

---

# LES MODÈLES PENDANT LA COVID-19

Mark A. Lewis et Daniel Coombs | 28 mai, 2020

La pandémie de COVID-19 qui sévit dans le monde entier depuis les six derniers mois attire plus que jamais l'attention sur la modélisation mathématique quantitative. Les décideurs politiques et le grand public se tournent vers la science, et la modélisation en particulier, pour comprendre la dynamique complexe de l'épidémie, d'un point de vue local et mondial, ainsi que pour prévoir les conséquences des interventions possibles sur le nombre de cas, d'hospitalisations et de décès. La modélisation joue un rôle essentiel, comme nous le constatons au quotidien dans les interventions de nos médecins en chef, des responsables gouvernementaux et des élus.

Un exemple particulièrement remarquable est le modèle utilisé par Neil Ferguson et ses collègues de l'Imperial College, London, pour conseiller le gouvernement britannique sur les conséquences probables d'une propagation incontrôlée de la maladie [1]. Les mesures politiques en faveur de la distanciation sociale, influencées par les résultats de cette modélisation, ont certainement sauvé des milliers de vies au Royaume-Uni et ont soutenu l'adoption de politiques semblables aux États-Unis. Tout modèle a ses propres limites, et de nombreux groupes de modélisation (en fait la plupart) s'empressent de souligner les lacunes de leur propre travail et de celui des autres. Le très influent modèle de l'Imperial College s'appuyait sur des estimations de paramètres qui étaient, et sont toujours, profondément incertains.

Devrions-nous nous inquiéter ? Cela dépend de ce qui nous préoccupe. Si notre souci est de concevoir des stratégies à court terme pour lutter contre la pandémie la plus dévastatrice de l'histoire récente, alors la réponse est « non ». Le modèle de l'Imperial College, et d'autres modèles similaires, ont amplement démontré la nécessité d'une distanciation sociale et les conséquences catastrophiques du non-respect de cette mesure, une belle illustration du mantra populaire « Tous les modèles sont faux, mais certains modèles sont utiles ». Cependant, si nous nous soucions de faire des prévisions à long terme plus précises afin de pouvoir préparer la relance de l'économie, la réponse est « nous pouvons faire mieux ». Nous pouvons faire mieux parce que nous disposons dorénavant de plus de données, ce qui nous permet de développer de nouveaux types de modèles.

Des modèles mathématiques sur la dynamique des maladies infectieuses existent depuis longtemps. Le mathématicien Daniel Bernoulli a développé un modèle dynamique sur la transmission et le contrôle de la variole en 1760 [2]. Ronald Ross, médecin et prix Nobel de mathématiques, a développé une théorie mathématique sur la dynamique des épidémies de malaria en 1911 [3].

Deux décennies plus tard, W.O. Kermack et A.G. McKendrick collaboraient pour développer une nouvelle théorie de la transmission des maladies infectieuses [4]. Leur modèle scindait la population en trois groupes : les personnes susceptibles (S), les personnes infectées (I) et les personnes immunisées/mortes (R). Cela leur a permis de se concentrer sur les taux de transfert de personnes entre les groupes en fonction de l'évolution biologique de l'infection, faisant passer des individus S en individus I et finalement en groupe R.

---

Ce modèle SIR repose sur une idée très simple : celle de l'infection par « action de masse », qui suppose que chaque rencontre aléatoire entre des individus susceptibles et des individus infectés engendre un certain risque de nouvelle infection. Ils ont utilisé leur modèle pour mieux comprendre l'épidémie de peste bubonique de Bombay en 1906.

Ce style de « modélisation compartimentale » est resté en vigueur et les bases sont enseignées aux épidémiologistes et aux étudiants en mathématiques appliquées dans les classes de premier cycle du monde entier. Bien que l'approche soit simple, elle est parvenue à expliquer de manière remarquable les schémas d'apparition des infections, y compris les épidémies de grippe dans les internats britanniques. Les extensions incluent davantage de catégories d'individus (par exemple, les exposés, ceux en quarantaine ou les asymptomatiques), ainsi que les risques différentiels pour différents groupes (par exemple, ceux qui présentent un risque différentiel de contracter le sida en fonction de leurs relations sexuelles). En effet, la structure de nombreux modèles épidémiques modernes sur la COVID-19 est basée sur des modifications de la structure compartimentale du SEIR, ajoutant la catégorie « Exposé » entre « Susceptible » et « Infecté ».

