MVA: Biostatistics (2020/2021)

Final Exam

Autor: *Charbel-Raphaël Segerie* *(March 2021)*

Articles choisis
J'ai choisi d'étudier l'article portant sur le Diamond Princess cruise ship de Mizumoto et al. [2020] ainsi que l'article portant sur la prédiction de la mortalité associée à la Covid de Pourhomayoun and Shakibi [2020] analysée à l'aide de l'outil PROBAST, Wolff et al. [2019].

1 Diamond Princess cruise

Le but de cet article est de trouver **la proportion de cas asymptotiques** parmi les nouveaux infectés de la Covid sur le navire de croisière.

1.1 Décrivez l'approche de modélisation utilisée

Cadre de l'étude Sur le navire de croisière, après l'apparition des premiers cas, le navire a été mis en quarantaine. Des tests PCR ont été conduits entre le 5 et le 20 février 2020.

Modélisation de la vraisemblance pour les testés positifs Pour chaque personne testée positive, l'article fait l'hypothèse qu'il n'y a que deux possibilités :
Soit l'individu comporte des symptômes, soit l'individu ne comporte pas de symptômes. Dans le cas où l'individu ne comporte pas de symptômes, soit l'individu est réellement asymptomatique, c'est-à-dire qu'il ne présentera jamais de symptômes, soit l'individu se révèle symptomatique dans le futur. La proportion de personnes qui bien que positives ne développera jamais de symptômes est posée comme valant p.

Temps d'incubation des symptômes Le temps d'apparition des symptômes à partir de la contamination suit une loi de Weibull de moyenne et d'écart-type 6.4 et 2.3 jours, et qui est donc une distribution en forme de cloche, telle une gaussienne.
Ainsi, pour chaque individu i testé positivement, on pose un intervalle $[a_i, b_i]$ qui contient la date X_i de l'infection de l'individu. Les b_i sont toujours choisis comme étant le moment où ils sont testé positivement. L'article pose c la date de censure de l'individu asymptomatique.

Vraisemblance totale Ainsi, l'article explique que pour un individu i exposé pendant l'intervalle $[a_i, b_i]$, la probabilité pour eux d'être asymptomatique est de $g(x, p)$ qui vaut $p + (1-p)(1-F_D(c-x))$ si l'individu n'a pas présenté de symptôme jusqu'au temps de censure et vaut $(1-p)F_D(c-x)$ si l'individu présente des symptômes avant la date de censure. Mais même si les formules sont correctes l'article n'est pas exact dans son explication : $g(x, p)$ n'est pas la probabilité d'être asymptomatique, mais est **la vraisemblance du modèle dans les deux cas**. (En effet, sinon la probabilité d'être asymptotique est nulle s'ils ont des symptômes !). On peut ensuite faire l'hypothèse de l'indépendance entre les différents tests positifs et multiplier les vraisemblances pour chacun des tests positifs.

Implémentation Au niveau computationnel, l'article utilise par la suite un sampling par Hastings-Metropolis et un NUTS sampler pour trouver la distribution postérieure de tous les paramètres X_i et D_i pour chaque individu, et la proportion d'asymptomatique parmi les porteurs p.

1.2 Décrivez les résultats obtenus

1.2.1 Principal résultat

Le résultat principal est celui qui donne la vraie proportion p des asymptomatiques parmi les porteurs qui serait de 17.9% (l'intervalle de confiance à 95% est de 15.5-20.2%). Mais cela semble **contradictoire avec l'analyse de sensibilité** en faisant varier la moyenne du temps d'incubation, l'article propose un intervalle pour p entre 20.6 et 39.9%, intervalle qui ne contient pas les 17.9% ! Il y a donc soit un problème, soit une non-monotonicité dans l'analyse de sensibilité.

1.2.2 Heatmaps

L'article présente également en annexe les heatmaps des temps d'infections pour les individus symptomatiques et asymptomatiques.

Erreur dans l'interprétation des heatmaps L'article semble se tromper sur l'interprétation de la heatmap des asymptomatiques : "Among the symptomatic cases, the infection timing appears to have occurred just before or around the start of the quarantine period (Supplementary Figure S1), while the infection timing for asymptomatic cases *appears to have occurred well before the start of the quarantine period*". Ce que j'ai mis en italique me semble faux, car la heatmap montre que le gros des infections des cas asymptomatiques a eu lieu *après* le 5 février soit après la mise en quarantaine.

