indiquer si un cadre méthodologique partagé encadre la production des statistiques sanitaires, ne serait-ce qu'au niveau européen ? Par exemple, la prise en compte ou non des personnes décédées au domicile ou en établissement d'hébergement pour personnes âgées dépendantes (EHPAD) peut changer sensiblement la donne et biaiser les comparaisons internationales.

Pouvez-vous nous indiquer si votre Autorité travaille avec les « GAFAs », comme Google, qui a construit des indicateurs permettant de suivre au jour le jour les chiffres de l'épidémie ? Comment veiller à une bonne exploitation de la statistique publique par ces plateformes ?

**Mme Jeanine Dubié.** J'ai pris connaissance avec beaucoup d'intérêt du rapport 2020 de l'ASP puisque je méconnaissais ses activités. Vous soulignez dans ce rapport les limites des statistiques sanitaires révélées par la crise covid. Selon vous, ces statistiques ne doivent pas se limiter à l'aspect structurel mais s'élargir grâce à un projet d'ensemble défini par les différents opérateurs et nécessairement piloté par le service statistique ministériel concerné. Vous évoquez la DREES, qui mérite, toujours selon ce rapport, d'être renforcée.

Pourriez-vous expliciter votre point de vue sur cette question des statistiques en matière de santé ? Quelles pistes d'évolution suggérez-vous pour améliorer nos connaissances en matière sanitaire ? Quels sont les liens avec la Commission nationale de l'informatique et des libertés (CNIL) en matière de traitement des données de santé ? Enfin, vous suggérez de renforcer la DREES. Parlez-vous de moyens humains, financiers ou techniques ?

Selon votre rapport toujours, la crise a mis en lumière la nécessité d'une plus grande transparence statistique et donc le besoin d'une meilleure utilisation par les autorités administratives ou gouvernementales de la statistique publique pour l'efficacité des politiques publiques. Vous recommandez de renforcer la confiance. Quelles seraient vos propositions, à la fois pour améliorer l'utilisation des statistiques par les décideurs publics et pour faciliter l'accès aux données statistiques et leur déchiffrement par les pouvoirs publics et le grand



public ? Comment concevez-vous une meilleure coopération autour des statistiques publiques ?

Enfin, vous avez évoqué l'international. Comment pourrions-nous améliorer les comparaisons internationales ? En tant qu'Autorité de la statistique publique, avez-vous des liens avec nos voisins européens ?

**M. Pierre Dharréville.** Vous avez évoqué des angles morts. Dans quels domaines devrions-nous progresser afin de mieux connaître, par le biais de statistiques, l'état de notre société ? Je pense à la situation des jeunes, dont beaucoup sont hors des radars du point de vue de leur situation sociale. Je pense aussi à la qualité de notre connaissance de la santé psychique de nos concitoyennes et concitoyens. Je ne sais pas si vous êtes en mesure de porter un regard plus précis sur ces angles morts, ce qui serait intéressant.

J'imagine que le premier point pour faire de la statistique est de savoir ce que nous voulons évaluer pour mettre ensuite en place les outils nécessaires. Avez-vous constaté, notamment à la faveur de la crise, une évolution des demandes en matière de statistiques ?

Enfin, nous avons tout de même été beaucoup submergés de chiffres durant cette période, au risque qu'il devienne difficile de faire le tri dans la vie publique. Que pensez-vous de l'usage qui a été fait des statistiques durant cette période dans le débat public mais, aussi, de l'usage gouvernemental ?

**M. Bernard Perrut.** Je salue tout d'abord la capacité de réaction de l'INSEE et des services statistiques ministériels pour fournir des indicateurs qui permettent d'apprécier la situation et, plus généralement, de documenter un certain nombre de répercussions.

Quelles leçons pouvez-vous tirer de cette période sanitaire que nous venons de vivre ? Comment garantir l'information nécessaire en temps de crise ? Quelles adaptations du cadre technique et juridique sont-elles à prévoir ?

