

Par Frédéric Lemoine, PhD en bio-informatique, chief data officer de DataCovID

Et Mathieu Moslonka-Lefebvre, PhD en épidémiologie mathématique, président de DataCovID,

s'exprimant ici à titre personnel

Avant-propos

Ce billet, à visée scientifique et didactique, illustre un **cas d'usage des données du Baromètre Covid 19** avec pour objectif de suivre **l'évolution des symptômes des Français susceptibles de conduire à une orientation vers les services de santé**. Ce suivi est ici réalisé en mobilisant un algorithme d'orientation développé par l'Institut Pasteur et l'Assistance Publique - Hôpitaux de Paris (APHP), et validé par le ministère des Solidarités et de la Santé¹. Cet algorithme, originellement conçu pour orienter "les personnes pensant avoir été exposées au Coronavirus"², a été ici ré-employé en population générale à des fins de suivi épidémiologique.

Les analyses mises en oeuvre n'engagent que leurs auteurs et n'ont pas fait l'objet d'une relecture par les pairs. Les éléments rapportés sont susceptibles d'évoluer en fonction des éventuels commentaires qui seront effectués par la communauté scientifique. Ce billet a vocation à être actualisé chaque semaine et enrichi par des analyses plus élaborées, notamment dans le cadre de travaux scientifiques spécifiques ultérieurs qui seront résumés sur le blog.

En matière de santé publique, il est recommandé de suivre et consulter les instructions officielles disponibles sur <https://www.gouvernement.fr/info-coronavirus>.

Les composantes de l'algorithme d'orientation

Afin d'aider les personnes pensant avoir été infectées par le Covid-19 à bénéficier d'une orientation médicale adaptée, le ministère des Solidarités et de la Santé a publié **un algorithme d'orientation**. Son principe et les données qui le nourrissent sont ici exposés.

Sur la base du **questionnaire en ligne** www.maladiecoronavirus.fr permettant de recueillir les symptômes (par exemple la toux) et les facteurs de risque (par exemple l'obésité), l'algorithme **propose une orientation adaptée au répondant** en suivant un arbre de décision déterministe.³ Au total huit orientations sont proposées par l'algorithme (cf. les deux premières colonnes du tableau en **Annexe 2**).

Construire des indicateurs de la demande potentielle de soins et de la gravité des symptômes potentiellement liés au Covid 19 à partir des sorties de l'algorithme.

L'algorithme précité ayant avant tout été conçu pour orienter les patients qui pensent avoir contracté le Coronavirus, ses prédictions gagnent à être regroupées pour fournir des indicateurs potentiellement adaptés à un **contexte d'application plus large** tel que celui du Baromètre Covid 19 de DataCovID qui couvre l'ensemble de la population française et pas uniquement les cas symptomatiques (voir aussi l'implémentation et la description des adaptations mineures présentées en **Annexe 1**).

Dans un tel contexte élargi, il est proposé d'agréger les sorties de l'algorithme pour proposer des indicateurs de la **demande potentielle de soins** liés à des symptômes pouvant évoquer le Covid 19. Par souci de lisibilité et afin d'optimiser les analyses à venir, il est proposé dans le présent billet d'opérer des regroupements suivant **deux logiques complémentaires** :

- Une **logique thématique** qui permet :
 - En premier niveau d'analyse, de **distinguer les cas qui ne relèveraient pas d'un Covid de ceux qui en relèvent potentiellement** ;
 - En second niveau d'analyse, de distinguer **quatre grands types d'orientation** parmi les cas qui relèvent potentiellement d'un Covid : rester à domicile; consulter un médecin dans le cadre d'une médecine dite "de ville"; consulter un médecin ou appeler le SAMU en cas d'indisponibilité; appeler le SAMU.
- Une **logique de gravité des symptômes**, qui reprend l'ensemble des catégories précédentes dans un format numérique plus adapté pour certains types d'analyse telles que la régression par exemple.

Les nomenclatures proposées ainsi que leur articulation avec les orientations de l'algorithme sont précisées plus en détail dans les deux dernières colonnes du tableau en **Annexe 2**.

Les indicateurs et résultats obtenus en appliquant l'algorithme aux données du baromètre.

-> Un nombre de cas nouveaux potentiels en baisse, avec probablement de nombreux "faux positifs" pour éviter de passer à côté de cas graves ("faux négatifs")

La part des personnes interrogées qui présenteraient un Covid potentiel et qui seraient

orientées en cas d'usage de l'algorithme "*maladiecoronavirus.fr*" serait de prêt d'un Français sur six, ce qui est considérable. Ce chiffre décroît de la vague 1 à la vague 7, de 22% à 16%, soit une baisse de 6 points en 2 mois.

