

Modélisation mathématique de l'épidémie de Covid-19 en France, et comparaison internationale

Par Samuel ALIZON

Directeur de recherche au CNRS et directeur de l'équipe
Écologie et Évolution de la santé au Centre de Recherche
interdisciplinaires en Biologie (CIRB) du Collège de France

Et Mircea T. SOFONEA

Maître de conférences en épidémiologie & évolution
des maladies infectieuses à l'Université de Montpellier,
chercheur au laboratoire PCCEI – U. Montpellier,
Inserm, EFS, U. Antilles

La pandémie de Sars-CoV-2 a mis sur le devant de la scène une discipline scientifique jusque-là discrète : la modélisation mathématique des maladies infectieuses. Pratique divinatoire pour certains, application triviale pour d'autres : elle est avant tout mal connue en France. Après avoir expliqué les fondements de ce champ et les différentes approches de modélisations, nous revenons sur le rôle de la modélisation mathématique au fil de l'épidémie en France. Afin de mettre en avant les variations de ce rôle, nous comparons les différentes périodes avec des politiques marquées mises en place dans d'autres pays. En conclusion, nous donnons des pistes pour une utilisation raisonnée de la modélisation en épidémiologie dans la prise de décision publique.

INTRODUCTION

Fin février 2020, le nombre quotidien de cas de Covid-19 rapportés dans toute l'Europe croît plus que linéairement (Roser *et al.*, 2020). Pour illustrer ce phénomène formalisé plus d'un siècle plus tôt dans la littérature scientifique, des modélisations se retrouvent propulsées devant une audience peu familière de l'exercice.

En France, c'est Arnaud Fontanet qui, rapporte-t-on (Chemin et Etchegoin, 2021), aurait présenté la première esquisse de modèle – une croissance exponentielle – au démarrage des travaux du Conseil scientifique. Didier Raoult en faisait encore parti à ce moment-là et aurait déclaré que « tout ça ne sont que des modèles », devant un parterre qui ne le contredit pas.

Ce texte a pour vocation de préciser quelle est cette approche utilisée par Arnaud Fontanet, pourquoi elle requiert une expertise spécifique, et comment elle a été utilisée depuis deux ans en France et dans le monde.

Encadré 1. Quels types de modélisation en épidémiologie

Trois grands types de modélisations sont utilisés en épidémiologie :

La modélisation mathématique consiste en une formalisation du système d'intérêt à l'aide d'outils mathématiques (équations différentielles ou équivalents). Elle permet d'explorer des scénarios prospectifs, ce qui explique sa popularité auprès du grand public qui les confond parfois avec des prédictions. Elle permet d'analyser des scénarios rétrospectifs et d'explorer des scénarios contrefactuels, ce qui constitue souvent le seul moyen de quantifier l'effet d'une politique de santé publique.

La modélisation statistique est moins connue mais probablement bien plus utilisée. Elle permet d'estimer les paramètres d'un modèle à partir de données. Fondée classiquement sur des séries temporelles d'incidence, elle implique de plus en plus des données originales (réseaux sociaux, génomique par exemple). Avec le déluge de données dans notre vie quotidienne, elle s'intéresse aussi à l'apprentissage profond pour extrapoler une série temporelle à partir des variations passées. Les deux désavantages majeurs de ces nouvelles approches sont la nécessité d'importants jeux de données (inexistants pour les maladies nouvellement émergentes ou négligées) et leur nature phénoménologique (empêchant toute interprétation biologique d'un écart entre la prédiction et la réalité).

Un dernier type de modélisation est **la modélisation informatique**, essentiellement orientée agents. Elle permet d'inclure pléthore de détails dans les simulations et ainsi de lever les hypothèses fortes des modèles mathématiques (comme l'absence de structuration spatiale). Ses points faibles relèvent des contraintes computationnelles et de la difficulté à réaliser des analyses de sensibilité pour explorer l'effet des différentes hypothèses (problèmes d'identifiabilité). On trouve ainsi des simulations inutiles impliquant des centaines de paramètres dont la majeure partie sont fixés au jugé. À l'inverse, le simulateur EPIDEMAP (Thomine *et al.*, 2021) parvient à simuler des épidémies réalistes en n'utilisant que six paramètres, et tout en exploitant directement le cadastre français.

Pour plus de détails sur cette typologie, voir par exemple Djidjou-Demasse, Selinger & Sofonea (2020).

