Utilisation des premières données pour estimer le taux de mortalité par infection du COVID-19 en France

Lionel Roques, Etienne E. Klein, Julien Papaïx, Antoine Sar, Samuel Soubeyrand

Utilisation des premières données pour estimer le taux de mortalité par infection du COVID-19 en France

Biology 2020, 9(5), 97

Lionel Roques, Etienne K. Klein, Julien Papaïx, Antoine Sar, Samuel Soubeyrand

Résumé

Le nombre de tests de dépistage réalisés en France et la méthodologie utilisée pour cibler les patients testés ne permettent pas de calculer directement le nombre réel de cas et le taux de mortalité infectieuse (IFR). L'objectif principal de ce travail est d'estimer le nombre réel de personnes infectées par le COVID-19 et de déduire l'IFR lors de la fenêtre d'observation en France. Nous développons une approche «mécaniste-statistique» couplant un modèle épidémiologique SIR décrivant la dynamique épidémiologique non observée, un modèle probabiliste décrivant le processus d'acquisition des données et une méthode d'inférence statistique. Le nombre réel de cas infectés en France est probablement supérieur aux observations: on retrouve ici un facteur × 8 (95% -CI: 5–12) qui conduit à un IFR en France de 0,5% (95% -CI: 0,3– 0.8) sur la base des données de comptage des décès à l'hôpital. En ajustant le nombre de décès dans les maisons de retraite, nous obtenons un IFR de 0,8% (IC 95%: 0,45-1,25). Cet IFR est conforme aux résultats précédents en Chine (0,66%) et au Royaume-Uni (0,9%) et inférieur à la valeur précédemment calculée sur les données du navire Diamond Princess (1,3%).

1. Introduction

L'épidémie de COVID-19 a commencé en décembre 2019 dans la province du Hubei, en Chine. Depuis lors, la maladie s'est propagée dans le monde entier pour atteindre le stade de la pandémie, selon l'OMS [1], le 11 mars. Les premiers cas ont été détectés en France le 24 janvier. Le taux de mortalité par infection (IFR), défini comme le nombre de décès divisé par le nombre de cas infectés, est une quantité importante qui nous renseigne sur le nombre attendu de victimes à la fin d'une épidémie, lorsqu'une proportion donnée de la population a été infecté. Bien que les données sur le nombre de décès dus au COVID-19 soient probablement exactes, le nombre réel de personnes infectées dans la population n'est pas connu. Ainsi, en raison du nombre relativement faible de tests de dépistage qui ont été réalisés en France (environ cinq personnes sur 10000 en France contre 50 sur 10000 en Corée du Sud jusqu'au 15 mars 2020; sources: Santé Publique France et Korean Centre pour Disease Control), le calcul direct de l'IFR n'est pas possible. Sur la base des cas confirmés par PCR chez des résidents internationaux rapatriés de Chine en janvier 2020, Verity et al. [2] ont obtenu une estimation du taux de mortalité par infection (IFR) de 0,66% en Chine et, en ajustant les taux d'attaque non uniformes par âge, un IFR de 0,9% a été obtenu au Royaume-Uni [3]. En utilisant les données du bateau de croisière Diamond Princess mis en quarantaine au Japon et en corrigeant les retards entre la confirmation et le décès, Russel et al. [4] ont obtenu un IFR de 1,3%.

En utilisant les premières données (jusqu'au 17 mars) disponibles en France, nos objectifs sont: (1) de calculer l'IFR en France, (2) d'estimer le nombre de personnes infectées par le COVID-19 en France, et (3) de calculer un nombre de reproduction de base R_0.

2. Matériels et méthodes

2.1. Les données

Nous avons obtenu le nombre de cas positifs et de décès en France, jour après jour, auprès du Johns Hopkins University Center for Systems Science and Engineering [5]. Les données sur le nombre de tests effectués ont

été obtenues auprès de Santé Publique France [6]. Certaines données (cas positifs, décès, nombre de tests) n'étant pas totalement fiables (exemple: 0 nouveau cas détecté en France le 12 mars 2020), nous avons lissé les données avec une moyenne mobile sur 5 jours. Les données officielles sur le nombre de décès par COVID-19 en France ne prennent en compte que les personnes hospitalisées. Environ 728 000 personnes en France vivent dans des maisons de retraite médicalisées (EHPAD, source: DREES [7]). Des données récentes de la région Grand Est (source: Agence Régionale de Santé Grand Est [8]), rapportent un total de 570 décès dans ces maisons de retraite, qu'il faut ajouter au décompte officiel (1015 décès au 31 mars).

2.2. Modèle mécanistique-statistique

Le formalisme mécaniste-statistique, qui devient un standard en écologie [9 , 10 , 11] permet à l'analyste de coupler un modèle mécaniste décrivant une variable latente, ici un modèle d'équation différentielle ordinaire (EDO) de type SIR, et des données incertaines non exhaustives. Pour combler le fossé entre le modèle mécaniste et les données, l'approche utilise un modèle probabiliste décrivant le processus de collecte de données. Une méthode statistique est ensuite utilisée pour l'estimation des paramètres du modèle mécaniste.