La crise a aussi révélé certaines fragilités et des lacunes, en premier lieu en matière de statistiques sanitaires. Quelles recommandations formuleriez-vous ? Le covid-19 est responsable de près de 108 000 décès en France, un peu plus d'un an après le début de l'épidémie. Ce bilan dramatique pour des dizaines de millions de familles ne correspond cependant pas exactement au nombre réel de patients décédés du covid-19 puisque le nombre total de décès, mis à jour quotidiennement par les autorités sanitaires, cumule les décès enregistrés à l'hôpital et ceux en EHPAD ou en établissements médico-sociaux à travers deux systèmes de remontée d'information différents, tandis que les morts à domicile ne sont pas pris en compte de la même façon. Savez-vous quelle est la proportion de ces décès ?

Le centre d'épidémiologie sur les causes médicales de décès de l'INSERM analyse les certificats de décès reçus par courrier, par mail ou par papier avec plusieurs mois de retard. Qu'en est-il de la prise en compte de ces décès en temps réel ? Avant la crise, il n'existait pas d'outil de décompte de décès en EHPAD au jour le jour. Si leur remontée se faisait vite entre l'EHPAD et la mairie, il fallait parfois plusieurs mois pour que l'information remonte jusqu'aux services à l'échelon national. En mode de fonctionnement normal, il faut un an et demi à deux ans pour avoir les vrais chiffres de la mortalité en EHPAD.

Par ailleurs, monsieur, nous savons que tous ces décès ne sont pas directement dus au covid-19. À l'hôpital, l'expression « patient covid décédé » a été utilisée pour désigner des

malades ayant perdu la vie alors qu'ils étaient porteurs du virus. En EHPAD, la situation est encore plus compliquée puisque des résidents décédés ont parfois été catégorisés parmi les morts du covid sans avoir été testés positifs, simplement parce qu'ils présentaient les symptômes de la maladie. Pensez-vous que cela ait conduit à une surévaluation de la mortalité covid, notamment au début, lorsque toute personne qui mourait avec de la fièvre ou des difficultés respiratoires était qualifiée « covid » par le médecin ? Sur l'ensemble de l'année, les statistiques sont-elles bien le reflet de la réalité ? Tel est le sens des questions que je voulais vous poser aujourd'hui.

**Mme Monique Limon.** Je voudrais appuyer ce qu'a dit ma collègue Bénédicte Pételle sur les statistiques que nous pouvons qualifier de « pas précises » en ce qui concerne les jeunes placés à l'aide sociale à l'enfance. Tout d'abord, comment faire pour arriver à obtenir des statistiques ? Il faut bien évidemment un travail avec les départements, dont c'est la compétence, mais est-ce suffisant ? Depuis le temps que nous en parlons, y arriverons-nous un jour ?

Comment faire pour disposer de statistiques fiables sur ces jeunes placés à l'aide sociale à l'enfance pendant la pandémie ? S'est-il produit plus de violences, plus de fugues ? J'ai eu quelques remontées de terrain dans ce sens. Au contraire, les placements ont-ils été plus stables ? Les séjours dans les familles ont-ils repris ? Comment pouvons-nous analyser avec des statistiques fiables cette période en ce qui concerne ce public ?

**M. Dominique Bureau.** Je ne pourrai pas répondre précisément à de nombreuses questions.

Je donne d'abord quelques éléments sur le fonctionnement de l'Autorité de la statistique publique. L'Autorité a été créée en 2009 et nous en sommes à la troisième présidente. Notre système statistique français présente la particularité que la base du système est constituée d'une part de l'INSEE, une direction générale du ministère de l'économie, et, d'autre part, des systèmes statistiques ministériels (SSM). Cette organisation a été créée par la loi de 1951.

Ce système peut sembler un peu curieux du fait de la grande proximité du politique, puisque quatre chefs de SSM, dont celui de la DREES, sont des directeurs d'administrations centrales nommés selon les règles applicables aux directeurs d'administrations centrales. Pour autant, la France a dans ce cadre développé un système de qualité appuyé notamment sur les corps de statisticiens de l'INSEE et cette conviction est assez largement partagée en France. Certains soulignent qu'un bon système statistique doit d'abord être constitué de bons professionnels et à l'écoute des attentes des citoyens et du public en général ; c'est quand même la base.