Ce nombre de cas nouveaux potentiels, appelé *incidence*, tels que détectés au travers de cet algorithme sur des échelles de temps intermédiaires (du jour à la semaine) excède largement les estimations réalisées à ce jour pour la France. En effet, le nombre total de cas cumulés (notion de *prévalence* en épidémiologie), depuis le début l'épidémie jusqu'à la sortie du confinement, a été projeté à seulement 4,4 % dans le cadre d'une étude coordonnée par une équipe de modalisation de l'Institut Pasteur.⁴ En outre, la prévalence perçue par les Français dans le cadre de la 7e vague du baromètre Covid 19 est seulement de 5 % pour les personnes les plus sûres d'avoir été touchées.

Il s'ensuit que l'algorithme a probablement été conçu dans l'optique de présenter une sensibilité élevée (peu de faux positifs présentant un caractère de gravité), avec pour contrepartie une faible spécificité (beaucoup de faux positifs). Ce type de paramétrage, qui correspond à une très forte aversion au risque, n'est pas illogique dans une optique de santé publique où il est essentiel de ne pas passer à côté de cas graves, que ces derniers concernent ou non le Coronavirus.

-> Des indicateurs de la charge potentielle des systèmes de santé qui évoluent de façon contrastée : baisse à domicile et en médecine de ville, stabilité globale pour les urgences

A un niveau de lecture plus fin, il est possible de suivre l'évolution de la charge théorique qui s'applique sur les services de santé en ville et d'urgence. Lors de la période sous revue comprenant sept vagues, il est observé une baisse des orientations conseillées vers la médecine de ville et éventuellement relayée par le SAMU dans le cas où le médecin de ville serait indisponible (passage de 9 % à 5 %). La part des orientations directes vers le SAMU suit une trajectoire globalement stable (6 à 7 %). En dehors du système de santé, la part des patients orientés à domicile passe de 6 à 4 %.

A nouveau, on note que l'algorithme sous-tend probablement une approche "prudente" avec un fléchage préférentiel des personnes vers les services de santé, en particulier directement vers le SAMU et donc vers l'hôpital (7 % pour le 7e vague ; soit un chiffre proche des orientations impliquant prioritairement la médecine de ville, de l'ordre de 5 %). Dans la mesure où le questionnaire a été adressé à un échantillon représentatif de la population générale, la très grande fraction de personnes qui seraient renvoyées vers "le 15" (SAMU)⁵, avec environ un Français sur dix concerné pour chacune des trois vagues, peut paraître

étonnante.

-> Une gravité des symptômes déclarés en légère diminution.

La gravité moyenne des symptômes déclarés baisse légèrement de 0,6 à 0,5 sur 4 lors de la période sous revue. Ce chiffre permet de mesurer le très grand “bruit de fond” qui règne dans les données de symptômes dont nombreux ne sont pas spécifiques du Covid 19.

Éléments de conclusion

Au total, il sera intéressant de **suivre les quelques indicateurs ici proposés dans la durée**, notamment **pour caractériser un éventuel rebond épidémique à l’automne** et **mesurer un surcroît potentiel des demandes de prise en charge** par le système de santé dans ses différentes composantes (ville et urgences vitales).

Les premières analyses ici rapportées ont vocation à être confirmées et précisées dans le cadre de travaux ultérieurs. En particulier, il sera nécessaire de **conduire un important travail pour aboutir à des indicateurs fiables de l’incidence**, par exemple en se fondant sur des symptômes qui semblent plus spécifiques du Covid-19 tels que la perte de goût et/ou de l’odorat.⁶

Tableau. Evolution de la demande potentielle de soins des Français liée au Covid 19. Les classifications directement issues de l’algorithme d’orientation sont présentées en **Annexe 3**.

Vague du baromètre	Part des Français qui présenteraient un Covid potentiel et qui serait orientée en cas d’usage de l’algorithme	dont part orientée à domicile	dont part orientée en médecine de ville exclusivement	dont part orientée vers la médecine de ville ou le SAMU	dont part orientée vers le SAMU	Gravité moyenne des symptômes
Vague 1 du 7 au 11 avril	22 %	6 %	5 %	4 %	6 %	0,6 / 4
Vague 2 du 15 au 21 avril	20 %	6 %	5 %	3 %	7 %	0,6 / 4

Vague 3 du 22 au 27 avril	19 %	4 %	5 %	2 %	7 %	0,6 / 4
Vague 4 du 28 avril au 4 mai	18 %	4 %	5 %	3 %	6 %	0,5 / 4
Vague 5 du 5 au 11 mai	16 %	3 %	4 %	2 %	6 %	0,5 / 4
Vague 6 du 12 au 19 mai	17 %	4 %	4 %	2 %	7 %	0,5 / 4
Vague 7 du 26 au 31 mai	16 %	4 %	3 %	2 %	7 %	0,5 / 4