Ces modèles donnent naissance à un concept clé dans l'étude des épidémies, le taux de reproduction de base ( $R_0$ ), qui est le nombre moyen de nouvelles infections engendrées par un seul individu infecté. Un  $R_0$  supérieur à 1 signifie que la maladie peut croître de manière exponentielle, avec des conséquences potentiellement dévastatrices. Cependant, un  $R_0$  inférieur à 1 signifie que la maladie devrait disparaître. Ainsi, la réduction du  $R_0$  devient un objectif de gestion en soi. Dans le modèle de Kermack-McKendrick, le  $R_0$  est le taux auquel un individu infecté infecte d'autres personnes, multiplié par la période d'infectiosité. La distanciation sociale peut réduire le premier facteur, tandis que la quarantaine raccourcit le second. Les deux stratégies de contrôle sont donc des méthodes puissantes pour réduire le  $R_0$  et ainsi faire baisser les niveaux de maladie. Cependant, la valeur du  $R_0$  peut changer quotidiennement en fonction de l'évolution des politiques sociales. Ainsi, lorsque le nombre de cas quotidiens est disponible, il est possible de calculer les valeurs quotidiennes de  $R_0$  qui permettent de rapprocher les résultats du modèle aux données observées. Ce processus, appelé calage d'un modèle, est la méthode utilisée pour produire la plupart des estimations de  $R_0$  relatives à la pandémie de COVID-19. En ajustant simultanément les paramètres essentiels aux données disponibles, le modèle devient progressivement un outil de projection de l'évolution future de l'épidémie. Dans le contexte actuel, nous pouvons utiliser cette approche pour faire des projections des taux d'hospitalisation ou d'admission en soins intensifs pendant l'épidémie, et pour adapter les conseils prodigués aux responsables.

On serait tenté de penser que plus un modèle est réaliste, mieux il est. Donc, si nous voulons de meilleurs modèles dynamiques, devrions-nous simplement les rendre plus compliqués ? Se pose ici un dilemme classique de la modélisation. Plus un modèle est réaliste, et donc complexe, plus le résultat est incertain, car le comportement d'un modèle complexe dépend d'une myriade de paramètres d'entrée détaillés, dont beaucoup sont inconnus ou connus de manière approximative. En effet, il faut toujours faire un compromis entre l'exactitude (le réalisme avec lequel le modèle intègre toutes les différentes entrées possibles) et la précision (le niveau de certitude associé aux prédictions du modèle). Le problème rencontré avec la COVID-19 est que certains paramètres, comme le niveau de distanciation sociale, peuvent être très variables d'un endroit à l'autre et d'une semaine à l'autre. D'autres paramètres, tels que

---

---

la proportion de personnes infectées qui ne développent pas de symptômes, peuvent être difficiles à mesurer. Ces problématiques complexifient le calage du modèle. En effet, la connexion des modèles aux données et l'évaluation de l'incertitude sont les parties les plus chronophages et les plus complexes d'un projet de modélisation.

Les modèles peuvent-ils indiquer quels sont les groupes les plus vulnérables ? Où sont les maillons faibles de nos systèmes de santé publique ? L'urgence intrinsèque à ces questions explique pourquoi nous devons développer des méthodes d'évaluation des risques avant que toutes les preuves ne soient réunies. Partout au Canada, de nombreux établissements de santé pour personnes âgées ont été submergés par la COVID-19, avec des conséquences tragiques, tandis que d'autres s'en sont sortis indemnes. En Alberta, les statistiques provinciales relatives au nombre de cas ont été fortement influencées par deux usines de conditionnement de viande et, depuis début mai, ces épidémies locales sont responsables d'un quart de tous les cas de COVID-19 signalés dans la province. Des épidémies similaires dans des usines de conditionnement de viande ont eu lieu en Colombie-Britannique, mais de nombreuses usines voisines ont jusqu'à présent évité les infections.

Les détails de ces événements ne peuvent être prédits avec certitude. Nous pouvons chercher des réponses à partir de modèles beaucoup plus complexes qui visent à simuler l'activité humaine dans un ordinateur. De tels modèles détaillés, structurés géographiquement et socialement, ont déjà été développés pour prédire la propagation de la grippe pandémique aux États-Unis [5] et la propagation du virus de la dengue en Asie du Sud-Est [6]. Rétrospectivement, les modèles mathématiques auraient pu et auraient dû faire beaucoup mieux en ce qui concerne la COVID-19. Pour faire mieux, les modélistes doivent utiliser les connaissances et les données relatives aux lieux, aux tailles et aux conditions de travail dans les espaces confinés tels que les usines, les prisons et les établissements de santé. Cette conception est une rupture par rapport au paradigme compartimental de la modélisation, où ces détails sont souvent négligés au profit d'une moyenne sur l'ensemble d'une population.

Qu'en est-il des « supercontamineurs » ? Un patient de 38 ans a peut-être été le premier cas de COVID-19 en Lombardie, une région d'Italie à l'épicentre de l'épidémie européenne. Les parquetiers italiens ont ouvert une enquête afin de déterminer si un retard dans son traitement aurait pu déclencher un premier pic d'infections. De même, une grande partie des cas déclarés en Corée du Sud pourraient provenir d'une seule église. Des études modélisant l'influence des supercontamineurs de COVID-19 [7], ainsi que d'autres maladies respiratoires telles que le SRAS et la grippe, ont montré l'immense impact de ces événements imprévisibles sur l'évolution de la maladie. Bien que nous puissions prédire les impacts, nous ne sommes pas en mesure de répondre à la question clé qui permettrait le contrôle, à savoir connaître où et quand de tels événements se produiront.