Le dispositif européen a apporté une attention à la qualité, au fait de respecter des règles strictes de méthodologie... Le système français a donc innové en conservant la tradition d'un système statistique publique fondé sur le principe que le citoyen devait avoir confiance dans l'État et qu'il faut organiser l'État pour que l'on puisse avoir confiance en lui. Sans jeter le bébé avec l'eau du bain, nous avons essayé de tirer profit des exigences européennes tout en conservant les atouts du système français.

Il manquait toutefois une pièce maîtresse : un contrôle de l'indépendance. Un contrôle externe a donc été créé et la France est l'un des rares pays à avoir mis en place un contrôle

externe par une autorité indépendante. Elle est très légère puisque constituée d'un président à temps partiel et d'un rapporteur mais elle permet d'avoir une bonne séparation des tâches.

Le système est finalement équilibré. La production statistique est réalisée par le système statistique public, c'est-à-dire l'INSEE et les services statistiques ministériels. La tendance actuelle, comme beaucoup de vos questions conduisent à le souligner, est le besoin de coordination. Le système statistique, comme toute l'administration, était un peu fragmenté et cloisonné. Le fait que le système statistique doit être coordonné est une des leçons de la crise, en particulier pour celles de vos questions sans réponse immédiate ; sur les jeunes par exemple, le système statistique de l'éducation nationale s'est impliqué.

La marque « statistique publique » doit être identifiée à l'égal de l'INSEE. Parfois, les gens croient que la source est l'INSEE et ne savent pas que c'est la statistique publique. C'est une bonne chose car cela prouve qu'ils ont la même confiance dans la qualité des données mais, fondamentalement, nous avons besoin de renforcer ce système. Le règlement européen assigne à l'INSEE la mission de coordonner le système statistique public. L'une des tâches de l'ASP ces dernières années a d'ailleurs été de mieux insérer les services statistiques ministériels dans leur ministère avec des décrets mais des décrets qui reconnaissent l'indépendance professionnelle et le rôle de coordination de la statistique publique. Je pense que cela sera précieux pour les années à venir.

Sur un sujet tel que la délinquance par exemple, nous n'avons pas pu faire cette année l'enquête sur la victimisation, qui est une enquête en face-à-face, car il faut être sûr que la personne qui répond est vraiment la personne enquêtée, en particulier pour la mesure des violences intrafamiliales. Curieusement, c'est au travers de l'enquête EpiCov que nous avons essayé de rattraper ce manque en insérant dans cette enquête des questions sur ce sujet. Nous avons donc tout intérêt à essayer de mieux coordonner et de mieux mailler le travail des différentes statistiques ministérielles sous l'égide de l'INSEE.

Le rôle de l'ASP est vraiment d'assurer un contrôle externe. Le rôle exécutif est confié à l'INSEE et aux SSM sous la coordination de l'INSEE. En revanche, à chaque incident, l'ASP peut rendre publics tous ses avis. Elle a une compétence assez large pour aller examiner tout ce qui pourrait constituer un incident ou une entorse au respect des différents principes. Elle est chargée à la fois de l'indépendance professionnelle et de la qualité. Elle fait par ailleurs chaque année un rapport adressé au Président de la République, aux présidents des deux assemblées... Finalement, le système français est donc bien équilibré.

L'une des questions posées en 2008 était de savoir s'il fallait regrouper le contrôle extérieur avec le CNIS, une instance d'orientation et d'écoute des attentes des différentes parties prenantes qui définit le programme de la statistique publique. Finalement, ce sont bien deux métiers assez différents. Le cadre institutionnel mis en place en France est intéressant car il oblige à un dialogue, l'ASP devant rendre son rapport annuel après avoir auditionné le directeur général de l'INSEE et le président du CNIS.

L'idée est donc d'avoir un cadre institutionnel balancé, assez similaire à celui qui existe au niveau européen, puisque l'équivalent de l'INSEE, Eurostat, est un coordonnateur de données – mais pas un producteur – et est accompagné par deux instances, d'une part le conseil consultatif européen pour la gouvernance statistique (ESGAB), lequel est l'équivalent de l'ASP, d'autre part le comité consultatif européen de la statistique (ESAC), qui est une instance de consultation en amont. L'organisation est donc la même avec l'exécutif, la consultation en amont pour définir les programmes d'enquête et le contrôle *ex post*.