Modèle mécaniste. La dynamique de l'épidémie est décrite par le modèle compartimental SIR suivant:

$$S'(t) = -\frac{\alpha}{N} S(t) I(t), \quad I'(t) = \frac{\alpha}{N} S(t) I(t) - \beta I(t), \quad R'(t) = \beta I(t), \tag{1}$$

avec S la population sensible, I la population infectée, R la population rétablie (individus immunisés) et N=S+I+R la population totale, censée être constante. Le paramètre α est le taux d'infection (à estimer) et 1/β est le temps moyen jusqu'à ce qu'un infecté se rétablisse. Sur la base des résultats de [12], la période médiane d'excrétion virale est de 20 jours, mais l'infectiosité a tendance à se disparaître avant la fin de cette période: les résultats de [13] montrent que l'infectiosité commence à partir de 2,5 jours avant l'apparition des symptômes et diminue dans les 7 jours suivant le début de la maladie. Sur la base de ces observations, nous supposons ici que1/β=10 journées.

Les conditions initiales sont $S(t_0)=N-1$, $I(t_0)=1$ et $R(t_0)=0$, où $N=67\ 10^6$ correspond à la taille de la population. Le modèle SIR est démarré à un moment donné $t=t_0$, qui sera estimée et devrait se rapprocher de la date d'introduction du virus en France (ce point est brièvement discuté à la fin de cet article). Le système EDO (1) est résolu grâce à un algorithme numérique standard, utilisant le solveur Matlab ® ode45 .

Ensuite, nous désignons par D(t) le nombre de décès dus à l'épidémie. Notez que l'impact du compartiment D(t) sur la dynamique du système SIR et sur la population totale est ici négligée. La dynamique de D(t) dépend de I(t) par l'équation différentielle:

$$D'(t) = \gamma I(t), \tag{2}$$

avec γ(t) le taux de mortalité des personnes infectées.

Modèle d'observation. On suppose que le nombre de cas testés positifs au jour t, noté $\hat{\delta}_t$, suit des lois binomiales indépendantes, conditionnellement au nombre de tests n_t effectuée le jour t et à p_t la probabilité d'être testé positif dans cet échantillon:

$$\hat{\delta}_t \sim Bi(n_t, p_t). \tag{3}$$

La population testée comprend une fraction des cas infectés et une fraction des susceptibles: $n_t = \tau_1(t)\, I(t) + \tau_2(t)\, S(t)$. Donc,

$$p_t = \frac{\sigma \tau_1(t) I(t)}{\tau_1(t) I(t) + \tau_2(t) S(t)} = \frac{\sigma I(t)}{I(t) + \kappa_t S(t)} \tag{4}$$

avec $\kappa_t := \tau_2(t)/\tau_1(t)$, la probabilité relative de subir un test de dépistage pour un individu de type S par rapport à un individu de type I (probabilité d'être testé conditionnellement à être S / probabilité d'être testé conditionnellement à être I). Nous supposons que le ratio κ ne dépend pas de t au début de l'épidémie (c'est-

à-dire sur la période que nous utilisons pour estimer les paramètres du modèle). Le coefficient σ correspond à la sensibilité du test. Dans la plupart des cas, des tests RT-PCR ont été utilisés et les données existantes indiquent que la sensibilité de ce test utilisant des écouvillons pharyngés et nasaux est d'environ 63 à 72% [14]. Nous prenons ici σ= 0,7 (70% de sensibilité).

2.3. Inférence statistique

Les paramètres inconnus sont α, t_0 et κ. Le paramètre γ(t) est calculé indirectement, en utilisant la valeur estimée de I(t), les données sur D(t) (supposé exact) et la relation (2). La probabilité L est définie comme la probabilité des observations (ici, les incréments $\{\hat{\delta}_t\}$) conditionnellement aux paramètres. En utilisant le modèle d'observation (3), et en supposant que les incréments $\hat{\delta}_t$ sont indépendants conditionnellement au processus SIR sous-jacent et que le nombre de tests n_t est connu, on obtient:

$$L(\alpha, t_0, \kappa) := P(\hat{\delta}_t|\alpha, t_0, \kappa) = \prod_{t=t_i}^{t_f} \frac{n_t!}{(\hat{\delta}_t)!(n_t - \hat{\delta}_t)!} p_t^{\hat{\delta}_t}(1-p_t)^{n_t-\hat{\delta}_t}, \quad (5)$$

avec t_i la date de la première observation et t_f la date de la dernière observation. Dans cette expression, L(α ,t_0, κ) dépend de α ,t_0, κ à travers p_t.

L'estimateur du maximum de vraisemblance (MLE, c'est-à-dire les paramètres qui maximisent L), est calculé à l'aide de l'algorithme de minimisation contraint BFGS, appliqué à -ln(L), via la fonction Matlab ® fmincon . Afin de trouver un maximum global de L, nous appliquons cette méthode à partir de valeurs initiales aléatoires pour α ,t_0, κ choisies uniformément dans les intervalles suivants: α ∈ (0 , 1), t_0∈ (1 , 31), (1er-31 janvier) et κ ∈ (0 , 1). L'algorithme de minimisation est appliqué à 10 000 valeurs initiales aléatoires des paramètres.