LA MODÉLISATION MATHÉMATIQUE EN ÉPIDÉMIOLOGIE EST UN MÉTIER

L'approche utilisée par le Professeur Fontanet, qui le fait aujourd'hui paraître visionnaire par rapport aux autres membres du Conseil, a pour nom la modélisation mathématique. Il s'agit d'une catégorie de modélisation parmi d'autres (*cf.* l'Encadré 1 ci-dessus), particulièrement adéquate pour appréhender le futur à l'aide de ce que l'on qualifie de « scénarios ». Contrairement aux méthodes produisant des « prévisions », celle-ci repose explicitement sur des hypothèses fortes (voir la Figure 1 en haut ci-contre).

La modélisation n'est évidemment pas l'apanage des épidémiologistes. Toutefois, le processus de contagion, tel que le met en avant dans son ouvrage Adam Kucharski, relève largement des maladies infectieuses. Ainsi, les alternances entre phases de croissance (quasi) exponentielles, saturation, pic et décroissance (quasi) exponentielles, si elles peuvent s'observer pour les effets de mode ou les tendances sur les réseaux sociaux, font partie des enseignements que retiennent les médecins (souvent sous le terme pseudo-technique de « courbe en cloche »).

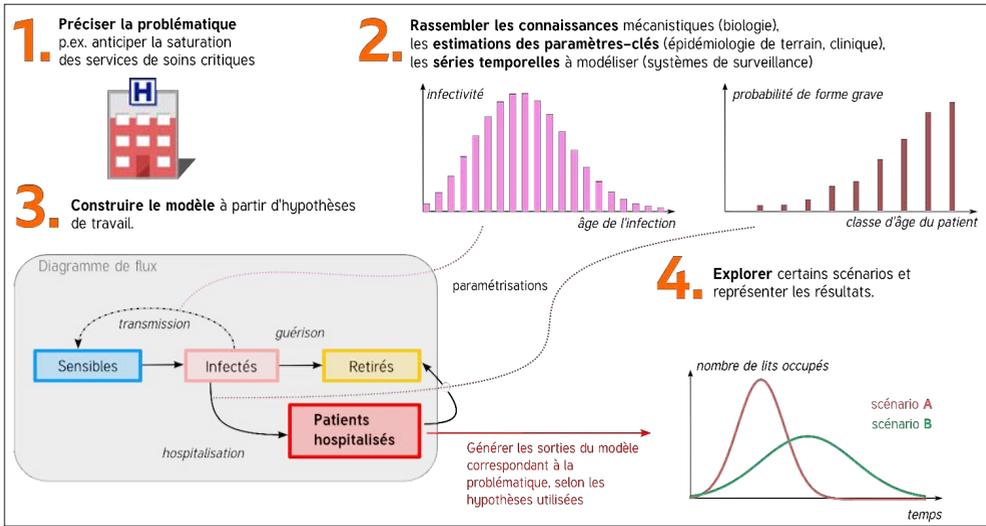


Figure1. Protocole simplifié de la modélisation mathématique d'une maladie infectieuse. D'après Bastien Reyné (Reyné, Saby & Sofonea, 2022), avec sa permission.

Conséquence du confinement forcé ou de la popularité soudaine de notre champ, bien des passionnés sont venus donner des leçons aux équipes de recherche, non sans condescendance. Montrer que le nombre journalier de décès hospitaliers en France pouvait s'approcher par une croissance exponentielle était assez simple (voir la Figure 2 ci-dessous). De manière plus générale, on ne comptait pas les modélisations capturant de manière précise une seule série temporelle d'incidence (nombre de nouveaux cas), que ce soit les dépistages, les hospitalisations ou les décès. Pour modéliser la prévalence (proportion de personnes dans la population présentant une infection active) ou anticiper le nombre de patients en soins critiques, on trouvait déjà moins de monde (notre plateforme COVID ici a été une des seules à le faire en continu). Quant à combiner deux séries temporelles ou plus, il ne restait que quelques équipes expertes.