Politique de test et différences entre les heatmaps On peut également remarquer que **les heatmaps pour asymptomatiques et symptomatiques sont très différentes**. Cela est inattendu et montre dans quelle mesure les symptomatiques ont été testés beaucoup plus souvent que les asymptomatiques au début de l'étude ce qui fausse grandement les résultats et sous estime la probabilité p. (Peut-être qu'il y a un moyen à partir de la comparaison entre les deux heatmaps de compenser la différence de tests, par exemple à l'aide d'un équilibrage des histogrammes, ou bien en utilisant la probabilité finale $p = 17.8\%$ pour estimer le nombre de cas asymptomatique que l'on a raté au début et compenser en conséquence la probabilité p). Mais en tout cas, les heatmaps sont censés avoir la même forme modulo le temps d'incubation de 6.4 jours qui explique le fondu vers le blanc sur droite de la heatmap des symptomatiques.

1.2.3 Tableau des résultats

Le fait que **la politique de test** ait un rôle primordial est attesté par le tableau principal de l'étude : les 5, 6, 7, 8, et 9 février, la proportion de ceux testés qui sont positif est de successivement approximativement 1/3, 1/7, 1/4, 1/10, 1/2 (nombre d'individus positifs divisé par le nombre de tests). Une telle variance n'est explicable que par une différence au jour le jour dans la politique de tests.

1.2.4 Autre

- On voit la grande **influence du prior uniforme** dans les heatmaps : toute la gauche des heatmaps est uniforme alors que l'épidémie a progressé de façon exponentielle les premiers jours et donc le prior devrait être exponentiel.
- Le **dernier jour utilisé dans la heatmap** est le jour 27 ce qui correspond au 15 février, alors que le tableau principal va jusqu'au 20 février.
- Vraisemblablement, ce ne sont pas des NA dans la colonne "number of tests" du tableau des résultats mais des 0.

1.3 Commentez l'approche de modélisation

1.3.1 Biais

Biais de sélections Il n'y a qu'une seule phrase sur la politique de test : "Laboratory testing by PCR had been conducted, prioritising symptomatic or high-risk groups.", ce qui est très flou, alors que c'est la politique de test menée qui détermine toute la probabilité inférée p.
La modélisation est entièrement basée sur **l'hypothèse que l'on repère tous les porteurs ce qui est vrai pour les symptomatiques, mais faux pour les asymptomatiques**. Même si l'article reconnaît ce problème rapidement, il ne discute pas assez de la manière dont les patients ont été sélectionnés pour être testés et ne discute absolument pas de la possibilité de corriger les biais de sélections.

Biais de mesures Les tests PCR ne sont sans doute pas sensible à 100%, mais à 70% d'après COV [2021]: "The patients were supposed to have two PCR tests at intervals to leave the hospital, but in many cases the second was positive, even though the first was negative. Tamura said that he felt the sensitivity of PCR test was about 70%". Cela montre que la proportion d'asymptomatique est encore une fois sous-estimée.

Correctifs possibles Ce n'est pas une critique de la modélisation, mais une critique de la politique de tests. Vu le nombre de tests cumulés menées sur les 17, 18, et 19 février qui représente à peu près 1500 tests soit la moitié du navire, il me semble qu'il existe une politique de tests bien plus efficace et sans biais en faisant le même nombre de tests : **il aurait suffi de tirer au sort 1000 personnes à tester au hasard parmi l'ensemble des personnes sur le navire**, puis à l'aide des 500 tests supplémentaires de tester les 270 personnes présentant des symptômes ces jours-là, afin que ces derniers soient soignés, mais pas compté dans l'étude. (statistiquement, 1/3 des 270 personnes présentant des symptômes feraient partie des gens testés parmi les 1000 personnes sélectionnées aléatoirement, mais ce n'est qu'un détail). Les auteurs de l'article auraient alors pu utiliser les personnes testées positives parmi les 1000 pour faire une étude en éliminant les biais de sélections relatifs à la sélection des symptômes. Cette étude sur 1000 personnes eût été presque aussi statistiquement précise que l'étude menée sur 3000 personnes, le biais de sélection en moins. Mais je suis bien conscient qu'à l'époque, on ne savait pas que les asymptomatiques étaient si prévalent. La critique est facile.