Nous pensons que le système statistique progresse dans ce cadre institutionnel et la manière dont il s'est saisi des nouvelles données est plutôt exemplaire. Les statisticiens sont par exemple assez fiers de ce qu'ils ont réalisé sur les mouvements de population au moment du premier confinement. Il se trouve que l'INSEE développait alors des partenariats avec les opérateurs de téléphonie mobile pour savoir comment utiliser leurs données. Cela s'est révélé immédiatement utile puisque cela a permis d'observer les mouvements de population. De manière générale, nous pensons que le système est équilibré et qu'il faut surtout le conforter, notamment par la capacité de professionnels.

Cela ne signifie pas que certains points ne doivent pas être changés. Ainsi, la mission Bothorel avait étudié dans le passé le passage au *big data*. Nous pouvons aujourd'hui passer assez facilement au *big data* puisque les services statistiques ministériels peuvent avoir accès aux données du SMDS (*switched multimegabit data services*), par exemple. Nous avons aussi la possibilité d'obliger à répondre aux enquêtes. La loi sur le numérique a assuré la transition puisqu'elle permet de passer au dématérialisé pour consolider les enquêtes dont les données étaient obtenues antérieurement par des méthodes non dématérialisées.

Toutefois, nous n'avons pas imaginé que le système statistique soit en fait en évolution permanente. Assurer uniquement la consolidation de l'existant ne répond donc pas tout à fait au problème. L'attente du public augmente et il faudrait pouvoir étendre la possibilité de rendre obligatoires certaines enquêtes ou pérenniser des dispositifs liés au covid qui sont appelés à durer. Il faut finalement agir à la marge d'un système institutionnel qui, fondamentalement, fonctionne assez bien et est équilibré.

Depuis toujours, le principe de la statistique est que les données produites par la statistique publique sont accessibles à tous, aux ministres comme aux citoyens et aux citoyens comme aux ministres, exactement dans les mêmes conditions. Il n'existe pas d'accès privilégié, mis à part les embargos qui permettent aux ministres de ne pas être mis en difficulté par des journalistes qui les interrogeraient un matin à huit heures et cinq minutes sur un chiffre sorti à huit heures. Ces accès privilégiés sont explicites et sont limités.

La contrepartie au fait que nous obligeons les gens à répondre à certaines enquêtes est que nous leur garantissons le secret et le respect de la vie privée. Le système statistique français est donc fondé sur un bon équilibre et ne peut pas être un moyen de contourner des obligations fixées par la CNIL pour la confection de certains fichiers. Le double principe reste l'accès de tous à ce qui est produit par la statistique dans le respect de la protection de la vie privée, cette protection étant plutôt renforcée dans le cas des enquêtes INSEE.

Cela n'empêche pas que nous cherchions à renforcer l'accès pour la recherche. L'INSEE est impliqué dans l'*open data* gouvernemental et s'est doté d'un dispositif dénommé Centre d'accès sécurisé aux données, qui vise à favoriser l'accès des chercheurs aux données de l'INSEE. À l'inverse, lorsque se pose un enjeu de protection de données, cela passe forcément par le comité du secret, qui se réunit régulièrement de façon à concilier ces deux objectifs contradictoires d'accès aux données et de protection.

S'agissant des causes médicales de décès durant le covid, une mission d'inspection est en cours. Je ne pense pas que des erreurs se soient produites. Simplement, au début du covid, les résidents des EHPAD n'ont pas été testés et la déclaration a été faite sur le fondement des symptômes, avec sans doute en cas de comorbidités une certaine variabilité selon la personne signant le certificat de décès pour savoir si elle privilégiait le covid ou la comorbidité connue.

L'objectif de la mission en cours sur les causes médicales de décès est de construire un dispositif permettant de répondre à ces questions.

