

La construction accélérée d'un système d'information épidémiologique

Par **Maurice RONAI**

Ancien membre de la CNIL

« Lors de la première vague, on avait des indicateurs assez rustiques », Jean Castex, Premier ministre, le 26 août 2020.

L'émergence du SARS-CoV-2 a pris de court le système de surveillance épidémiologique. En février 2020, aucun des indicateurs scrutés quotidiennement un an plus tard n'était encore calculé. Le Plan pandémie grippale de 2011, il est vrai, ne consacrait que deux lignes aux moyens de surveillance épidémiologique⁽¹⁾.

La Mission Pittet, chargée d'évaluer la gestion de la crise sanitaire, suggère que « la prise de conscience des risques de débordement des capacités hospitalières ne serait pas venue des canaux "normaux" de surveillance sanitaire, mais d'initiatives individuelles (à partir du 10 mars) de médecins cliniciens (en particulier en contact avec des collègues italiens), ainsi qu'appuyés par des travaux de modélisation, notamment ceux de l'Imperial College (Londres) et de quelques experts de l'AP-HP et de l'Institut Pasteur »⁽²⁾.

Il a fallu, en quelques semaines, adapter des systèmes d'information déjà existants, les faire converger, étendre la couverture de certains d'entre eux, mais aussi en créer de nouveaux, de toutes pièces, et en urgence.

Initialement centré sur les seuls cas graves et les morts, sur les hôpitaux et les urgences, le système de surveillance a été, au fil des mois, étendu à la population générale : aux millions de personnes testées et, depuis janvier 2021, aux personnes vaccinées.

Autant de systèmes d'information, d'applications, de questionnaires auto-administrés et d'enquêtes qui ont, progressivement, permis de rendre visibles :

- les capacités d'accueil et les tensions hospitalières ;
- la mortalité ;
- la circulation virale, sa cinétique et sa répartition géographique ;
- la proportion de Français qui avaient été touchés par le virus ;
- les *clusters* ;
- le déploiement des campagnes de tests, puis celui de la campagne vaccinale ;
- l'émergence ou la propagation des variants.

Février-Mars 2020 : un système épidémiologique bousculé

« La dynamique de l'épidémie a pris de vitesse tant les systèmes de surveillance sanitaire que les entités chargées de l'anticipation et du déploiement des contre-mesures. Le retard d'appréciation

(1) « Des systèmes performants de surveillance et de recueil des données sont nécessaires. Mais ces données ne permettent pas de prévoir l'évolution d'une situation, particulièrement pour un agent infectieux nouveau ou variable dont on ne connaît pas les caractéristiques (virulence, contagiosité, morbidité/mortalité associée) », Plan national de prévention et de lutte « Pandémie grippale », document d'aide à la préparation et à la décision, 2011.

(2) Rapport d'étape de la Mission indépendante nationale sur l'évaluation de la gestion de la crise Covid-19 et sur l'anticipation des risques pandémiques, 13 octobre 2020.

a été d'autant plus pénalisant que la croissance du nombre des hospitalisations s'est avérée de type exponentiel »⁽³⁾.

Santé publique France (SPF) a vite rencontré des difficultés pour centraliser les données « éparpillées dans le mille-feuille du système de santé français ». SPF « ne disposait pas d'un outil de crise permettant une gestion, au niveau national, d'une pandémie »⁽⁴⁾.

Nous l'avons dit : En février 2020, aucun des indicateurs scrutés quotidiennement un an plus tard n'était encore calculé. Leur suivi était réalisé par les ARS, sans méthodologie harmonisée. « Tout le monde était sous l'eau, ça a été une course permanente. Il y a eu du flou, des corrections, parce que les gens ont documenté comme ils pouvaient l'information »⁽⁵⁾.

SPF publie le 3 mars 2020 un premier bilan quotidien : une infographie résumant les informations disponibles à cette date. Fin mars, SPF commence à produire un volume croissant de données pour le suivi de l'épidémie. « On a travaillé comme des forcenés », se souvient Yann Le Strat, directeur de DATA⁽⁶⁾ à SPF.

De février à avril, les données disponibles, outre leur qualité inégale, concernaient essentiellement les cas graves :

- le nombre de patients hospitalisés et le nombre de patients en réanimation ou en soins intensifs, *via* le système d'information pour le suivi des victimes (« SI-VIC »)⁽⁷⁾ ;
- les données issues des tests (mais dans une période où l'on ne testait que les personnes hospitalisées, ou les symptomatiques sévères) ;
- les données de recours aux urgences et à SOS Médecins (*via* le système de surveillance « SURSAUD »⁽⁸⁾ et le réseau « OSCOUR »⁽⁹⁾) ;
- les déclarations de décès (avec des délais peu adaptés à la gestion d'une crise sanitaire) ;
- les capacités en respirateurs des hôpitaux, en lits disponibles et installés.

« Plus de 97 % des personnes infectées n'avaient alors pas mis le pied à l'hôpital. Pour autant, un seul indicateur hors hospitalier, issu des médecins de ville, était disponible à partir du 17 mars : le réseau Sentinelles, créé en 1984 pour surveiller les maladies virales »⁽¹⁰⁾. Un réseau performant pour surveiller la grippe ou la varicelle, mais sous-dimensionné (800 médecins généralistes, 0,8 % des praticiens) pour détecter les signaux émergents.