Pour améliorer le pouvoir prédictif des modèles mathématiques, nous avons besoin d'un partage rapide et détaillé des données entre les agences de santé publique et les équipes de modélisation. Les données requises dépendent du problème à étudier. Pour comprendre la vitesse de propagation de la maladie parmi les différentes populations, nous avons besoin d'informations détaillées sur le moment de l'apparition des symptômes, les tests et l'auto-isolement des personnes infectées. Pour mesurer le risque d'erreurs entre les différents groupes d'âge, ces données doivent être analysées par

---

âge. Si nous voulons comprendre l'importance des épidémies dans les établissements, qui se superposent à un contexte de transmission communautaire, nous avons besoin de détails géographiques sur le domicile, le travail et les événements où la transmission est possible. Ces données doivent être approfondies en réunissant les responsables de la santé publique et les équipes de modélisation, afin de mieux comprendre les limites de la collecte de données sur le terrain, mais aussi les possibilités offertes par l'apport rapide de données détaillées.

Toutefois, au-delà du besoin de données hautement qualitatives, les équipes de modélisation mathématique devraient chercher à élargir leurs boîtes à outils pour y inclure des méthodes et des flux d'informations provenant d'autres domaines que l'épidémiologie classique et la santé publique. À titre d'exemple, l'apprentissage automatique et l'intelligence artificielle se sont rapidement développés. Ces techniques fournissent des données et génèrent des prédictions sans faire de suppositions sur les mécanismes ou les hypothèses. Au fur et à mesure de la mise à disposition de grands ensembles de données (par exemple, grâce aux applications de suivi des téléphones intelligents adaptées à l'évaluation des risques de COVID-19), ces méthodes peuvent devenir de puissants outils, si balisés, pour faire des prévisions à court terme. Étant donné que l'objectif de l'apprentissage automatique est simplement de prédire des modèles basés sur des données, les approches qui construisent, par exemple, des outils pour évaluer le risque de crédit pourraient également être appliquées pour déterminer quelles villes ou quelles populations sont les plus à risque.

En second lieu, les économètres et les statisticiens du gouvernement possèdent des quantités impressionnantes de données sur les entreprises et l'activité économique, ainsi que des modèles qui peuvent être mis en parallèle aux prévisions épidémiologiques, lorsque nous visons un assouplissement des restrictions et une réouverture de l'économie. Lorsque les entreprises reprennent leurs activités normales, quels sont les autres secteurs qui sont leurs principaux fournisseurs et leurs clients essentiels ? Que se passe-t-il si ces secteurs opèrent dans une autre province ou un autre pays ? Quels sont les secteurs économiques et les entreprises dont les conditions de travail présentent les plus grands risques de transmission sur le lieu de travail ? Quels secteurs comptent de nombreux employés âgés de plus de 60 ans, plus exposés à des risques de maladie grave ? Dans quelle mesure devons-nous attendre des employés qu'ils respectent les instructions d'éloignement du lieu de travail ou de soumission à un test en cas de symptômes légers ? Les équipes engagées dans la modélisation mathématique de la désescalade doivent être prêtes à travailler avec des économistes et des données statistiques détaillées pour comprendre l'interaction complexe entre l'activité économique et la propagation des maladies.

Les scientifiques du monde entier appellent à utiliser des modèles mathématiques pour comprendre le fonctionnement de la propagation de COVID-19. Les modèles sont utilisés pour prendre des décisions de vie ou de mort, et les Canadiens ont été en première ligne pour mener à bien ces efforts, avec des groupes de chercheurs exceptionnels travaillant sans relâche. Toutes les victoires obtenues grâce aux modèles dans la compréhension de notre situation actuelle doivent être attribuées au travail minutieux de générations d'épidémiologistes mathématiques. Même si un travail intense a été mené pour comprendre la vague initiale d'infection, nous avons encore du beaucoup de travail à accomplir et nous pouvons continuer à faire mieux.

---

## Références

- [1] Ferguson, N., et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. (2020).
- [2] Bernoulli, D. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole. Mém. Math. Phys. Acad. Roy. Sci., Paris, (1766).
- [3] Ross, R. The Prevention of Malaria. Murray, London (1911).
- [4] Kermack, W.O., McKendrick, A.G. A contribution to the mathematical theory of epidemics. Proc. R. Soc. London, 115:700-721 (1927).
- [5] Germann, T.C., Kadau, K., Longini, I.M., Macken, C.A. Mitigation strategies for pandemic influenza in the United States. Proc. Nat. Acad. Sci. USA 103(15):5935-5940 (2006).
- [6] Chao, D.L., Halstead, S.B., Halloran, M.E., Longini, I.M. Controlling dengue with vaccines in Thailand. PLoS Negl. Trop. Dis. 6(10):e1876 (2012).
- [7] Reich, O., Shalev, G. Kalvari, T. Modeling COVID-19 on a network: super-spreaders, testing and containment. medRxiv <https://doi.org/10.1101/2020.04.30.20081828> (2020).