Je pense que cela occupera les chercheurs pendant un certain temps car les questions qui se posent sont très compliquées. Il existe, en plus de la surmortalité du covid, des surmortalités induites par le fait que des personnes ne se sont pas soignées mais, *a contrario*, un certain nombre de décès ne se sont pas produits. Des recherches avec des modèles de simulation seront nécessaires pour préciser progressivement cette surmortalité et toutes ses facettes afin de parvenir à un bilan net vraiment détaillé tenant compte des différents phénomènes.

Le dispositif d'enregistrement statistique des causes de décès ne répond qu'avec dix-huit mois de délai et pas du tout en temps réel. L'une des priorités d'un projet statistique est sans doute de remettre à niveau ce processus. En plus du certificat de décès, nous disposons d'autres sources ; des éléments sont notamment disponibles à la CNAM. Toutefois, le fait de disposer de beaucoup d'éléments partout ne permet pas de créer un système statistique sans une organisation qui étudie comment les champs se recouvrent.

La comparabilité des statistiques à l'échelon européen est un sujet récurrent. Malheureusement, il n'existe pas de définition européenne de ce qu'est un EHPAD par exemple et, même sur des sujets économiques comme celui des entrepreneurs individuels, la France et l'Allemagne n'ont pas exactement la même manière de comptabiliser la partie entreprise et la partie ménage de l'entrepreneur individuel. Procéder à cette normalisation est le travail d'Eurostat et c'est l'une des difficultés des comparaisons internationales. Le dispositif n'était pas en place antérieurement et, si ce type de pandémie s'inscrit à l'avenir dans une tendance, nous devons nous doter d'outils pour faire des comparaisons. Je pense que ce sera le cas dans les mois qui viennent.

En ce qui concerne la collecte des recensements, le problème est que les enquêtes en face-à-face n'ont pas pu avoir lieu. Il n'était pas imaginable que les enquêteurs de la statistique publique soient à l'origine d'un cluster. La question se posait de savoir si une situation de crise justifie de s'affranchir des règles. Finalement, les évolutions des programmes statistiques ont été approuvées par le CNIS avec des procédures d'urgence qui préexistaient et nous avons donc pu actionner des leviers sans que ce soit simplement le directeur général de l'INSEE qui décide de supprimer la collecte de recensement de l'année 2020. Il faudrait que vous interrogiez plus précisément le directeur général de l'INSEE à ce sujet. Sur un an, l'INSEE peut faire des projections et croiser différentes informations pour disposer des données de recensement les plus nécessaires, même si ce sont des données provisoires, de façon à assurer la continuité des données.

La question des modifications de comportement sur le télétravail, les transports, etc. sera le sujet de travaux par le système statistique. Le rapport *France, portrait social* a essayé de donner une première synthèse des éléments collectés durant la crise mais beaucoup d'autres viendront. Verrons-nous un retour à la normale sur le télétravail ? Sans doute pas et les enjeux sont importants puisque tous ne sont pas égaux devant la possibilité de télétravailler. Cela aura aussi des conséquences sur le système de transports, sur les formes des villes...

Madame Pételle, vous avez cité la baisse des moyens de 12 %, la santé scolaire... Notre premier message est tout de même le besoin d'un projet plus construit pour la statistique sanitaire. Il faut que les besoins soient adaptés à ce projet mais je suis sûr

que, lorsque nous remettrons par exemple à niveau le système sur les causes de décès, il faudra y mettre les moyens. L'ordre de marche de l'ASP n'est pas de venir en appui à des demandes budgétaires ; notre souci est que les projets soient pertinents, efficaces et que les moyens soient mis sur ces projets.

Le mandat du président de l'Autorité est de six ans et est non renouvelable. En six ans, nous n'avons jamais écrit qu'un SSM n'avait pas assez d'argent. Dans le cas de l'INSEE, nous avons plaidé voici quelques années pour qu'il ait un contrat d'objectifs et de moyens ; c'est le cas maintenant et tout le monde s'en félicite car cela permet une meilleure gestion mais nous n'avons pas écrit qu'il manquait structurellement d'argent. Le code de bonnes pratiques nous demande d'ailleurs à la fois de nous assurer que les moyens sont suffisants et que l'argent est utilisé efficacement.