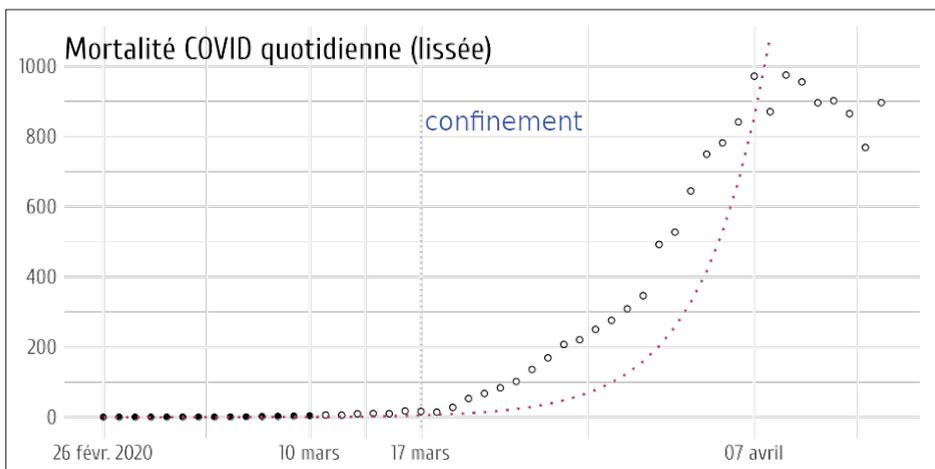


Figure 2. Extrapolation exponentielle (pointillés roses) de la borne inférieure de la mortalité hospitalière Covid quotidienne nationale estimée à partir d'une régression sur les données disponibles au 10 mars 2020 (disques noirs). Données SI-VIC, graphique Mircea Sofonea et Samuel Alizon.

DE L'IMPORTANCE DE LA « MÉMOIRE » DES PROCESSUS

Les difficultés viennent de ce que l'épidémiologie ne se limite pas au processus de contagion. Elle inclut aussi une dimension biologique. Ainsi, la contagiosité d'une personne varie au cours de son infection. De plus, les symptômes ne se déclenchent pas tout de suite : il faut attendre environ deux semaines pour observer les formes sévères. Inclure ces processus nécessite, d'une part, des études de terrain (études que la France n'a que peu, voire pas, financé), d'autre part, des modèles mathématiques adéquats. Plus techniquement, les modèles classiques d'équations aux dérivées ordinaires (EDO) font l'hypothèse que les processus n'ont pas de mémoire (propriété dite de Markov). Selon cette hypothèse, une personne infectée a la même probabilité de guérir si elle est infectée depuis une journée ou depuis une semaine. Si sur le temps long ces effets s'estompent, pour effectuer des scénarios à court ou moyen terme, cet irréalisme explique la difficulté à s'ajuster à plusieurs séries temporelles dépendantes (voir la Figure 3 ci-dessous).

Notre équipe a utilisé trois méthodes non-markoviennes. La plus précise utilise des équations aux dérivées partielles. La probabilité de quitter chaque état est alors régie par une fonction qui lui est propre et qui dépend du nombre de jours passés dans cet état. Autrement dit, au lieu de suivre les individus infectés à une date t , de nombre $I(t)$, on suit les individus infectés depuis a jours à une date t , de nombre $I(a, t)$. Le système s'écrit selon deux équations de flux (1a et 1b) et une condition aux bords (l'équation 1c) :

$$\frac{dS(t)}{dt} = -S(t) \int \beta(a) I(a, t) da \quad (1a)$$

$$\frac{\partial I(a, t)}{\partial a} = -\gamma(a) I(a, t) \quad (1b)$$

$$I(0, t) = S(t) \int \beta(a) I(a, t) da \quad (1c)$$

où $\beta(a)$ et $\gamma(a)$ sont respectivement les taux de transmission et de guérison d'une personne dont l'infection a démarré a unités de temps auparavant. Ceci est particulièrement important pour le second, car la probabilité de guérir n'est pas la même un jour et une semaine après infection. Dans le modèle susceptible-infecté-retiré (SIR) le plus courant, ces paramètres sont constants au cours de l'infection, et le système 1 peut alors s'écrire comme un système EDO classique.