Le suivi des décès était alors assuré par l'Insee (avec un décompte journalier de la mortalité par département en centralisant les certificats médicaux) et par l'Inserm, en charge de la statistique nationale des causes médicales de décès (mais qui ne pouvait produire de décompte en temps réel, vu la faible part des certificats de décès électroniques remplis par les médecins). SPF s'est tournée rapidement vers une troisième source de données, le « SI-VIC », mais qui ne couvrait que les seuls décès hospitaliers.

(3) Rapport d'étape de la Mission indépendante nationale sur l'évaluation de la gestion de la crise Covid-19 et sur l'anticipation des risques pandémiques, 13 octobre 2020.

(4) « Covid-19 : les données incomplètes de SPF compliquent le suivi de l'épidémie », *Le Monde*, 8 octobre 2020.

(5) « Urgence, cafouillages, "mille-feuille" d'indicateurs... Dans les coulisses des données très stratégiques sur le Covid-19 », FranceTVinfo, 10 septembre 2020.

(6) Direction appui, traitements et analyses de données (DATA).

(7) Mis en place à la suite des attentats terroristes de Paris de novembre 2015, le « SI-VIC » récupère le nombre de décès par Covid-19 transmis chaque jour par chaque hôpital. Il laisse de côté les décès survenus à domicile ou en maison de retraite (Ehpad).

(8) Surveillance sanitaire des urgences et des décès.

(9) Organisation de la surveillance coordonnée des urgences, mise en place à la suite de la canicule de 2003.

(10) BARNEOUD L. (2020), « Données épidémiologiques : la pénurie cachée », *Mediapart*, 11 juillet.

En l'absence d'un système centralisé de remontée des résultats des tests, l'indicateur de la disponibilité des lits de réanimation va ainsi rester (jusqu'en mai 2020) la principale boussole des autorités sanitaires.

Une consolidation tardive des indicateurs de tension hospitalière

Les commissions d'enquête du Sénat et de l'Assemblée nationale ont consacré une partie de leurs travaux aux « circuits d'information » qui avaient permis de gérer l'afflux de nouveaux patients dans les semaines critiques de mars-avril 2020, avec une double préoccupation : la remontée d'information vers les cellules nationales de crise et la circulation des données entre acteurs locaux (hôpitaux, ARS, SPF, préfets et collectivités).

Les sénateurs dressent le constat « d'importantes lacunes qui ont fortement diminué la qualité du pilotage *bottom-up* par le ministère de la Santé ». Se penchant de manière détaillée sur les systèmes d'information mobilisés par les établissements de santé (répertoire opérationnel de ressources ⁽¹¹⁾ et « SI-VIC » ⁽¹²⁾, notamment), ils déplorent leur hétérogénéité, à l'origine d'une « consolidation tardive des informations sur le capacitaire » ⁽¹³⁾ et « d'une incapacité notable à fournir des données de qualité en temps réel ».

Les difficultés rencontrées dans le traitement des alertes du local vers le national trouvent leur source, selon les sénateurs, dans la faiblesse des effectifs des 17 antennes régionales de Santé publique France, les CIRE ⁽¹⁴⁾, « totalement sous-dimensionnées pour faire face à des situations sanitaires exceptionnelles ». Des difficultés aggravées par un déficit de « coordination entre les circuits d'informations traitées par les CIRE et par les ARS ». Si des cellules régionales d'appui et de pilotage sanitaire (Craps) furent activées, ce dispositif « ne garantit pas, faute d'harmonisation des circuits de remontées d'informations épidémiologiques entre les CIRE et les ARS, une veille sanitaire suffisamment fine au niveau territorial ». Les sénateurs plaident, en conclusion, pour une « unicité de commandement sanitaire en région ».

L'imparfait décompte des victimes du Covid-19

La France fait partie des rares pays où le nombre de morts officiels liés au Covid-19 est supérieur à l'excès de mortalité sur la période. Ce qui laisse penser que le décompte français a été plus complet que dans d'autres pays ⁽¹⁵⁾.

Au début de l'épidémie, des voix se sont fait entendre pour s'inquiéter du non-décompte des morts dans les Ehpad et, au-delà, d'une sous-estimation du nombre de morts liées au Covid-19, seules celles survenues à l'hôpital étant alors recensées ⁽¹⁶⁾.

Le système statistique des décès, avec ses deux piliers, l'Insee (qui enregistre les décès, mais ignore les causes) et l'Inserm (qui centralise les causes de décès, mais avec un délai très long, seuls 20 %

(11) S'agissant du répertoire opérationnel de ressources (ROR), ils observent que « la matrice des ROR diffère selon les régions, ce qui interdit *a priori* toute consolidation nationale pertinente et empêche leur interopérabilité ».

(12) S'agissant du SI-VIC, le système d'identification et de suivi des victimes d'attentats et de situations sanitaires exceptionnelles, ils observent qu'il ne permet pas de distinguer les places de réanimation et d'USC (unités de soins critiques).