Dans la gestion de crise, que la DREES soit au front et accompagne le ministère pour aider à la remontée des données était évidemment utile. L'arbitrage est souvent assez difficile pour savoir qui doit être intégré ou non à la cellule de crise. Il est délicat de trouver le partage optimal entre ceux qu'il faut protéger pour qu'ils fournissent de la statistique de qualité et ceux qui sont plus au four et au moulin. Le choix fait a été d'assurer d'abord l'appui nécessaire pour la gestion de crise mais nous voyons bien que, pour avoir de la statistique de qualité, il faut que certains aient le temps de se concentrer sur la production des chiffres et d'assurer la méthodologie.

CovidTracker a montré une grande agilité pour assurer une diffusion de chiffres correspondant très bien aux canons de la statistique, pour produire des chiffres interprétables dans lesquels les variations hebdomadaires sont bien corrigées. Pour avoir des projets qui fonctionnent, il faut que les responsabilités et les tâches soient bien affectées, que les gens puissent se concentrer sur leur tâche principale.

En tant qu'ancien président de l'ASP, je ferai passer le message des questions que vous posez sur la protection de l'enfance comme sujet de préoccupation particulière. Je pense que vous avez raison car la question de savoir comment les enfants sortiront de cette crise est un sujet très grave. Vous avez ainsi raison de dire qu'il faut développer les statistiques dans ce domaine.

Le problème des certificats de décès pose aussi la question des polyopathologies. Pour faire de la bonne statistique sanitaire, il faut des personnes qui sachent produire de la bonne statistique, faire remonter les données des EHPAD et des mairies, mais aussi quelqu'un qui vérifie la cohérence du certificat de décès. C'est l'une des raisons qui avaient conduit à s'appuyer sur le CépiDc, notamment pour avoir des capacités en sémiologie. La statistique thématique nécessite à la fois des compétences de personnes qui connaissent ce qui est mesuré et de personnes qui appliqueront des méthodes statistiques aux données. Le dispositif doit conserver cet équilibre. Sans rendre plus administratif le certificat de décès, il faut toutefois qu'il fournisse une information exploitable. En ce qui concerne les données de mortalité, la normalisation internationale est assurée en grande partie par l'Organisation mondiale de la santé et pas seulement par Eurostat.

Une question importante a été posée sur l'utilisation des GAFAs. Le service statistique a mis en place un *data lab* et essaie d'utiliser au mieux les nouvelles données. Toutefois, pour avoir des données qui permettent réellement d'éclairer la décision publique, il est important qu'elles soient sans biais, robustes dans ce qu'elles mesurent et comparables sans rupture de

série. Or, les données provenant du numérique, même si elles sont intéressantes, sont souvent affectées de biais, en particulier parce qu'elles sont financées par la publicité.

C'est une question lancinante pour la statistique publique depuis des années. Même dématérialisées, les enquêtes sont chères ; la question se pose donc de se passer des enquêtes et d'utiliser plus directement les chiffres. Comme l'explique notre rapport, c'est ce que nous faisons et un projet a été conçu pour utiliser les données de caisse pour l'indice des prix. Malgré les espoirs que nous avons sur la possibilité d'éviter de grosses enquêtes en utilisant mieux les sources de données administratives au sens large, y compris celles des opérateurs, la mise en perspective avec l'augmentation des attentes des citoyens conduit tout de même à la conclusion que les différentes sources sont complémentaires plutôt que substituables. Nous ne pouvons pas lâcher la proie pour l'ombre et, pour avoir des données robustes qui renseignent vraiment sur les évolutions structurelles, il faut avoir mis en place l'instrumentation statistique nécessaire. Les données des GAFA permettent donc d'enrichir la statistique mais pas de se passer d'enquêtes qui restent irremplaçables.