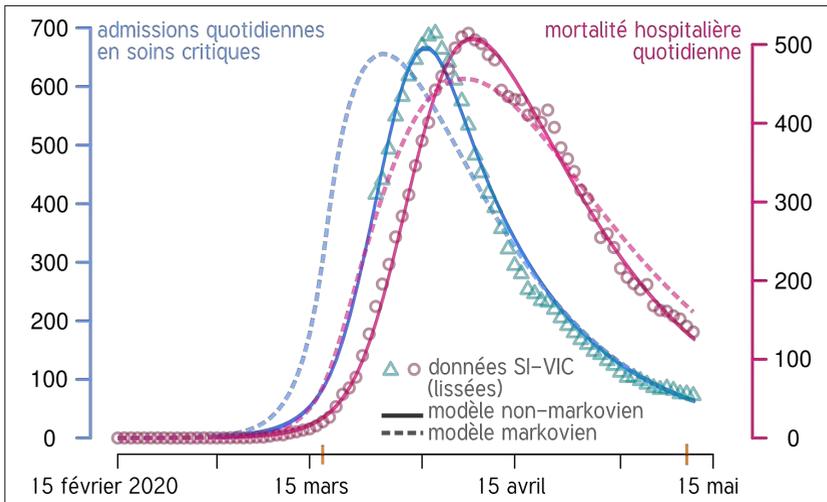


Figure 3. Meilleurs ajustements joints des données SI-VIC quotidiennes lissées (admissions en soins critiques et mortalité hospitalière) selon un modèle non-markovien (traits pleins) et son équivalent markovien (pointillés), sur les données disponibles au 18 mai 2020. D'après Sofonea, Reyné *et al.* (2021).

Ce formalisme ne permet pas d'avoir accès à toute la palette d'outils analytiques offerts par les EDO. Plus problématique, l'implémentation numérique est assez lourde et les temps de calculs importants. Nos modélisations de routine impliquaient donc des modèles en temps discret, qui permettent de modéliser le nombre d'individus infectés depuis a jours $I(a,t)$, mais avec une vitesse de calcul bien supérieure. Cerise sur le gâteau, l'ajustement avec les données empirique est évident puisque les données sont agrégées par jour et non instantanées. En revanche, les analyses statistiques pour estimer les paramètres du modèle à partir des données ont nécessité des développements délicats.

Enfin, une dernière méthode, utilisée par exemple par nos collègues de l'équipe de Simon Cauchemez (Institut Pasteur), consiste à discrétiser un compartiment en plusieurs sous compartiments. Ainsi, si au lieu de passer des états susceptible (S) à infecté (I) à guéri / retiré (R), vous passez successivement de S à I_1 , puis I_2 , puis I_3 , puis I_4 , puis à R le temps de résidence dans l'état I ne suivra plus une distribution exponentielle (comme dans le modèle SIR à EDO classique), mais une loi Gamma de même moyenne que le processus markovien.

EXTENSIONS SPATIALES

Impossible de détailler ici tous les raffinements de la modélisation. L'un des plus débattus concerne la structure spatiale. Même non-markoviens, les modèles font souvent l'hypothèse simplificatrice que chaque personne est potentiellement en contact avec toutes les autres. Suivant cette hypothèse, dans un modèle SIR classique (markovien), la proportion q^* de la population infectée à la suite du passage d'une vague épidémique incontrôlée de paramètre R_0 (le nombre d'infections secondaires engendrées par une personne infectée dans une population entièrement susceptible) est donnée par la formule $q^* R_0 + \log(1 - q^*) = 0$. Pour $R_0 = 2,5$, q^* avoisine les 90 %. À l'aide de simulations individuelles centrées à haute résolution, basées sur le cadastre français, nous montrons que l'approximation de mélange homogène fonctionne bien dans les communes les plus denses (voir la Figure 4 ci-dessous). En revanche, lorsque la densité est plus hétérogène, d'autres types d'approches sont nécessaires. Précision importante, de telles simulations orientées agents ne sont pas incontournables : des structures de type méta-population, où différentes populations sont connectées par un graphe, sont aussi envisageables. On peut aussi utiliser des approximations de paires et suivre, en plus des nombres d'individus S, I ou R, les nombre d'arêtes S-S, S-I, S-R, I-I, I-R ou R-R sur un réseau de contact.