(13) « L'hétérogénéité des systèmes d'information au sein des établissements est à l'origine de plusieurs difficultés, au rang desquelles figurent notamment la consolidation tardive des informations sur capacitaire installé par secteur et la quantification des déprogrammations ».

(14) CIRE : cellules interrégionales d'épidémiologie.

(15) CheckNews (2020), « La France a-t-elle mieux compté les morts du Covid-19 que les autres pays ? », *Libération*, 18 août.

(16) « Dans le champ de la mortalité, la sous-information est totale, pas seulement en France », *Le Monde*, 28 mars 2020.

des certificats de décès lui parvenant en mode numérique) était incapable d'assurer le suivi en temps réel de la mortalité liée au Covid-19. D'où le recours au « SI-VIC » (dont ce n'était pas la vocation⁽¹⁷⁾), devenu la principale source du décompte journalier, communiqué chaque jour par les autorités de santé. Le décompte du « SI-VIC » reste incomplet, puisque seuls les décès survenus à l'hôpital sont décomptés, et non ceux survenus à domicile ou en maison de retraite.

La mise en place accélérée par SPF d'un système de signalement, « Vozanoo », à partir du 2 avril, a permis d'inclure les chiffres des Ehpad dans le décompte des décès quotidiens, mais avec des informations moins détaillées que celles des personnes décédées⁽¹⁸⁾ en dehors de ces établissements.

Depuis le 27 mars 2020, l'Insee (destinataire des bulletins de décès qu'elle reçoit pour 90 % d'entre eux par voie électronique) publie régulièrement un décompte qui retrace l'évolution de la mortalité « toutes causes confondues » et « rend visible » l'excès de mortalité. Mais sans permettre d'estimer avec précision les morts dues au Covid-19.

Cette inadaptation du système statistique à suivre en temps réel une épidémie a conduit les sénateurs à recommander « une modernisation en profondeur du système de déclaration des cas et des décès ». La numérisation de la chaîne de transmission des bulletins de décès état-civil des mairies-ARS-Inserm pourrait épargner à l'Inserm d'avoir à coder chaque année plusieurs centaines de milliers de certificats de décès (ce qui lui prend deux à trois ans pour une année).

Reste un problème plus fondamental qui résulte de la partition du système statistique des décès en deux filières : l'une administrative, autour de l'Insee (qui reçoit et centralise le volet administratif du bulletin de décès) et l'autre sanitaire, avec l'Inserm (qui reçoit et centralise le volet médical du bulletin de décès). On sait grâce à l'Insee qui meurt (âge, sexe, etc.) et par l'Inserm de quoi on meurt (les causes de décès) : sans appariement de ces deux systèmes, on ignore « qui meurt de quoi »⁽¹⁹⁾.

L'ouverture des données a favorisé une effervescence d'initiatives

« Nous ne savons pas nous servir d'un stéthoscope, mais nous pouvons peut-être aider ceux qui savent en leur apportant de la *data* facilement exploitable et de la *dataviz* », Lior Perez, ingénieur chez Météo France.

Les autorités de santé s'étaient assez largement tenues à l'écart du mouvement d'ouverture des données publiques.

Début mars 2020, les seules données disponibles sur l'épidémie étaient éparpillées dans les communiqués de presse des ARS et de Santé publique France. Éclatées et fragmentaires, ces données ne permettaient pas de tracer des courbes et de suivre l'évolution de l'épidémie. La publication de données n'était pas non plus une priorité pour la direction générale de la santé (DGS), concentrée sur la gestion de crise.

(17) Le « SI-VIC » avait initialement été déployé pour répertorier les victimes des attentats de Paris de novembre 2015, dans une perspective nominative et administrative, avec pour objectif principal l'identification des victimes, parfois critique en situation d'attentats, sans description clinique détaillée. Il n'a pas été conçu pour être utilisé dans la durée. « Il a cependant dû être utilisé par les établissements de santé, malgré ses insuffisances, liées notamment à l'absence d'interfaçage avec leur propre système d'information, ce qui a donc nécessité une saisie manuelle des données par les équipes de soins, processus chronophage et source d'erreurs », Cour des comptes (2021), « Les établissements de santé face à la première vague de Covid-19 : exemples néo-aquitains et franc-comtois », mars.

(18) La Commission d'enquête de l'Assemblée nationale juge « regrettable que la comptabilisation des décès dans les Ehpad n'ait pas fait partie des premières données sur l'épidémie : on peut logiquement penser que cela a contribué au retard de prise de conscience de la gravité de la situation dans ces établissements ».

(19) PISON G. & MESLÉ F. (2020), « La statistique publique face à l'urgence du décompte des morts », *The conversation*, 15 avril.