En ce qui concerne les moyens humains et financiers, notre diagnostic est qu'il faut construire un projet. C'est aux responsables de l'exécutif de le construire et c'est eux que vous devrez auditionner pour savoir si le projet a avancé et connaître les directions prises. Selon les endroits, les besoins peuvent porter sur les moyens humains ou sur les moyens financiers. Les moyens financiers sont des moyens pour faire des enquêtes. Les moyens humains consistent à avoir des capacités statistiques. Nous avons besoin de moyens humains, c'est-à-dire de personnes compétentes en statistique. Malgré tout, certaines enquêtes sont faites avec des tours de table trop compliqués, ce qui rend malheureusement plus aléatoire leur fréquence.

En ce qui concerne les angles morts, la tendance jusqu'à présent a été d'utiliser au mieux les données facilement accessibles en espérant que, avec un peu d'astuce, nous parviendrions à construire un système. Nous avons par exemple labellisé les données de médicaments de la CNAM mais elles ne concernent que les médicaments qu'elle rembourse. Ces données ne tiennent compte ni de l'hôpital, ni des médicaments non remboursés. Nous pouvons essayer de compléter mais il faut à un moment s'assurer de la cohérence des données et leur fragmentation devient problématique.

Nous insistons donc vraiment sur cette idée de projet. Il existe beaucoup de données dans le système sanitaire français et dans le système social français mais elles sont trop fragmentées. Il faut créer une architecture et que quelqu'un pilote l'ensemble. Il ne faut pas simplement mieux bricoler mais aussi s'assurer des angles morts, par exemple les jeunes.

La prolifération des chiffres est réelle. Effectivement, nous ne devons pas seulement être bavards et avoir plein de chiffres. Par exemple, nous avons labellisé certains chiffres de Pôle emploi. *A contrario*, nous n'avons pas labellisé des données de gestion ou de suivi de plans d'action qui ne nous semblent pas avoir suffisamment de pérennité pour venir dans la statistique comme instrument d'éclairage.

Si nous mesurons deux objets différents, nous devons avoir deux séries différentes. Si nous mesurons un seul objet, plusieurs sources peuvent être utilisées mais le code de bonnes pratiques de la statistique insiste que pour les différentes sources soient comparées et réconciliées. Il existe suffisamment de raisons pour que les personnes soient submergées de chiffres et que les mauvais chiffres l'emportent sur les bons pour que nous évitions d'avoir trop de chiffres.

Nous avons par exemple, dans le passé, deux services qui produisaient des chiffres de l'emploi à partir des mêmes données mais selon deux méthodologies statistiques distinctes. Ils arrivaient à des résultats assez proches mais différents ce qui était totalement incompréhensible par le public. Nous y avons mis fin. Pour les chiffres du chômage, les chiffres de Pôle emploi et les chiffres de l'enquête INSEE/Bureau international du travail (BIT) sont complémentaires mais nous avons demandé à l'INSEE de réconcilier les chiffres et de vérifier que les personnes répondaient sur leur situation au chômage dans l'enquête BIT d'une façon conforme à leur enregistrement à Pôle emploi. Si ces chiffres diffèrent, c'est parce qu'ils ne mesurent pas la même chose. Il est donc important de coordonner les résultats pour disposer de statistiques et non pas de plein de chiffres.

Sur l'adaptation du cadre juridique, j'ai évoqué la pérennisation des données mais, pour moi, le premier problème est de bien faire fonctionner le système statistique dans le cadre institutionnel existant. Nous pouvons l'améliorer par le développement de la cartographie des risques pour nous poser par exemple la question de la résilience. Je ne pense pas que la mesure des causes de décès puisse se faire en temps réel mais j'espère que nous aurons *ex post* des analyses plus fines et des systèmes statistiques plus appropriés.

Madame Limon, je transmettrai à Fabrice Lenglard, chef du SSM, et à Jean-Luc Tavernier, directeur général de l'INSEE, votre question sur ce qui touche aux jeunes dans toutes les facettes : le sanitaire, la victimisation, le travail, l'éducation... C'est une question tout à fait actuelle et légitime que de savoir comment les jeunes sortiront de cette crise.

**Mme la présidente Fadila Khattabi.** Je vous remercie pour votre belle présentation et vos réponses aux questions des collègues.

*La séance s'achève à dix-sept heures vingt-cinq.*