Annexe 1. Implémentations de l'algorithme d'orientation et disponibilité des scripts d'analyse

L'algorithme d'orientation a été implémenté indépendamment dans deux langages de programmation, PHP et R. L'implémentation PHP a notamment permis d'intégrer le calcul aux résultats donnés par l'API de DataCovId. Il est accessible ici <https://datacovid.org/api> (documentation accessible via le téléchargement des données ici <https://datacovid.org/data>), et l'implémentation R de travailler directement sur les données brutes (les scripts d'analyse sont disponibles sur GitHub <https://github.com/datacovid/analyses>).

Les sondages effectués dans le cadre du baromètre autorisent les répondants à ne pas répondre aux différentes questions. Pour certaines réponses, il manque donc des variables nécessaires au calcul de l'orientation. Afin de rester conservateur dans les résultats, il a été choisi de calculer tout d'abord les facteurs de risques majeurs, et de classer comme FIN0 le reste des données pour lesquelles au moins une variable entrant dans le calcul de l'orientation n'est pas renseignée.

Une limite peut être soulevée concernant l'application directe de l'algorithme d'orientation sur les données du baromètre COVID19. Un âge supérieur à 70 ans est défini comme un facteur pronostique défavorable dans la spécification de l'algorithme. En revanche, les tranches d'âges définies dans le baromètre ne permettent pas d'appliquer à l'identique ce facteur pronostique défavorable, et il a été choisi de la remplacer par la tranche d'âge "65 ans et plus". Il convient toutefois de souligner que ce seuil de 65 ans correspond bien aux dernières recommandations en date du Haut Conseil de la Santé Publique.²⁾

Il convient enfin d'être prudent dans l'interprétation des évolutions intervenant entre les vagues 1 et 2 car le périmètre temporel des questions posées a été modifié à partir de la deuxième vague pour permettre des jointures temporelles à l'échelle de la semaine.⁸

Annexe 2. Nomenclature des orientations de l'algorithme

www.maladiecoronavirus.fr et regroupements par thème et par niveau de gravité proposés par les auteurs.

Orientation proposée par l'algorithme validé ⁹	Code de l'orientation ¹⁰	Regroupement thématique proposé par les auteurs (en cas de Covid potentiel, le type d'orientation est précisé en second niveau d'information)	Regroupement numérique par degré de gravité proposé par les auteurs (de 0 pour une gravité minimale à 4 pour une gravité maximale)
<p>Pour tous : Restez chez vous au maximum en attendant que les symptômes disparaissent. Prenez votre température deux fois par jour. Rappel des mesures d'hygiène. Un dispositif national grand public de soutien psychologique au bénéfice des personnes qui en auraient besoin est accessible via le numéro vert : 0 800 130 000.</p>	FIN0 : non classable ¹¹	Non classable	NA (valeur manquante)
<p>Cette application n'est pas faite pour les personnes de moins de 15 ans. Prenez contact avec votre médecin généraliste au moindre doute. En cas d'urgence, appelez le 15.</p>	FIN1	Sans objet (car le panel ne comprends que des personnes dont l'âge est supérieur ou égal à 18 ans).	Sans objet (car le panel ne comprends que des personnes dont l'âge est supérieur ou égal à 18 ans).

<p>Votre situation peut relever d'un COVID 19 qu'il faut surveiller. Si de nouveaux symptômes apparaissent, refaites le test ou consultez votre médecin. Nous vous conseillons de rester à votre domicile.</p>	FIN2	Covid potentiel / Domicile	1
<p>Votre situation peut relever d'un COVID 19. Demandez une téléconsultation ou un médecin généraliste ou une visite à domicile. Appelez le 15 si une gêne respiratoire ou des difficultés importantes pour vous alimenter ou boire apparaissent pendant plus de 24 heures.</p>	FIN3	Covid potentiel / Médecine de ville (puis SAMU si aggravation)	2
<p>Votre situation peut relever d'un COVID 19. Demandez une téléconsultation ou un médecin généraliste ou une visite à domicile. Si vous n'arrivez pas à obtenir de consultation, appelez le 15.</p>	FIN4	Covid potentiel / Médecine de ville ou SAMU	3
<p>Appelez le 15</p>	FIN5	Covid potentiel / SAMU	4
<p>Votre situation peut relever d'un COVID 19. Demandez une téléconsultation ou un médecin généraliste ou une visite à domicile (SOS médecins, etc.)</p>	FIN6	Covid potentiel / Médecine de ville	2
<p>Votre situation peut relever d'un COVID 19. Un avis médical est recommandé. Au moindre doute, appelez le 15. Nous vous conseillons de rester à votre domicile.</p>	FIN7	Covid potentiel / Médecine de ville ou SAMU	3

Votre situation ne relève probablement pas du COVID 19. N'hésitez pas à contacter votre médecin en cas de doute. Vous pouvez refaire le test en cas de nouveau symptôme pour réévaluer la situation. Pour toute information concernant le COVID 19, composer le 0 800 130 000.