C'est un groupe de citoyens qui va s'en charger. En quelques jours, un collectif (l'initiative « OpenCOVID19 ») puis une communauté de 200 *datascientists* s'organisent pour extraire manuellement les données issues de 20 ARS et de 100 préfectures, et les intégrer dans un tableau. Une charge très lourde : tout est à refaire chaque jour⁽²⁰⁾. Le tableau de bord « veille-coronavirus.fr » prend forme et propose une première vision consolidée des données officielles. La mise en œuvre de cet outil contributif reçoit rapidement l'appui de développeurs d'Etalab, puis le soutien du ministère chargé de la Santé. Consécration : le 28 mars 2020, Édouard Philippe, alors Premier ministre, et Olivier Véran, ministre des Solidarités et de la Santé, en conférence de presse, s'appuient sur les infographies de « veille-coronavirus.fr ».

En avril 2020, 177 jeux de données relatifs au Covid-19 étaient en ligne sur « data.gouv.fr ».

Depuis le 18 mars 2020, Santé publique France publie chaque jour des jeux de données actualisées sur la plateforme de données publiques « data.gouv.fr » et sur « Geodes », l'observatoire cartographique de Santé publique France.

Cette ouverture des données a nourri les travaux des chercheurs et ouvert la voie à une myriade de visualisations dans la presse. Les datavisualisations de « gouvernement.fr » et des tableaux de bord comme « CovidTracker », « coronaboard.fr », « covinfo.fr » ou « vaccinator.fr » s'appuient sur les données de « Geodes », « CovidTracker », suivi par 40 000 abonnés sur Twitter, totalise à lui seul 15 millions de pages vues par mois.

Début mars 2021, SPF proposait sur son site « Geodes » 123 indicateurs, à toutes les échelles (nationale, régionale, départementale, territoriale).

Un double cahier des charges épidémiologique pour la sortie du confinement

Dans le Plan de préparation de la sortie du confinement (dont il était alors le coordinateur), Jean Castex dresse, le 27 avril 2020, la liste des indicateurs qu'il juge nécessaires « pour adapter les mesures de maîtrise de la transmission à la sortie du [premier] confinement⁽²¹⁾ ».

Le Conseil scientifique, pour sa part, avait recensé, le 20 avril 2020, les critères épidémiologiques « pour que la levée du confinement puisse se faire dans de bonnes conditions »⁽²²⁾. Il dessinait, à cette occasion, les grandes lignes d'un (nouveau) système de surveillance épidémiologique « capable de détecter les nouveaux cas et une reprise de l'épidémie », et notamment « d'identifier les lieux à risque de transmission, voire d'épidémie ».

Entre avril et mai 2020, la délégation au numérique en santé (DNS) et SPF vont s'atteler à la mise en œuvre de ce double cahier des charges.

(20) « Coronavirus : en France, l'administration a été forcée à une numérisation accélérée pour gérer l'épidémie », *Le Monde*, 15 avril 2020.

(21) « Ils devront refléter, de la manière la plus précoce possible, la survenue des nouveaux cas d'infection et leur évolution spatio-temporelle pour permettre d'évaluer et ajuster au mieux l'offre de dépistage de ces cas, d'organiser les isolements, la détection des contacts et la mise en quatorzaine de ces derniers au plus près du territoire où ces personnes vivent ; de prévoir les besoins de prises en charge hospitalière en particulier en réanimation », Plan de préparation de la sortie du confinement, Jean Castex, 27 avril et 6 mai 2020.

(22) Un faible nombre journalier d'hospitalisations et d'admissions en réanimation, un nombre de reproduction inférieur à 1 ($R < 1$) et un nombre de lits de réanimation capable de prendre en charge les malades en cas de reprise épidémique, Conseil scientifique Covid-19, Avis du 20 avril 2020.

La campagne de tests et le « SI-DEP » rendent enfin visible la circulation du virus

Jusqu'au 13 mai 2020, les pouvoirs publics n'avaient qu'une vision partielle du nombre de tests réalisés en France, une partie de ceux réalisés en ville, notamment, n'étant pas centralisés.

Tout change avec l'apparition du système d'information de dépistage (« SI-DEP »), le 13 mai 2020. En centralisant la quasi-totalité des tests réalisés dans les laboratoires privés et hospitaliers, « SI-DEP » permet de suivre les opérations de dépistage. « SI-DEP » a été développé en un mois et demi. « En temps normal, cela nous aurait pris plusieurs années », reconnaît Yann Le Strat⁽²³⁾.

Outil central de la stratégie tester-tracer-isoler, le suivi exhaustif de l'ensemble des patients testés permet de construire toute une série d'indicateurs quotidiens, comme les taux de positivité, de dépistage et d'incidence de la maladie.

Rendre visible la prévalence du virus

Le professeur William Dab s'étonnait, le 11 avril 2020 : « Comment se fait-il que ce soient des épidémiologistes britanniques qui aient estimé la proportion de Français infectés ? Comment lutter contre une épidémie sans connaître son étendue ? Des enquêtes par sondages hebdomadaires par téléphone ou Internet permettraient de suivre son évolution. C'est facile à réaliser. Ce n'est pas complètement fiable, mais c'est mieux d'être dans le brouillard que dans le noir absolu. En attendant que des tests sérologiques soient déployés à grande échelle, même avec des imperfections, ce type d'enquête par sondages répétés nous donnerait une tendance sur l'évolution de la prévalence de l'infection⁽²⁴⁾ ».