La distribution postérieure des paramètres (α ,t_0, κ) est calculée avec une méthode bayésienne, en utilisant des distributions a priori uniformes dans les intervalles donnés ci-dessus (une distribution a priori plus informative a également été testée, voir l' annexe A). Cette distribution postérieure correspond à la distribution des paramètres conditionnellement aux observations:

$$P(\alpha, t_0, \kappa|\hat{\delta}_t) = \frac{L(\alpha, t_0, \kappa)\pi(\alpha, t_0, \kappa)}{C}, \quad (6)$$

où π(α ,t_0, κ) correspond à la distribution a priori des paramètres (donc uniforme) et C est une constante de normalisation indépendante des paramètres. Le calcul numérique de la distribution postérieure est effectué avec un algorithme Metropolis – Hastings (MCMC), en utilisant quatre chaînes indépendantes, chacune avec 10^6 itérations, à partir de valeurs aléatoires proches de la MLE (seule la seconde moitié des itérations est utilisée pour générer le postérieur). Les codes Matlab ® sont disponibles en tant que matériaux supplémentaires .

Les données $\hat{\delta}_t$ utilisés pour calculer la MLE et la distribution postérieure sont ceux correspondant à la période du 29 février au 17 mars.

3. Résultats

Ajustement du modèle. Pour évaluer l'ajustement du modèle, nous avons comparé les observations, c'est-à-dire le nombre cumulé de cas $\Sigma_t := \Sigma_{s=1,\ldots,t}\hat{\delta}_s$ avec l'espérance du modèle d'observation associé au MLE (espérance d'un binôme). À savoir, nous avons comparé Σ_t et $\Sigma_{s=1,\ldots,t}\, n_s p_s^*$ avec

$$p_s^* = \frac{\sigma I^*(s)}{I^*(s) + \kappa^* S^*(s)}. \quad (7)$$

et I*(s), S*(s) les solutions du système (1) (au temps s) associés à la MLE. Les résultats sont présentés dans la figure 1. Nous observons une bonne correspondance avec les données.

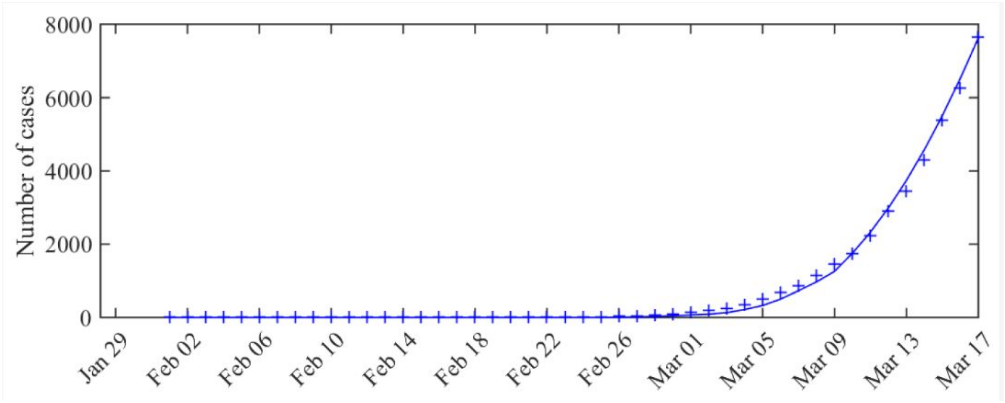


Figure 1. Nombre attendu de cas observés associés à la MLE par rapport au nombre de cas réellement détectés (total des cas). La courbe correspond aux valeurs cumulées de l'observation attendue $n_t\, p_t^*$ données par le modèle, et les croix correspondent aux données (valeurs cumulées de $\hat{\delta}_t$.

Taux de mortalité par infection et nombre réel de cas infectés. En utilisant la distribution a posteriori des paramètres du modèle (les distributions par paires sont présentées à l' annexe A , voir la figure A1), nous avons calculé la distribution quotidienne du nombre réel de personnes infectées. En utilisant la relation (2) avec les données sur $D(t)=\Sigma_t$, on en déduit la distribution du paramètre $\gamma(t)$, à chaque date. L'IFR correspond à la fraction des infectés décédés, soit:

$$IFR_t := \gamma(t)/(\gamma(t)+\beta). \qquad (8)$$

On obtient ainsi, le 17 mars, un IFR de 0,5% (95% -CI: 0,3-0,8), et la distribution de l'IFR est relativement stable dans le temps (voir figure A3 en annexe A).

De plus, la distribution du nombre cumulé de cas infectés (I(t) + R(t)) à travers le temps est présentée dans la figure 2. On observe qu'il est beaucoup plus élevé que le nombre total de cas observés (comparer avec la figure 1). Le rapport moyen estimé entre le nombre réel d'individus infectés et les cas observés$(I(t) + R(t)) / \Sigma_t$ est de huit (IC à 95%: 5–12) sur la période considérée.