FIN8	Non Covid	0
------	-----------	---

Annexe 3. Evolution des prédictions des orientations de l'algorithme www.maladiecoronavirus.fr sur la base des données de DataCovid.

Vague du baromètre	FIN0	FIN1	FIN2	FIN3	FIN4	FIN5	FIN6	FIN7	FIN8
Vague 1 du 7 au 11 avril	559	0	309	258	5	314	13	178	3364
Vague 2 du 15 au 21 avril	655	0	283	239	6	331	23	143	3321
Vague 3 du 22 au 27 avril	729	0	224	248	8	359	19	114	3299
Vague 4 du 28 avril au 4 mai	666	0	200	226	7	301	17	138	3445
Vague 5 du 5 au 11 mai	669	0	172	190	5	293	13	109	3549
Vague 6 du 12 au 19 mai	704	0	210	199	4	331	10	119	3423
Vague 7 du 26 au 31 mai	774	0	178	167	0	330	7	94	3450

1. Cf. <https://delegation-numerique-en-sante.github.io/covid19-algorithme-orientation/>[↵]
2. Cf. <https://www.pasteur.fr/fr/espace-presse/documents-presse/lancement-du-site-internet-maladiecoronavirusfr>[↵]
3. Déterministe signifie entièrement prédictible car ici l'algorithme ne comporte aucune composante probabiliste et donne ainsi toujours le même résultat pour une entrée donnée. Le diagramme de l'arbre est présenté ici : <https://raw.githubusercontent.com/Delegation-numerique-en-sante/covid19-algorithme-orientation/master/diagramme-algorithme-orientation-covid19.png>. [↵]
4. Cf. Salje et al. (2020) *Estimating the burden of SARS-CoV-2 in France*. *Science*[↵]
5. directement ou en relai de la médecine de ville[↵]
6. Cf. <https://www.pasteur.fr/fr/espace-presse/documents-presse/covid-19-premiere-etude-seurologique-france-deja-beaucoup-enseignements>, et notamment : "l'étude [au niveau d'un cluster de cas qui ne reflète pas nécessairement la population générale] révèle

deux symptômes majeurs permettant d'identifier qu'une personne a été contaminée : 84,7% des personnes ayant eu une perte d'odorat et 88,1% ayant eu une perte du goût sont infectées". [↵]

7. Dans le cadre de la préparation de la phase de déconfinement, le Haut Conseil de la santé publique (HCSP) a récemment actualisé l'avis relatif aux personnes à risque de forme grave de Covid-19. Aux termes de cette mise à jour, il apparaît que les personnes considérées comme présentant un risque de développer une forme grave de Covid-19 sont notamment les personnes âgées de 65 ans et plus (même si les personnes âgées de 50 ans à 65 ans doivent être surveillées de façon plus rapprochée). Cf. <https://www.hcsp.fr/explore.cgi/avisrapportsdomaine?clefr=807> (référence de l'avis : "hcspa20200420_covperrisetmesbarspccesper.pdf"[↵]
8. Les questions relatives aux symptômes demandant d'indiquer une réponse "Dans les dernière 24 heures" ou "Ces derniers jours" ont été remplacées à compter de la vague 2 par la formule "Au cours des 7 derniers jours". Cette évolution permet d'assurer la couverture d'une période entière d'une vague à l'autre.[↵]
9. Toutes les orientations décrites correspondent strictement à l'algorithme officiel mis à part pour le code orientation "non classable" (FIN0) proposée par les auteurs afin de traiter les données manquantes pour lesquelles la documentation de l'algorithme ne prévoit pas un traitement explicite.

[↵]

10. D'après le pseudo-code de l'algorithme accessible ici : <https://github.com/Delegation-numerique-en-sante/covid19-algorithme-orientation-arbre-de-decision/blob/master/covid19-orientation-arbre-de-decision.txt>. L'arbre de décision pris pour référence dans ce billet est daté du 26 avril 2020 et porte le numéro de commit github « 3966ac8 ».

[↵]

11. Catégorie ajoutée par les auteurs mais correspondant au message d'orientation proposé à tous

[↵]