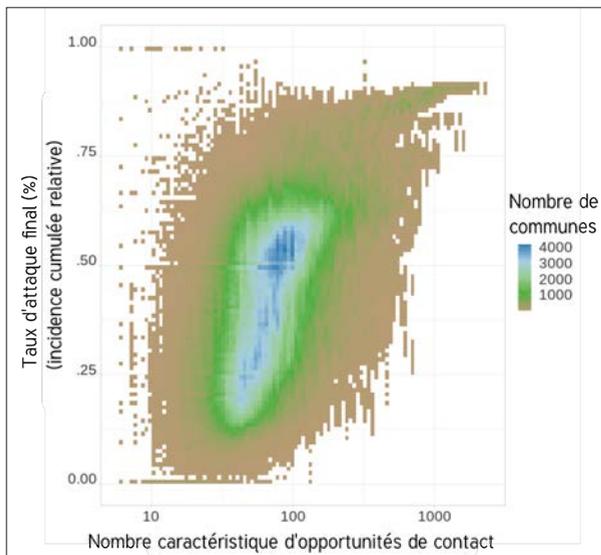


Figure 4. Taux d'attaque final d'une première vague de Covid-19 non contrôlée fictive ($R_0 = 3$), en fonction du nombre de contacts potentiels dans une commune, tel que simulé par EPIDEMAP. Les valeurs les plus élevées du taux d'attaque, tel que calculé par les modèles non spatiaux, sont essentiellement retrouvées dans les communes les plus densément peuplées. D'après Thomine *et al.* (2021).

Outre les effets mémoires liés à l'histoire naturelle de la maladie et aux effets spatiaux induits par la saturation locale des réseaux de contacts, évoqués ci-dessus, d'autres défis se posent à la modélisation. Parmi eux, le changement spontané de comportement est un des plus ardues. Respect des gestes barrières, fréquence des contacts, rappels vaccinaux... Chacun l'a ressenti et les études le montrent : ils sont liés à l'augmentation de cas dans l'entourage, voire à la médiatisation de l'épidémie. La formalisation et la paramétrisation de ces processus sont loin d'être triviales, et nécessitent une expertise et des données en temps réel issues des sciences humaines et sociales. Ceci soulève le problème chronique du manque de financement des études de terrain, en particulier l'absence en France (contrairement au Royaume-Uni) de dépistages réguliers d'échantillons aléatoires de la population générale selon un plan d'échantillonnage statistique bien conçu. Car les relevés d'incidence repris sans filtres par les journalistes sont hyper biaisés : indications, effort ou incitation au dépistage dépendent de l'âge, des symptômes, des cas contacts, des conditions d'accès, ou des fêtes de fin d'année. Au contraire, un suivi aléatoire permet d'estimer la prévalence (donc le nombre de personnes potentiellement contagieuses). Cette donnée cruciale permettrait d'informer en temps réel les autorités comme les modèles, réduisant le nombre d'inconnues et conduisant, éventuellement, à inférer des paramètres plus fins sur le plan spatial, temporel ou interventionnel.

QUELQUES PISTES POUR LA MODÉLISATION EN SANTÉ PUBLIQUE

L'avenir de la modélisation n'échappe pas à la question budgétaire puisque comme le fait remarquer William Dab (*Le Monde*, 11 juillet 2020), l'ancien directeur de la direction générale de la Santé (DGS), en matière de dépenses nationales de santé, chaque fois que nous dépensons 100 euros, 96 euros vont aux soins individuels et 4 euros seulement à la prévention organisée. Le déséquilibre est très fort alors que les pays à forte politique de santé publique ont plutôt un ratio de 90/10. En France, la santé publique a toujours été considérée comme un secteur purement administratif et sans valeur ajoutée. Au-delà des aspects financiers, une agence comme Santé publique France n'emploie personne pour faire de la modélisation mathématique à temps plein. Pour la DGS ou la Haute autorité de santé (HAS), cette expertise, avec celle des biostatistiques, est tout aussi rare. Ceci reflète la nature de la santé publique en France, qui est considérée comme une discipline avant tout médicale. De plus, cette spécialité est peu valorisée et, en pratique, peu sélective au concours de l'internat. Bref, une des urgences est probablement une reconnaissance de la santé publique comme discipline à part, qui inclut, mais aussi dépasse la médecine.

La modélisation mathématique peut présenter de nombreux atouts. Pour n'en citer que trois : un premier concerne l'élaboration de scénarios prospectifs, qui, sans des modèles quantitatifs éprouvés, ne reposent que sur des modèles verbaux ou, pire, d'autorité. Un second concerne la surveillance et le développement de stratégies d'échantillonnage optimisées, maximisant la qualité de l'information avec un minimum de prélèvements. Le troisième concerne les scénarios rétrospectifs. Lorsque qu'une crise sanitaire survient, soit elle est sous-évaluée et de nombreux décès surviennent, soit elle est bien anticipée et contrée, mais les critiques sont tout aussi nombreuses car il n'existe alors aucune manifestation tangible de cette crise. Comme il est impossible de rejouer une crise sanitaire, seule la modélisation permet d'échapper à cette tragédie bien connue de la santé publique à l'aide de scénarios contrefactuels, et ainsi de juger le bienfondé de l'action publique. Enfin, quelle que soit leur utilisation, les modélisations ne sont qu'un élément parmi d'autres dans la prise de décision, mais il est important d'éviter le maximum d'intermédiaires entre les porte-paroles des équipes expertes en modélisation et de celles amenées à prendre les décisions.