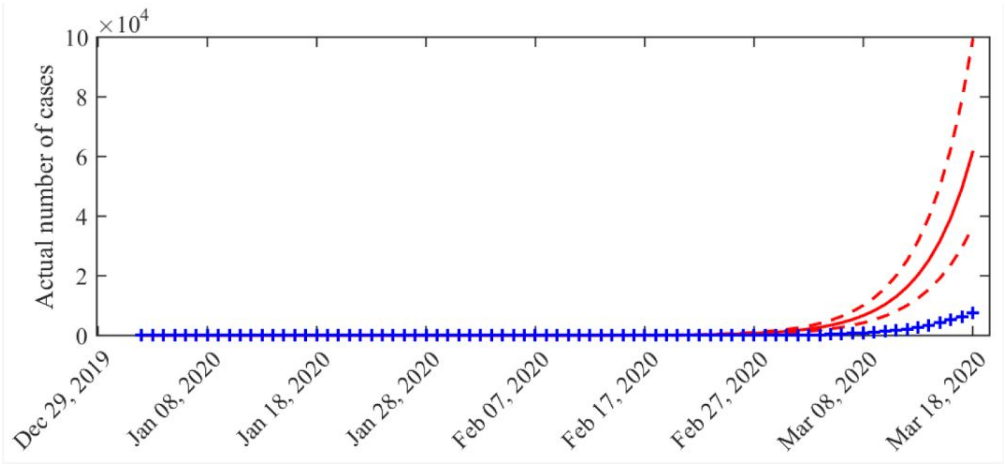


Figure 2. Distribution du nombre cumulé de cas infectés (I(t) + R(t)) à travers le temps. Ligne continue: valeur moyenne obtenue à partir de la distribution postérieure des paramètres. Courbes en pointillés: quantiles postérieurs ponctuels de 0,025 et 0,975. Croix bleues: données (valeurs cumulées de $\hat{\delta}_t$).

Prise en compte des données dans les maisons de retraite médicalisées. Le calcul de l'IFR ci-dessus est basé sur le décompte officiel des décès par COVID-19 en France, qui ne prend pas en compte le nombre de décès dans les maisons de retraite. Sur la base des données locales de la région Grand Est, nous en déduisons que

l'IFR que nous avons calculé a été sous-estimé par un facteur d'environ (1015 + 570) / 1015 ≈ 1,6, conduisant à un IFR ajusté de 0,8% (IC 95%: 0,45-1,25).

Nombre de reproduction de base. Avec les systèmes SIR de la forme (1), le nombre de reproduction de base R_0 peut être calculé directement, sur la base de la formule $R_0 = \alpha / \beta$ [15]. Quand $R_0 < 1$, l'épidémie ne peut pas se propager dans la population. Quand $R_0 > 1$, le compartiment infecté I augmente tant que $R_0 S > N = S+I+R$. Nous avons calculé la distribution postérieure marginale du nombre de reproduction de base R_0. Cela conduit à une valeur moyenne de R_0 de 3,2 (IC à 95%: 3,1–3,3). La distribution complète est disponible dans l' annexe A (figure A4).

Sensibilité des résultats par rapport aux paramètres fixes du modèle. Nous avons calculé la MLE avec une période infectieuse plus longue ($1 / \beta$) de 20 jours estimé par [12]. Cela conduit à un nombre reproductif de base beaucoup plus grand $R_0 = 4,8$ et un facteur ×15 entre les cas signalés et le nombre réel de cas. Cependant, la valeur de l'IFR reste inchangée (0,5%). Nous avons également vérifié si la largeur de la fenêtre du lissage (moyenne mobile sur 5 jours) avait un impact sur nos résultats. Des calculs du MLE avec une largeur de fenêtre de 3 jours (et $\beta = 1 / 10$) conduit aux mêmes résultats que ceux présentés ci-dessus, à savoir une $R_0 = 3,2$ et un IFR de 0,5%.

4. Discussion

Sur l'IFR et le nombre de cas infectés. Le nombre réel d'individus infectés en France est probablement beaucoup plus élevé que les observations (on retrouve ici un facteur × 8), ce qui conduit à un taux de mortalité inférieur à celui calculé sur la base des cas observés: on retrouve ici un IFR de 0,5 % basé sur les données de comptage des décès à l'hôpital, à comparer avec un taux de létalité (TFC, nombre de décès sur nombre de cas diagnostiqués) de 2% le 17 mars. En ajustant le nombre de décès dans les maisons de soins infirmiers, nous avons obtenu un IFR de 0,8%. Ces valeurs pour l'IFR sont cohérentes avec les résultats de [2] (0,66% en Chine) et [3] (0,9% au Royaume-Uni). La valeur de 1,3% estimée sur le bateau de croisière Diamond Princess [4] se situe au-dessus de l'extrémité supérieure de notre IC à 95%. Cela reflète la répartition par âge du navire, qui était biaisée vers les individus plus âgés (âge moyen: 58 ans), parmi lesquels l'IFR est plus élevé [3 , 4].