Cependant, ne prendre en compte que la fin de l'étude avec la politique de test proposée précédemment et ne pas prendre en compte les débuts hasardeux de la politique de test aurait permis d'éviter aux auteurs d'allouer de manière aléatoirement uniforme les 35 asymptomatiques entre le 5 février et le 13 février, ce qui est surement une mauvaise chose à faire compte tenu du fait que **la progression de l'épidémie dans le navire a été exponentielle, et non constante et uniforme**.

Il suffit de diviser par 0.7 la proportion p finale obtenue pour corriger le biais de mesure des tests PCR dans l'hypothèse où chaque individu n'a été testé que par un seul test PCR.

1.3.2 Nombre de variables et features choisies

Au niveau de la modélisation computationnelle sur le logiciel R, les auteurs auraient pu **réduire considérablement le nombre de variables** en rassemblant tous les individus asymptomatiques et tous les individus symptomatiques testés le même jour sous une même variable pondérée de manière appropriée.

Il n'y a **pas eu de différence de traitement suivant les âges** or, on sait aujourd'hui que c'est la variable principale pour déterminer la gravité de l'évolution des symptômes. Étant donné la modélisation bayésienne en grande dimension utilisée, il n'aurait pas couté grand-chose vu la taille de l'étude de rajouter une dépendance linéaire de la probabilité d'être asymptomatique suivant l'âge, par exemple $p(\text{age}) = \sigma(a * \text{age} + b)$, avec a et b qui eurent été simplement deux paramètres supplémentaires et σ la fonction sigmoid. Cela permettrait donc de partiellement corriger le problème du biais de sélection relevant de la population sur le navire, qui est en moyenne âgée.

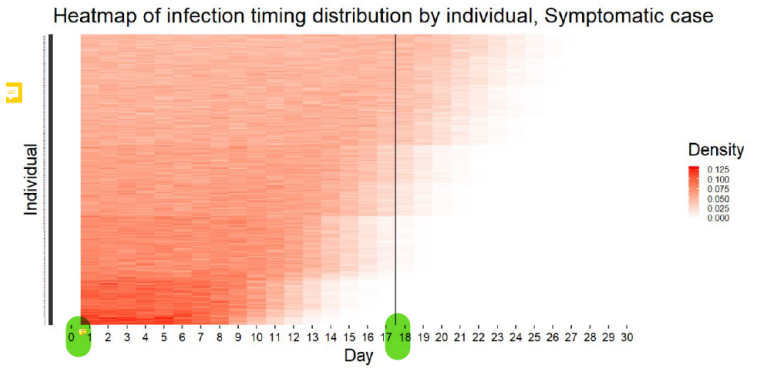


Figure 1: La heatmap des symptomatiques devrait être davantage diagonale et non triangulaire car le temps d'incubation de la Covid-19 est de 5 jours et les symptomatiques sont repérées rapidement sur le navire en quarantaine.

1.3.3 Oublis de la modélisation

Un autre problème avec la heatmap des symptomatiques 1 est que la distribution est en forme de triangle ce qui me parait très bizarre. En effet, je pense que la distribution devrait être en forme de "diagonale" et non "triangulaire" 1, car les individus symptomatiques sont aussitôt repérés et il parait invraisemblable que des individus présentant des symptômes soient infectés le premier jour, mais testés autant de temps après. Cela vient du fait que la modélisation utilise un prior uniforme.

De la même manière qu'il y a un temps d'incubation pour les symptomatiques, il y a aussi certainement un temps d'incubation pour les asymptomatiques, temps nécessaire avant d'être repéré par les tests PCR, que l'article ne prend pas en compte. Ainsi, une fois de plus, l'article fait l'hypothèse qu'il repère tous les asymptomatiques, ce qui est faux et la proportion d'asymptomatiques est encore une fois sous-estimée.

1.4 Conclusion

Cet article présente un paradigme de modélisation intéressant, mais fait des erreurs regrettables dans l'interprétation des résultats et a la malchance de cumuler des biais qui sous estiment systématiquement la proportion d'asymptomatiques : L'article fait l'hypothèse que tous les asymptotiques sont repérés Les biais de mesures, car la sensibilité des tests PCR n'est que de 70% et n'est sans doute pas égale entre les symptomatiques et les asymptomatiques. L'article fait l'hypothèse que les asymptomatiques sont repérables immédiatement (durée d'incubation négligée), ce qui est sans doute faux.