Le professeur Dab pointait ici deux enjeux distincts pour le pilotage de la crise : connaître la proportion de la population qui a déjà été infectée (et développé des anticorps, décelables par les tests sérologiques) et connaître la proportion de la population infectée (porteuse du virus et donc détectable par les tests de dépistage), à un moment donné ou sur une période donnée (la prévalence).

Connaître la séroprévalence (proportion de la population qui a déjà été infectée, même si certains individus n'ont jamais présenté de symptômes) permet de mieux calculer le taux de mortalité ou le taux de reproduction, de déterminer le chemin qui reste à parcourir avant que la population ait acquis une immunité (au moins partielle et de court terme). À cette fin, deux enquêtes sérologiques nationales ont été conduites en France : EpiCOV⁽²⁵⁾ (avec 12 400 personnes testées) et Sapis⁽²⁶⁾ (16 000). Face à la difficulté de conduire des enquêtes sérologiques régulières, à grande échelle et représentatives, une équipe française d'épidémiologistes⁽²⁷⁾ a développé une méthode d'analyse qui croise les résultats de 22 enquêtes épidémiologiques, les données françaises d'hospitalisation

(23) « Covid-19 : les données incomplètes de SPF compliquent le suivi de l'épidémie », *Le Monde*, 8 octobre 2020.

(24) « En matière de prévention, nous ne sommes pas à la hauteur de l'épidémie », *Le Monde*, 11 avril 2020.

(25) EpiCOV avait pour objectif de fournir une cartographie précise du statut immunitaire de la population, de la santé, des conditions de vie et des inégalités sociales. 135 000 personnes ont répondu à l'enquête. Un kit d'auto-prélèvement sanguin à réaliser à domicile était proposé à une partie des répondants, afin de détecter la présence d'anticorps (sérologie). Au total, 12 400 personnes ont été testées.

(26) Sapis visait à recueillir des données auprès des participants de grandes cohortes épidémiologiques en population générale existantes. Au total, 280 000 adultes au sein de ces cohortes ont ainsi été invités à participer à l'enquête. 42 % d'entre eux ont rempli le premier questionnaire, et 39 % le second. 16 000 d'entre eux avaient été sélectionnés au hasard pour la réalisation de tests, et 14 628 inclus finalement dans l'analyse.

(27) Faute de données sérologiques, cette équipe avait entrepris très tôt de reconstruire l'évolution du nombre total d'infections en France, en croisant les données d'hospitalisations françaises conjointement avec les résultats d'autres enquêtes épidémiologiques.

et de séroprévalence⁽²⁸⁾. C'est cette méthode qui a donné lieu à l'Institut Pasteur d'établir, en février 2021, qu'environ 17 % de la population métropolitaine âgée de plus de 20 ans avait été infectée par le coronavirus depuis un an⁽²⁹⁾.

L'autre enjeu, c'est de connaître la proportion de la population infectée, à un moment donné et dans le temps. On dispose, certes, désormais, de toute une série d'indicateurs, pour étudier la dynamique de la circulation virale au sein du territoire national comme à une plus petite échelle, mais ils restent difficiles à analyser, car le nombre de tests reste soumis à des fluctuations. On peut aussi déduire le taux de prévalence de toute une série de calculs. La meilleure manière de le cerner reste encore de conduire régulièrement des enquêtes auprès d'un échantillon de la population. C'est ce que fait, au Royaume-Uni, le programme « React » de l'Imperial College. Tous les mois, environ 150 000 Britanniques réalisent un auto-prélèvement. Selon ces enquêtes (1,5 million de tests analysés en 9 vagues), le taux de prévalence chute régulièrement. Seule 0,2 % de la population anglaise serait en mars 2021 porteuse du virus, contre 6 % au début de l'été 2020. Les enquêtes « React » ont, en outre, permis d'apporter une réponse à la question qui taraude les scientifiques depuis le début de la pandémie : combien de temps l'organisme reste-t-il immunisé ?

Rendre visibles les *clusters* et les circonstances de contamination

Au début de l'épidémie, les équipes d'investigateurs de SPF et des ARS, très réactives, sont parvenues à circonscrire les premiers foyers de contamination, comme aux Contamines-Montjoie, ou encore dans l'Oise. Elles parviennent, encore, dans un premier temps, à repérer les cas qui se multiplient et remontent les données récoltées *via* l'application « Godata » de l'OMS⁽³⁰⁾. Le passage en phase 3 signe l'arrêt des opérations de suivi des contacts dans la plupart des régions⁽³¹⁾. Conformément au Plan pandémie grippale de 2011 qui n'envisage l'investigation des cas contacts suspects que lors des phases 1 (freiner l'introduction du virus) et 2 (freiner sa propagation).

Cette interruption du suivi des contacts suspects et des foyers ne va pas de soi, car l'identification des *clusters* vise un double objectif : contenir l'épidémie, en premier lieu (ce qui devient impossible, à la mi-mars 2020, faute de moyens humains suffisants, quand le virus circule largement), mais aussi comprendre comment le virus se propage et chemine au sein de la population : le type de lieux propices, le type d'évènements déclencheurs.