L'objectif de notre étude était d'estimer l'IFR sur la base des premières données, avant que des enquêtes à grande échelle ne deviennent disponibles. Fin avril, de nouvelles données et études préliminaires sont disponibles et peuvent être comparées à nos résultats. Une étude sur les anticorps à New York publiée le 24 avril 2020 montre qu'environ 14% ont été testés positifs, correspondant à 2,7 millions de cas, à comparer avec les 271000 cas confirmés et un total de 15500 décès dans tout l'État. Cela correspond à un IFR de 0,6%. En France, une autre étude préliminaire menée par l'Institut Pasteur [16], et sur la base d'une estimation conjointe des données françaises jusqu'au 14 avril (données de comptage des décès dans les hôpitaux) et des données des navires de croisière Diamond Princess trouve un IFR de 0,5% confirmant ainsi notre résultat. Au 28 avril, le nombre de décès dus au COVID-19 en Lombardie (Italie) est de 13575 (source: Ministero della Salute), pour une population de 10 millions de personnes, ce qui montre que l'IFR est d'au moins 0,14%. En revanche, en Corée du Sud où le nombre de cas détectés a rapidement atteint un plateau, suggérant une faible proportion de cas non détectés, le rapport entre le nombre de décès et le nombre de cas positifs est de 244/10 752 ≈ 2,3% (source: Johns Hopkins University Center for Systems Science and Engineering [5]), qui peut être considérée comme une limite supérieure de l'IFR, bien que surestimée.

Si le virus conduisait à contaminer 80% de la population française [3], le nombre total de décès à déplorer en l'absence de variation du taux de mortalité (augmentation induite par exemple par la saturation des structures hospitalières, ou diminution liée à de meilleur soins) serait de 336 000 (IC 95%: 192 000–537 000), en excluant le nombre de décès dans les maisons de retraite. Cette estimation pourrait être corroborée ou invalidée lorsque 80% de la population sera infectée, éventuellement sur plusieurs années, en supposant qu'un individu infecté est définitivement immunisé. Il faut noter que les mesures de confinement ou de distanciation sociale peuvent diminuer à la fois le pourcentage d'individus infectés dans la population et le degré de saturation des structures hospitalières.

Sur la valeur de R_0. La distribution estimée en France est élevée par rapport aux estimations récentes (2,0–2,6, voir [3]) mais cohérente avec les résultats de [17] (2,24–3,58). Une estimation directe, par une méthode non mécaniste, des paramètres (ρ ,t_0) d'un modèle de la forme $\hat{\delta}_t = e^{\rho(t-t_0)}$ donne t_0= 36 (5 février) et ρ =0.22. Avec le modèle SIR, I'(t) ≈ I(α - β) pour les petits temps (S≈ N), ce qui conduit à un taux de croissance égal à $\rho \approx \alpha - \beta$, et une valeur de $\alpha \approx 0{,}32$, c'est-à-dire R_0= 3,2 , ce qui est cohérent avec notre distribution de R_0. Notez que nous avons supposé ici une période d'infectiosité de 10 jours. Une période plus courte conduirait à une valeur inférieure de R_0.

Sur l'incertitude liée aux données. L'incertitude sur le nombre réel d'infectés et donc l'IFR est très élevée. Il faut donc interpréter avec prudence les inférences qui peuvent être faites à partir des données dont nous disposons actuellement en France. De plus, nous ne faisons pas ici de prévisions: la dynamique future sera fortement influencée par les mesures de confinement qui seront prises et devra être modélisée en conséquence.

Sur la sensibilité des résultats par rapport aux paramètres fixes du modèle. Nous avons délibérément choisi un modèle parcimonieux avec quelques paramètres pour éviter les problèmes d'identifiabilité. Cependant, nous devions corriger certaines valeurs de paramètres. En particulier, nous avons supposé une durée moyenne de la période infectieuse (1 / β) de 10 jours. Une période infectieuse beaucoup plus longue de 20 jours (correspondant à la période médiane d'excrétion virale trouvée dans [12]) conduirait à un nombre de reproduction de base beaucoup plus grand R_0= 4,8 (mais toujours compris entre 1,4 et 6,49 décrit dans [18]) et un facteur ×15 entre les cas signalés et le nombre réel de cas. Cependant, notre principal résultat sur l'IFR resterait inchangé (0,5%). Nous avons également évalué la sensibilité de l'inférence par rapport aux connaissances antérieures, en proposant un ensemble de distributions antérieures uniformes plus informatives que l'ensemble spécifié dans le texte principal. Dans l'ensemble, cette modification préalable n'influence pas significativement les distributions postérieures; voir l' annexe A .

Sur les hypothèses sous-jacentes au modèle. Les données utilisées ici contiennent peu d'informations, d'autant plus que la période d'observation considérée est courte et correspond à la phase initiale de la dynamique épidémique, qui peut être fortement influencée par des événements discrets. Cette limite nous a conduit à utiliser un modèle particulièrement parcimonieux afin d'éviter des problèmes d'identifiabilité des paramètres. Les hypothèses qui sous-tendent le modèle sont donc relativement simples et les résultats doivent être interprétés au regard de ces hypothèses. Par exemple, la date de l'introduction t_0 doit être considérée comme une date d'introduction efficace pour une dynamique où une seule introduction serait déterminante pour l'éclosion et les autres introductions (antérieure et postérieure) auraient un effet insignifiant sur la dynamique.

Un modèle épidémiologique plus complexe de l'épidémie de COVID-19 en Chine a été proposé dans [19], avec une classe infectieuse divisée en plusieurs compartiments (individus asymptomatiques, infectieux symptomatiques non observés et infectieux symptomatiques observés). Les auteurs utilisent ce modèle dans [20] pour faire des prévisions sur le nombre cumulé de cas en Chine, tout en tenant compte des stratégies de gestion. Dans ces deux études, les auteurs soulignent l'importance de pouvoir estimer la fraction de cas infectieux non observés afin de prévoir la dynamique de l'épidémie. Notre étude, bien que basée sur un modèle SIR plus simple, montre que cette fraction peut être estimée sur la base de données précoces.