2 Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making

2.1 Comment évaluez-vous le risque de biais sur l'item "Participants" ?

2.1.1 Were appropriate data sources used, e.g., cohort, RCT, or nested case–control study data?

(?) Non, la donnée venant des 146 pays est volumineuse mais non contrôlée.

2.1.2 Were all inclusions and exclusions of participants appropriate?

(-) "We have also removed the unlabeled data samples." D'où vient cette donnée non labellisée ? C'est un peu suspect.
De plus, l'article utilise le data balancing, hors on sait que la Covid frappe les personnes âgées beaucoup plus que les jeunes. Le dataset utilisé par les auteurs surreprésente donc les personnes très âgées, même

parmis les personnes constituant le datasets, qui vient de personnes admise à l'hôpital et qui donc sont déjà âgées. On aurait aimé savoir quel est le pourcentage exact de positif dans le dataset et la procédure exacte pour équilibrer le dataset.

2.2 Comment évaluez-vous le risque de biais sur l'item "Predictors" ?

2.2.1 Were predictors defined and assessed in a similar way for all participants?

(-) La donnée venant de 146 pays différents, il est peu probable que les critères aient été définis de la même manière tout autour du globe. Cependant, si l'on conserve les features "pays" ainsi que "laboratoire", puisque les différents laboratoires sont certainement consistants au moins en interne, cela devrait réduire le biais des prédicteurs.
D'autre part, certaines caractéristiques des prédicteurs sont très suspectes. On peut voir en figure 2 que le nombre de diabétiques est **anormalement bas**, alors que celui-ci devrait toucher environ une personne sur 10 de plus de 60 ans parmi les 12000 que contient cette analyse. Ce genre d'anomalie est également présent sur les autres features, et montre que les différentes features n'ont pas été récoltées dans les différents centres.

2.2.2 Were predictor assessments made without knowledge of outcome data?

(+) Cela n'est pas précisé dans l'article, mais ne semble pas être le cas, les analyses ayant été faites je pense du vivant des individus.

2.2.3 Are all predictors available at the time the model is intended to be used?

(?) Les prédicteurs se décomposent en trois familles : les symptômes, les comorbidités ainsi que les informations démographiques, tous semblants bien être disponibles du vivant de l'individu, au moment des analyses. Mais le diagnostic de certaines maladies n'est potentiellement pas aussi facilement disponible dans tous les pays et tous les centres.