Encadré 2. Retours historiques et internationaux

Mars 2020 : le modèle asiatique

Le début d'une épidémie est un processus stochastique. En France, plusieurs chaînes de transmission initiées pendant les mois de janvier et février se sont éteintes et n'ont pas contribué à la première vague (Nadeau *et al.*, 2021). Avec son suivi génomique et des capacités d'analyse bien supérieures aux nôtres, l'Angleterre a pu détecter des centaines de ces introductions. La France, comme la plupart des pays européens, a dû s'en remettre à un contrôle strict de l'épidémie, suivant par-là les modélisations du Pr Fontanet ou celles, un peu plus élaborées, de l'équipe de Neil Ferguson à l'Imperial College de Londres. Ce faisant, elle a suivi ce que l'on pourrait appeler un modèle asiatique consistant à contrôler strictement l'épidémie jusqu'à sa disparition. Toutefois, une différence regrettable a été le délai avant l'instauration des mesures.

Automne 2020 : cafouillis à la tête de l'État

La fin 2020 marque la disgrâce de la modélisation en France pour des raisons qui étaient évitables. À la suite du relâchement quasi complet des mesures de prévention à la rentrée scolaire, la deuxième vague annoncée par le Conseil scientifique à partir de la fin juillet et détectable dès le mois d'août ne peut plus être ignorée. L'équipe de l'Institut Pasteur a réalisé plusieurs scénarios prospectifs sur lesquels s'appuie l'exécutif sans consulter l'équipe en question. Hélas, les cabinets ministériels semblent n'en avoir retenu qu'un scénario pour référence : le scénario sans réaction. On se retrouve ainsi avec un discours du Président de la République annonçant que quoi qu'il arrive, nous aurons 9 000 personnes en réanimation au mois de novembre. N'importe quel modélisateur en épidémiologie aurait pu éviter cette bévue en expliquant qu'avec la mise en place des mesures de santé publique, *a fortiori* un confinement, ce scénario ne se réaliserait pas.

L'hiver 2020-2021 : l'aveuglement à la brésilienne

La deuxième année de la pandémie marque le divorce entre la tête de l'État et le Conseil scientifique. Face aux modélisations indiquant que la contagiosité accrue du variant Alpha est susceptible d'engendrer une nouvelle vague épidémique, le gouvernement se contente de mettre en place un couvre-feu à 18 h 00 en janvier. À la fin du mois, des scénarios de l'équipe de Vittoria Colizza à l'Inserm indiquent que cela ne suffira pas à contenir la vague. S'ensuivent plusieurs semaines qui aboutiront au confinement strict, inéluctable et anticipé du 3 avril. Nos scénarios rétrospectifs montrent que si un confinement de même durée avait été mis en place début février comme le demandait le Conseil scientifique, il y aurait eu 14 000 à 20 000 décès de moins, même en prenant en compte une reprise épidémique avec le déconfinement début mars. Au niveau international, le Brésil est un des pays qui a illustré les tragiques conséquences à laisser circuler l'épidémie dans une population d'âge moyen non vaccinée.

2021 : la couverture vaccinale, un succès à la française

Si les critiques ne manquent pas concernant la gestion de la crise, force est de constater que, malgré les hésitations de 2020 et la gestion catastrophique par les cabinets de conseils, l'implémentation de la vaccination en France en 2021 a été un succès, que la modélisation mathématique permet de quantifier. À l'aide d'un scénario rétrospectif, nous estimons qu'au moins 36 000 vies avaient été sauvées sur les neuf premiers mois de vaccination (Sofonea, Boennec *et al.*, 2021). Notons au passage que ce type de scénario n'est pas si trivial à élaborer, car il faut tenir

compte du fait que face à une vague épidémique de grande ampleur liée à l'absence de vaccination, les pouvoirs publics ne seraient pas restés indifférents. Des collègues ont ainsi tendance à surestimer l'effet de la vaccination dans leurs analyses, ce qui n'est pas sans soulever des problèmes politiques (Watson *et al.*, 2022).