Matériel supplémentaire

Les informations suivantes sont disponibles en ligne à l' adresse https://www.mdpi.com/2079-7737/9/5/97/s1 .

Contributions d'auteur

LR, EKK, JP, AS et SS ont conçu le modèle et conçu l'analyse statistique. LR et SS ont rédigé l'article, LR a effectué les calculs numériques. Tous les auteurs ont lu et accepté la version publiée du manuscrit.

Le financement

Ce travail a été financé par INRAE: Réseau MEDIA.

Remerciements

Nous remercions Laurent Desvillettes pour sa suggestion concernant le calcul d'une borne inférieure pour l'IFR à partir des données de la province de Lombardie en Italie. Nous remercions également les réviseurs anonymes pour leurs suggestions et commentaires.

Les conflits d'intérêts

Les auteurs ne déclarent aucun conflit d'intérêt.

Abréviations

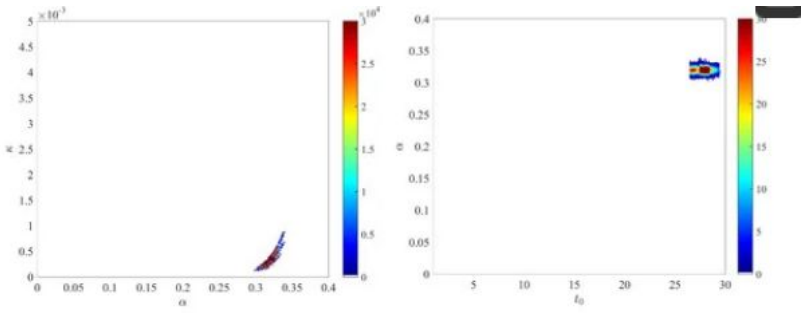
Les abréviations suivantes sont utilisées dans ce manuscrit:

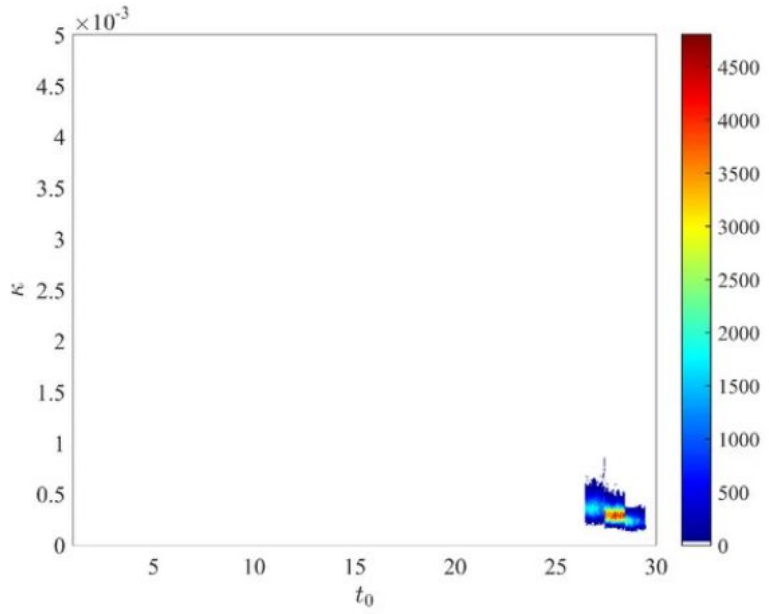
- IFR Taux de mortalité par infection
- CFR Taux de létalité
- EDO Équation différentielle ordinaire
- ARS Agence Régionale de Santé
- OMS Organisation Mondiale de la Santé
- MLE Estimateur du maximum de vraisemblance
- DREES Direction de la recherche, des études, de l'évaluation et des statistiques

Annexe A

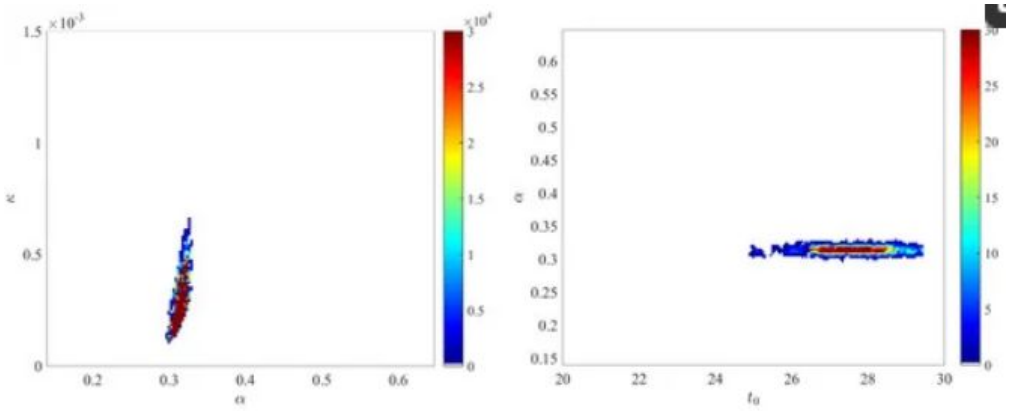
- Les distributions postérieures conjointes des trois paires de paramètres (α , κ), (t_0, α) et (t_0, κ) sont représentés sur la figure A1 .
- Pour vérifier la robustesse de nos résultats vis-à-vis du choix de la distribution a priori, nous avons également considéré le cas d'un a priori plus informatif. À savoir, nous avons supposé les distributions antérieures uniformes suivantes:
    - $\alpha \in (0{,}14 , 0{,}65)$, correspond à $\beta \times R_0$ avec $\beta = 1/10$ et des valeurs de R_0 comprises entre 1,4 et 6,49 (la plage décrite dans [18]);
    - $t_0 \in (20 , 31)$ correspondant à une introduction fin janvier;
    - $\kappa \in (0 , 10^{-2})$, correspondant à une faible probabilité d'être testé pour les cas sensibles, par rapport aux cas infectés.