D'autres *clusters* auraient-ils pu être circonscrits fin février-début mars ? Selon la Commission d'enquête du Sénat, « il semble qu'à la fin du mois de février la gestion de crise centralisée ait été insuffisamment attentive aux alertes émises du terrain ». Elle ajoute que « l'efficacité des opérations de traçage a été fortement compromise par une absence de communication entre les organismes chargés de la recherche des contacts et de la détection des *clusters*, d'une part, et ceux habilités à prendre les mesures pour les limiter, d'autre part »⁽³²⁾.

(28) « Proportion de la population ayant été infectée par SARS-CoV-2 », Institut Pasteur, 9 avril 2021.

(29) Sans disposer, à l'époque, de données sérologiques, sur la base d'une modélisation, cette équipe avait estimé cette proportion à 5,7 % en juillet 2020 : "Estimating the burden of SARS-CoV-2 in France", *Science*, 10 juillet 2020.

(30) Quelque 11 142 cas confirmés y étaient ainsi recensés au 17 mars 2020.

(31) « Nous n'en avons pas fait assez, car nous manquons de capacités d'action sur le terrain... J'en ai parlé à la directrice générale de Santé publique France. Elle m'a dit : "Nous n'avons pas les moyens. Il y a trop de malades. Nous ne pouvons pas investiguer les sources de contagion" », professeur William Dab, Audition devant la commission d'enquête de l'Assemblée nationale, 23 juin 2020.

(32) Plusieurs acteurs ont regretté devant la Commission d'enquête que les préfètes et les collectivités territoriales, responsables du maintien de l'ordre public et habilitées à limiter les rassemblements, ne soient pas informées des *clusters* lorsque ces derniers sont repérés par les agents de l'assurance maladie ou par l'ARS. Interrogée sur ce point, la CNAM a effectivement rappelé que « sur la transmission aux collectivités territoriales des informations concernant les *clusters*, les textes nous interdisent de passer ce type d'informations... Les rapporteurs ne peuvent que déplorer ce défaut de transmission, qui dépourvoit les opérations de traçage de l'effectivité qu'elles requièrent sur le terrain », Commission d'enquête du Sénat.

Les opérations de traçage des contacts reprendront, à grande échelle, en mai 2020, à la sortie du premier confinement. Le dispositif mis en place par l'assurance maladie, s'il a permis d'identifier des « cas groupés », n'a, en revanche, pas été conçu pour retracer et remonter les chaînes de contamination. « Les agents de l'assurance maladie, lors de leur enquête, ont bien eu connaissance des circonstances dans lesquelles le "patient zéro" a potentiellement contaminé ses contacts. En allant chercher l'information dans "SI-DEP", il serait aussi possible de savoir lesquels de ces contacts ont été infectés. Mais, aucun croisement n'est fait »⁽³³⁾.

Quand les agents de l'assurance maladie identifient, grâce au « SI-DEP », des « cas groupés », ils passent le relais à SPF et aux ARS (le « niveau 3 ») pour des campagnes de dépistage ciblées dans des lieux identifiés comme à risque, ou pour détecter d'éventuels *clusters* dormants. SPF déploie, à cette fin, à partir du 11 mai 2020, un nouvel outil, « MONIC » (MONItorage des Clusters) pour identifier les nouveaux *clusters* en temps réel et suivre ceux en cours d'investigation. Les ARS, chargées du traçage de niveau 3⁽³⁴⁾ adoptaient, pour leur part, « Sormas », un logiciel *open source*, développé en Allemagne en 2014 pour aider le Nigeria à faire face à l'épidémie d'Ebola (et adapté, entre-temps, aux spécificités du Covid-19).

« MONIC » a permis, de fin mai jusqu'à mi-novembre 2020, à SPF de publier dans ses points épidémiologiques hebdomadaires des chiffres sur les *clusters* par type de lieu (Ehpad, établissements pénitentiaires, entreprises, écoles...). Ces données étaient alors qualifiées par SPF de « pertinentes pour identifier les collectivités pour lesquelles la proportion de *clusters* à criticité élevée est la plus importante, contribuant à prioriser les mesures de prévention et de contrôle ». Un thermomètre qui reste lacunaire⁽³⁵⁾. En novembre 2020, SPF a cessé de publier des données détaillées sur les foyers de contagion, faisant valoir que leur nombre était « très fortement sous-estimé »⁽³⁶⁾. Les contaminations identifiées grâce à la détection de ces foyers de contagion ne représentaient alors qu'une très faible part de l'ensemble des personnes déclarées positives sur le territoire.

Rendre visible la campagne vaccinale

Afin d'assurer la traçabilité des vaccins et des étapes de la vaccination, de la consultation pré-vaccinale aux injections du vaccin, l'assurance maladie a ouvert, le 4 janvier 2021, un téléservice baptisé « Vaccin Covid »⁽³⁷⁾. Ce téléservice permet de connaître le nombre quotidien de personnes ayant reçu une ou deux doses de vaccin, par date d'injection, par tranche d'âge et par sexe : des données mobilisées pour modéliser l'évolution de l'épidémie.