2022 : Omicron, ou ignorer la pandémie à la suédoise

La Suède a fait couler beaucoup d'encre pour son refus de « confiner ». D'abord, notons que les mots recouvrent des réalités différentes selon les pays. Si l'on en croit les données de Google Mobility, les déplacements dans la Suède « déconfinée » étaient assez proches de ceux en Allemagne « confinée » (la baisse de fréquentation des transports publics était de 40 % en Suède contre 55 % en Allemagne ou 80 % en France (Roser *et al.*, 2020)). Le flou de ce terme fait qu'il est désormais peu utilisé tel quel dans la littérature scientifique. D'autre part, on le sait moins mais il existe des analyses violentes accusant l'exécutif suédois d'avoir sciemment menti sur les chiffres (Brusselsaers *et al.*, 2022). En France, heureusement, de tels extrêmes n'ont jamais été atteints et si des graphiques ont été modifiés lors de leur présentation, c'est plus par incompétence que par volonté de dissimulation. Il n'en reste pas moins que depuis 2022, l'épidémie en France semble avoir été reléguée au rang des infections respiratoires saisonnières, avec un effort « à la suédoise » pour ne rien anticiper.

RÉFÉRENCES

- BRUSSELAERS N., STEADSON D., BJORKLUND K., BRELAND S., STILHOFF SÖRENSEN J., EWING A., BERGMANN S. & STEINECK G. (2022), "Evaluation of science advice during the Covid-19 pandemic in Sweden", *Humanit. Soc. Sci. Commun.*, 9(1), <https://doi.org/10.1057/s41599-022-01097-5>
- CHEMIN A. & ETCHEGOIN M.-F. (2021), *Raoult. Une folie française*, Éditions Gallimard.
- DJIDJOU-DEMASSE R., SELINGER C. & SOFONEA M. T. (2020), « Épidémiologie mathématique et modélisation de la pandémie de Covid-19 : enjeux et diversité », *Revue Francophone des Laboratoires*, n°526, pp. 6369, [https://doi.org/10.1016/S1773-035X\(20\)30315-4](https://doi.org/10.1016/S1773-035X(20)30315-4)
- NADEAU S. A., VAUGHAN T. G., SCIRE J., HUISMAN J. S. & STADLER T. (2021), "The origin and early spread of Sars-CoV-2 in Europe", *Proc. Nat. Acad. Sci., USA*, 118(9), p. e2012008118, <https://doi.org/10.1073/pnas.2012008118>
- REYNÉ B., SABY N. & SOFONEA M. T. (2022), "Principles of mathematical epidemiology and compartmental modelling application to Covid-19", *Anaesthesia Critical Care & Pain Medicine*, 41(1), 101017, <https://doi.org/10.1016/j.accpm.2021.101017>
- ROSER M., RITCHIE H., ORTIZ-OSPINA E. & HASELL J. (2020), "Coronavirus pandemic (Covid-19)", *Our world in data*, <https://ourworldindata.org/coronavirus>
- SOFONEA M. T., BOENNEC C., FONTALIRANT A., REGNAT M., LEFRANT J.-Y., ALIZON S. & FISCHER M.-O. (2021), "Quantifying the real-life impacts of vaccination on critical Covid-19", *OSF Preprints*, <https://doi.org/10.31219/osf.io/6ebxu>
- SOFONEA M. T., REYNÉ B., ELIE B., DJIDJOU-DEMASSE R., SELINGER C., MICHALAKIS Y. & ALIZON S. (2021), "Memory is key in capturing Covid-19 epidemiological dynamics", *Epidemics*, n°35, 100459, <https://doi.org/10.1016/j.epidem.2021.100459>

THOMINE O., ALIZON S., BOENNEC C., BARTHELEMY M. & SOFONEA M. (2021), “Emerging dynamics from high-resolution spatial numerical epidemics”, *eLife*, n°10, e71417, <https://doi.org/10.7554/eLife.71417>

WATSON O. J., BARNSLEY G., TOOR J., HOGAN A. B., WINSKILL P. & GHANI A. C. (2022), “Global impact of the first year of Covid-19 vaccination: A mathematical modelling study”, *Lancet Infect. Dis. sous presses*, [https://doi.org/10.1016/S1473-3099\(22\)00320-6](https://doi.org/10.1016/S1473-3099(22)00320-6)