2.3 Comment évaluez-vous le risque de biais sur l'item "Outcome" ?

2.3.1 Was the outcome determined appropriately ?

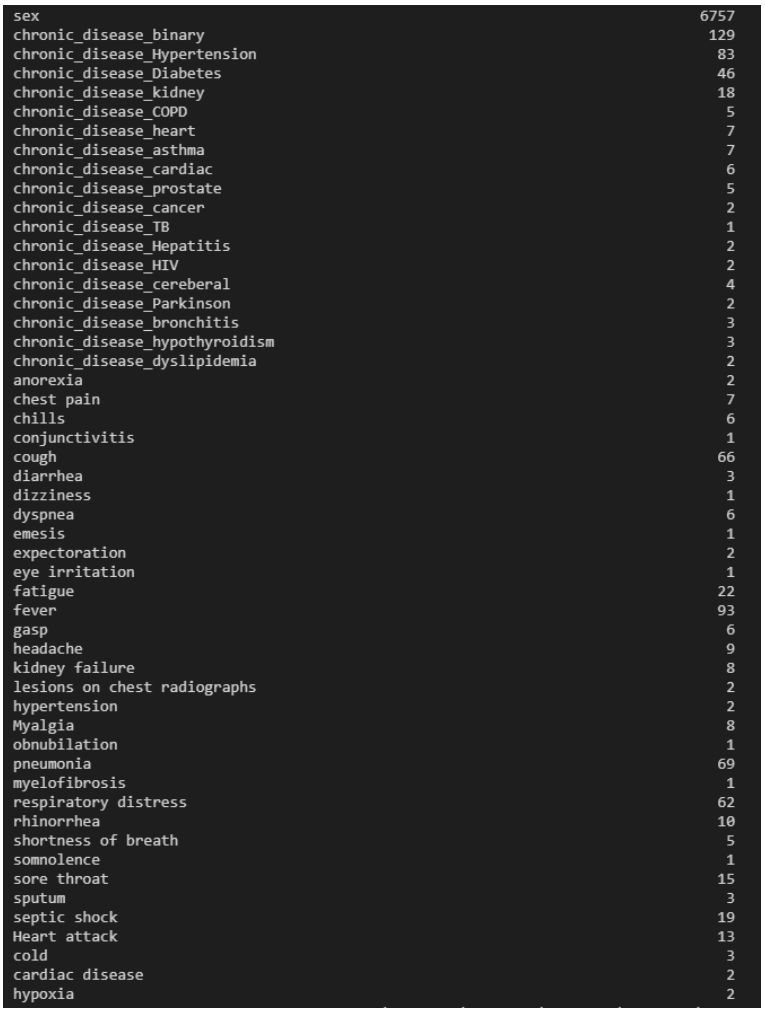
(?) On peut voir que le dataset original *Labeled_only_dataset* ne contient pas que des variables booléennes comme vu en figure 3. L'article ne précise pas comment les labels délicats ont été intégrés. On peut ensuite voir dans le code que le traitement fait par l'article est en fait très simple comme indiqué en figure 4, mais exagère le nombre de morts.

2.3.2 Was a prespecified or standard outcome definition used?

(?) La mort est un concept relativement non ambigu. Par contre l'article à traité de la même manière les patients en états critique et le critère de la sortie des hôpitaux est potentiellement délicat : "We considered patients that were discharged from hospital or patients in stable situations with no more symptoms as recovered patients". Il faut faire attention au fait que peut-être, les patients dans les pays pauvres sortent de l'hôpital plus rapidement que dans les pays riches par faute de moyens, il a donc potentiellement **des biais entre les différents pays**. On peut également imaginer que les personnes non assurées dans les pays sans service social ne vont pas rester longtemps à l'hôpital malgré les risques qu'elles encourent. Étant donnée la grande diversité des pays étudiés, cela mérite des examens supplémentaires.

2.3.3 Were predictors excluded from the outcome definition?

(?) La définition de l'outcome "mort" ne fait pas intervenir les prédicteurs utilisés. Par contre, il aurait fallu vérifier que l'outcome "sortie d'hôpital" ne dépende pas des caractéristiques physiologiques des différents patients. Par exemple, les médecins peuvent soigner différemment les personnes, suivant la gravité de l'avancement de la maladie.



sex	6757
chronic_disease_binary	129
chronic_disease_Hypertension	83
chronic_disease_Diabetes	46
chronic_disease_kidney	18
chronic_disease_COPD	5
chronic_disease_heart	7
chronic_disease_asthma	7
chronic_disease_cardiac	6
chronic_disease_prostate	5
chronic_disease_cancer	2
chronic_disease_TB	1
chronic_disease_Hepatitis	2
chronic_disease_HIV	2
chronic_disease_cereberal	4
chronic_disease_Parkinson	2
chronic_disease_bronchitis	3
chronic_disease_hypothyroidism	3
chronic_disease_dyslipidemia	2
anorexia	2
chest pain	7
chills	6
conjunctivitis	1
cough	66
diarrhea	3
dizziness	1
dyspnea	6
emesis	1
expectoration	2
eye irritation	1
fatigue	22
fever	93
gasp	6
headache	9
kidney failure	8
lesions on chest radiographs	2
hypertension	2
Myalgia	8
obnubilation	1
pneumonia	69
myelofibrosis	1
respiratory distress	62
rhinorrhea	10
shortness of breath	5
somnolence	1
sore throat	15
sputum	3
septic shock	19
Heart attack	13
cold	3
cardiac disease	2
hypoxia	2

Figure 2: Petite analyse que j'ai menée pour chacune des features dans le dataset utilisé (Balanced-Dataset) qui contient 12020 cas. Les chiffres donnent le nombre de cas où la features est positive. On peut voir que la majorité des features utilisées dans les heatmaps ne sont présentes que pour moins de 5 individus comme pour l'HIV, cérébral, Parkinson, hypothyroidism et dyslipidemia, dont les chiffres sont d'ailleurs anormalement bas. On peut remarquer qu'il n'y a qu'un seul individu obnubilé dans tout le dataset, feature pourtant utilisée par l'article et validée par Cross Validation. L'article aurait du trier en amont les features grâces aux connaissances cliniques.