Très scrutées, les données quotidiennes permettent, en particulier, d'évaluer le rythme moyen des injections tandis que les données cumulées offrent à voir quelle part de la population est protégée par ce biais, notamment parmi les plus âgés. Les données départementales et régionales peuvent dévoiler d'éventuelles disparités selon les territoires⁽³⁸⁾.

(33) « L'assurance maladie explique ce retard par des contraintes juridiques, liées à la loi Informatique et libertés », extrait de « Covid-19 : les données incomplètes de Santé publique France compliquent le suivi de l'épidémie », *Le Monde*, 8 octobre 2020.

(34) « Situations relevant des chaînes de transmission ou de *cluster* ainsi que les cas ayant eu des contacts multiples lors d'un rassemblement ».

(35) « Taux d'incidence, nombre de personnes vaccinées : les définitions et les limites des grands indicateurs du Covid-19 », *Le Monde*, 13 février 2021

(36) Santé publique France : point épidémiologique du 12 novembre 2020.

(37) « Vaccin Covid » permet, en outre, de créer, d'enregistrer et d'éditer un document après chaque étape. En fin de vaccination, le bilan de la vaccination peut être imprimé, signé et remis au patient auquel il servira de certificat de vaccination.

(38) « Taux d'incidence, nombre de personnes vaccinées : les définitions et les limites des grands indicateurs du Covid-19 », *Le Monde*, 13 février 2021.

La surveillance des variants

L'apparition de nouveaux variants est venue percuter en décembre 2020 la gestion de la crise. Depuis fin janvier 2021, le fichier de suivi « SI-DEP » intègre les résultats des tests RT-PCR « de criblage ». Ces tests détectent les principales mutations caractérisant les variants connus dits « d'intérêt », et la part qu'ils représentent parmi tous les tests positifs réalisés. En avril 2021, tous les tests positifs ne sont pas encore criblés, et des disparités existent entre les territoires. « Ces tests ne permettent de suivre que les variants considérés comme "d'intérêt" par les autorités sanitaires françaises, au travers des enquêtes Flash »⁽³⁹⁾.

La quête de nouvelles sources de données

Deux cas emblématiques illustrent l'intérêt de faire appel à d'autres sources de données pour alimenter et améliorer les modèles épidémiologiques : les données de mobilité telles que mesurées par les opérateurs télécoms et l'analyse des eaux usées.

La collaboration nouée dès mars 2020 entre l'Inserm et Orange pour exploiter les données anonymisées issues de l'activité des abonnés mobiles a permis d'établir qu'un million de personnes avaient quitté la région parisienne entre le 13 et le 20 mars 2020. Une fois connus, ces déplacements rendaient caduques les données usuelles des modèles épidémiologiques⁽⁴⁰⁾. Le partenariat entre l'Inserm et Orange n'est pas isolé. En Allemagne, Deutsche Telekom a transmis les données anonymisées de géolocalisation de ses 46 millions de clients à l'Institut Robert Koch.

Dès avril 2020, plusieurs laboratoires ont lancé le réseau de surveillance Obépine⁽⁴¹⁾, pour suivre la dynamique du virus sur le territoire grâce à l'analyse d'échantillons d'eaux usées provenant de plus de 150 stations d'épuration (parmi les 20 000 françaises), afin d'en tirer des prévisions sur sa circulation. Le 25 janvier 2021, le réseau Obépine a, pour la première fois, mis en ligne les résultats des derniers prélèvements dans une trentaine de stations d'épuration.

Un rendez-vous manqué avec l'épidémiologie collaborative

Au Royaume-Uni, le projet "Covid Symptom Study" proposait au public, *via* une application, de rendre compte, jour par jour, de leurs symptômes, ou de répondre régulièrement à des questionnaires ciblés sur divers sujets. Il comptait 4 millions de contributeurs en mai 2020⁽⁴²⁾.

(39) *Idem*.

(40) « Dans une pandémie de ce type, il y a une forte perturbation de la mobilité : les gens s'adaptent et ne voyagent plus à cause du confinement et de la restriction des déplacements, les trains circulent moins, les vols sont annulés. Il y a de très fortes perturbations de la mobilité : les modèles issus de la vie normale ne sont plus applicables et auraient donné des prédictions erronées. Il est important d'informer nos modèles avec des données qui suivent en temps réel ces changements », explique Victoria Colizza, directrice de recherches à l'Institut Pierre-Louis d'épidémiologie et de santé publique (Inserm-Sorbonne Université), extrait de « Pourquoi les données téléphoniques aident à comprendre la pandémie de Covid-19 », *Le Monde*, 27 mars 2020.

(41) Obépine : Observatoire épidémiologique dans les eaux usées.

(42) "Covid Symptom Study" a été conçu par des médecins et scientifiques du Massachusetts General Hospital, de la Harvard School of Public Health, du King's College London et de la Stanford University School of Medicine, en collaboration avec ZOE, une société britannique spécialisée dans les sciences de la santé.

C'est probablement à ce type de choses que pensait le Conseil scientifique quand il appelait, le 20 octobre 2020⁽⁴³⁾, au lancement de « programmes de science participative pour mieux comprendre les modalités de transmission ».