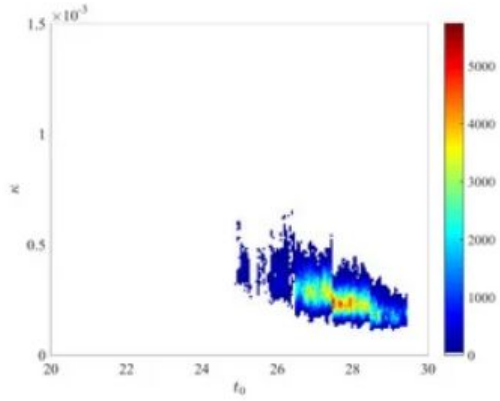
  Nous avons obtenu les distributions postérieures illustrées à la figure A2 , basées sur deux chaînes indépendantes avec 10^6 itérations (seule la seconde moitié des itérations est utilisée pour générer le postérieur). Dans l'ensemble, ces distributions sont relativement similaires à celles affichées sur la figure A1 et obtenues avec les distributions antérieures définies dans le texte principal.
- La dynamique de la distribution estimée de l'IFR est illustrée à la figure A3 .
- La distribution marginale postérieure de R_0 est représenté sur la figure A4 .



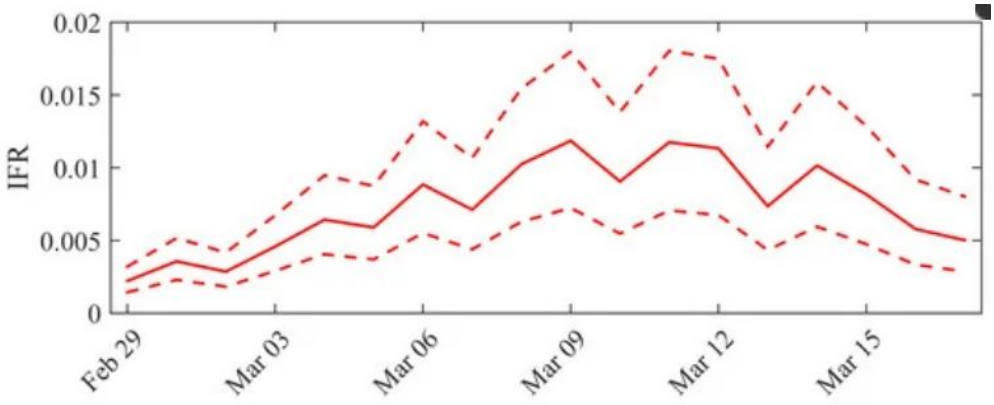


Graphique A1. Distributions postérieures conjointes de(α , κ), (t_0, α) et (t_0, κ).

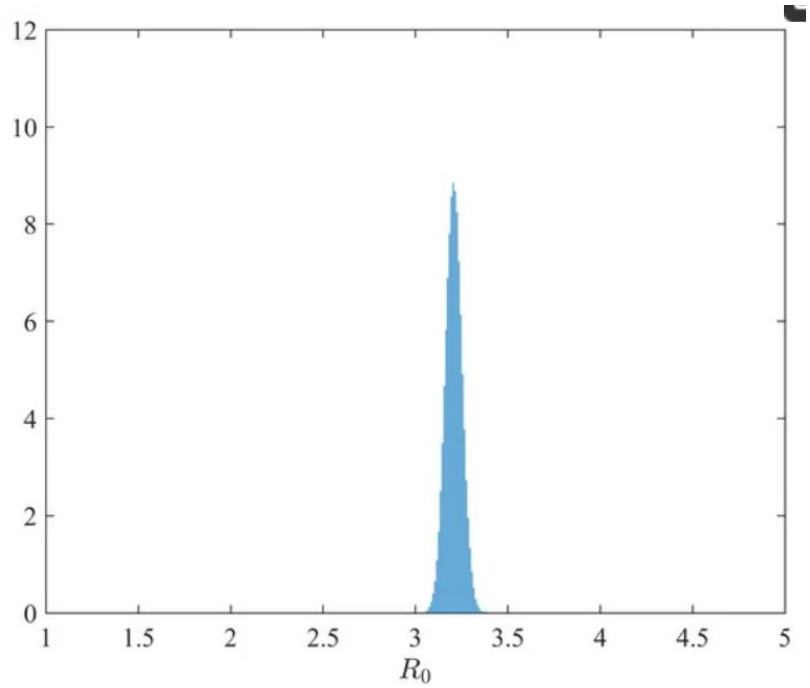




Graphique A2. Distributions postérieures conjointes de(α , κ), (t_0, α) et (t_0, κ) obtenu avec un préalable plus informatif.



Graphique A3. Dynamique de l'IFR en France. Ligne continue: valeur moyenne obtenue à partir de la distribution postérieure des paramètres. Courbes en pointillés: 0,025 et 0,975 quantiles ponctuels.



Graphique A4. Distribution postérieure du nombre de reproduction de baseR_0 En France.

Références

1. World Health Organization. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19—11 March 2020; WHO: Geneva, Switzerland, 2020.
2. Verity, R.; Okell, L.C.; Dorigatti, I.; Winskill, P.; Whittaker, C.; Imai, N.; Cuomo-Dannenburg, G.; Thompson, H.; Walker, P.; Fu, H.; et al. Estimates of the severity of COVID-19 disease. medRxiv 2020.
3. Ferguson, N.M.; Laydon, D.; Nedjati-Gilani, G.; Imai, N.; Ainslie, K.; Baguelin, M.; Bhatia, S.; Boonyasiri, A.; Cucunubá, Z.; Cuomo-Dannenburg, G.; et al. Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand; Imperial College: London, UK, 2020.
4. Russell, T.W.; Hellewell, J.; Jarvis, C.I.; van Zandvoort, K.; Abbott, S.; Ratnayake, R.; Flasche, S.; Eggo, R.M.; Edmunds, W.J.; Kucharski, A.J.; et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. Eurosurveillance 2020, 25.
5. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 2020.
6. Santé Pulique France. COVID-19: Points Épidémiologiques du 17 et 24 Mars 2020. https://www.santepubliquefrance.fr/recherche/#search=COVID-19%20:%20point%20epidemiologique&sort=dat

7. . DREES. 728,000 Résidents en Établissements d'Hébergement pour Personnes Agées en 2015. https://drees.solidarites-sante.gouv.fr/IMG/pdf/er1015.pdf (accessed on 8 May 2020).
8. Agence Régionale de Santé Grand Est. Dossier de Presse—COVID-19: Point de Situation Dans le Grand Est; Agence Régionale de Santé Grand Est: Nancy, France, 2020; https://www.grand-est.ars.sante.fr/system/files/2020-04/DP_point%20de%20situation%20COVID%2019%20en%20Grand%20Est_010420.pdf
9. Roques, L.; Soubeyrand, S.; Rousselet, J. A statistical-reaction-diffusion approach for analyzing expansion processes. J. Theor. Biol. 2011, 274, 43–51.
10. Roques, L.; Bonnefon, O. Modelling population dynamics in realistic landscapes with linear elements: A mechanistic-statistical reaction-diffusion approach. PLoS ONE 2016, 11, e0151217.
11. Abboud, C.; Bonnefon, O.; Parent, E.; Soubeyrand, S. Dating and localizing an invasion from post-introduction data and a coupled reaction–diffusion–absorption model. J. Math. Biol. 2019, 79, 765–789.
12. Zhou, F.; Yu, T.; Du, R.; Fan, G.; Liu, Y.; Liu, Z.; Xiang, J.; Wang, Y.; Song, B.; Gu, X.; et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020.
13. He, X.; Lau, E.H.; Wu, P.; Deng, X.; Wang, J.; Hao, X.; Lau, Y.; Wong, J.Y.; Guan, Y.; Tan, X.; et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. medRxiv 2020.
14. Wang, W.; Xu, Y.; Gao, R.; Lu, R.; Han, K.; Wu, G.; Tan, W. Detection of SARS-CoV-2 in different types of clinical specimens. JAMA 2020.
15. Murray, J.D. Mathematical Biology, 3rd ed.; Interdisciplinary Applied Mathematics 17; Springer: New York, NY, USA, 2002.
16. Salje, H.; Tran Kiem, C.; Lefrancq, N.; Courtejoie, N.; Bosetti, P.; Paireau, J.; Andronico, A.; Hoze, N.; Richet, J.; Dubost, C.L.; et al. Estimating the burden of SARS-CoV-2 in France. medRxiv 2020.
17. Zhao, S.; Lin, Q.; Ran, J.; Musa, S.S.; Yang, G.; Wang, W.; Lou, Y.; Gao, D.; Yang, L.; He, D.; et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int. J. Infect. Dis. 2020, 92, 214–217.
18. Liu, Y.; Gayle, A.A.; Wilder-Smith, A.; Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J. Travel Med. 2020.
19. Liu, Z.; Magal, P.; Seydi, O.; Webb, G. Understanding unreported cases in the 2019-nCov epidemic outbreak in Wuhan, China, and the importance of major public health interventions. Biology 2020, 9, 50.
20. Liu, Z.; Magal, P.; Seydi, O.; Webb, G. Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data. Math. Biosci. Eng. 2020, 17, 3040–3051.