SPF s'appuie, certes, sur la plateforme « GrippeNet.fr », rebaptisée le 30 mars 2020 « CovidNet.fr », pour recueillir des données auprès du public et suivre en temps réel l'évolution de l'épidémie⁽⁴⁴⁾. Bénéficiant d'une visibilité, « CovidNet.fr » ne compte pourtant que 9 335 participants. L'AP-HP, pour sa part, a fait appel à une communauté de patients pour la recherche (dans le cadre du programme « ComPaRe⁽⁴⁵⁾ ») pour évaluer la perception du risque d'une infection sévère ou pour identifier les symptômes du « Covid long » et leur effet sur la vie des patients.

Mentionnons, ici, cette étude participative, lancée en avril 2020, dans le Haut-Rhin pour comprendre quand et comment ce département s'était transformé en foyer épidémique⁽⁴⁶⁾. Les enquêteurs avaient appelé les Haut-Rhinois à répondre à une centaine de questions. Leur étude a confirmé l'hypothèse que le virus circulait dès fin janvier 2020 dans le département.

On a peut-être sous-estimé, en France, l'intérêt d'associer et de faire participer le public à la production de connaissances épidémiologiques, pour la compréhension des symptômes ou celles des mécanismes de circulation.

Le projet « SourceCovid.fr », initié par l'équipe à l'origine de « MaladieCoronavirus.fr », s'inscrit pleinement dans la démarche participative, préconisée par le Conseil scientifique. Afin de mieux identifier les sources de propagation du virus (magasins, restaurants, lieux de travail, cercles familiaux et personnels...), cette application *web* propose, depuis le 14 décembre 2020, aux personnes ayant été touchées par le virus ou à leurs proches « de prendre deux minutes de leur temps pour aider à la compréhension de la deuxième vague de la Covid-19 ».

Conclusion

« Aujourd'hui, on a quelque chose de quasi exhaustif. Personne en France n'aurait pensé qu'on puisse créer un système national comme celui-là en si peu de temps », Yann Le Strat, en septembre 2020⁽⁴⁷⁾.

Entre mai et juin 2020, brique par brique, un véritable système de surveillance épidémiologique a finalement été mis sur pied.

Ce système de surveillance génère chaque jour des flots de données. SPF reçoit et traite chaque jour 400 mégaoctets, qui alimentent les modèles et actualisent les 123 indicateurs. Début mars 2021, « Geodes » proposait 123 indicateurs, à toutes les échelles (nationale, régionale, départementale, territoriale).

(43) « Les modalités de transmission du virus restent encore mal connues, faute de données précises. Dans une démarche de science participative, garantissant le consentement et l'anonymat des personnes, et dans le cadre de projets de recherche opérationnelle, chacun pourrait contribuer grâce au numérique à faire mieux comprendre les modalités de transmission, en fonction des lieux ou des temps d'exposition par exemple », « Un nouvel ensemble numérique pour lutter contre le SARS-CoV-2 », Avis du Conseil scientifique Covid-19, 20 octobre 2020.

(44) « CovidNet.fr » propose aux volontaires de répondre chaque semaine à un questionnaire de 5 minutes sur d'éventuels symptômes constatés depuis leur précédente connexion. La participation est anonyme.

(45) « ComPaRe » (communauté de patients pour la recherche) rassemble les patients ayant choisi de faire avancer la recherche sur les maladies chroniques.

(46) L'objectif de l'enquête conduite par deux équipes du CNRS et de l'Université Clermont-Auvergne était de savoir si le virus évoluait déjà avant le fameux rassemblement protestant à Mulhouse en février 2020, et de comprendre quel rôle cet événement avait pu jouer.

(47) « Urgence, cafoillages, "mille-feuille" d'indicateurs... Dans les coulisses des données très stratégiques sur le Covid-19 », FranceTVinfo, 10 septembre 2020.

Même si des zones d'ombre subsistent, même si des *bugs* surviennent fréquemment (nombre de cas surestimé, écarts entre les chiffres publiés par SPF et par la DGS, doublons liés aux erreurs d'enregistrement... vite détectés par les réutilisateurs des données en *open data*), ces données et ces indicateurs permettent désormais, tant bien que mal, aux responsables nationaux et locaux de prendre des décisions informées. Ils ont apporté une réponse à la demande du public, de la communauté scientifique, des médias, et des professionnels de santé. Et aidé à nourrir le débat public, avec son lot de controverses sur la fiabilité des données et la pertinence des indicateurs.

Références

CheckNews, *Libération*.

INSEE, Santé publique France & ministère des Solidarités et de la Santé (2020), « Le suivi de l'épidémie de Covid-19, Livret de présentation », mars.

Les décodeurs, *Le Monde*.

Linc-CNIL (2020), « Coronoptiques : dispositifs de surveillance et gestion de l'épidémie ».

Santé publique France (2020), « Surveillance épidémiologique du Covid-19 », dossier pédagogique, avril.

Santé publique France (2020), « Surveillance épidémiologique de la Covid-19 : un dispositif au plus près des territoires », juin.