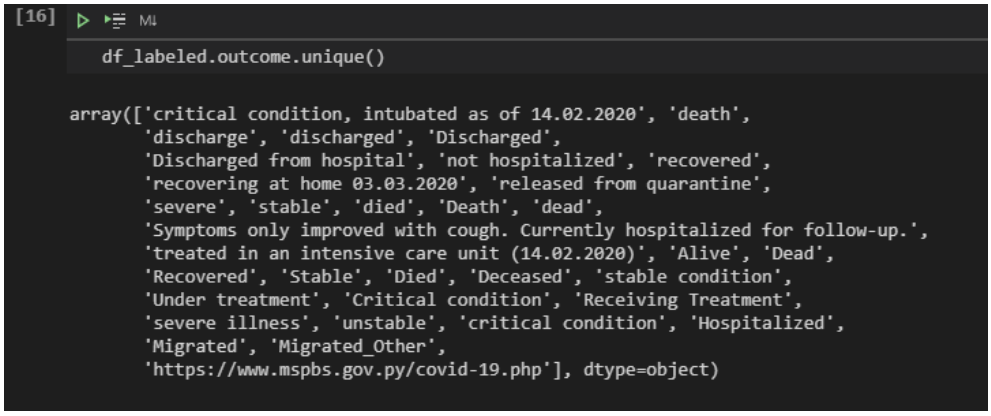


Figure 3: J'ai affiché ici les différents outcome possible. La principale difficulté ne sont pas les erreurs flagrantes qui sont inévitables étant donnée la taille des données. Par contre, on voit que de nombreuses personnes sont en situation critiques ou non stables, et leur attribuer une variable booléenne est compliqué.

```
df=df.replace('died',0)
df=df.replace('recovered',1)
df=df.replace('stable',1)
df=df.replace('severe',0)
```

Figure 4: Copie d'écran du code original des auteurs. Les individus en état critique ont été traités de la même manière que les individus décédés. Cela permet d'esquiver une étude de la donnée censurée. Mais cela aurait dû être précisé dans l'article.

2.3.4 Was the outcome defined and determined in a similar way for all participants?

(?) Idem à 3.2

2.3.5 Was the outcome determined without knowledge of predictor information?

(-) On sait que pendant la crise sanitaire les hôpitaux étaient malheureusement surmenés, et on a eu recours au **triage**. Les personnes ayant le plus de chance de survie ont été admises en réanimation, et on a alors modifié les outcomes des différents patients suivant leurs features, et il faudrait analyser à quel point ça a été fréquent.

2.3.6 Was the time interval between predictor assessment and outcome determination appropriate?

(?) Nous n'avons pas l'information. Le plus probable étant que l'étude a utilisé les données disponibles quel que soit le temps entre leur recueillement et l'issue finale du patient. Cependant, ce problème sera représentatif des problèmes qui se manifesteront également en pratique, donc cela ne semble pas être un grave problème.

2.4 Comment évaluez-vous le risque de biais sur l'item "Analysis" ?

2.4.1 Were there a reasonable number of participants with the outcome?

(?) Le nombre de participants est très élevé et se compte dans l'ordre des 3 millions. Mais étant donné le caractère de la Covid qui est une maladie qui ne tue qu'environ 1% des malades, le nombre de décès dans l'étude doit se compter dans l'ordre des 30 milliers, ce qui reste suffisant, mais nécessitera alors une étape de feature selections.
Lorsque l'on regarde le code, on s'aperçoit que **l'étude n'utilise que 6010 cas de guérison et 6010 décès**, ce qui est finalement assez peu, et encore moins que les 30 milliers estimés a priori. Cela n'est pas du tout précisé dans l'article.
On obtient donc une centaine de features pour seulement 10000 cas : Il faudra donc être vigilant lors de l'étape de feature selection. Mais a priori, le nombre de participants reste suffisant.

2.4.2 Were continuous and categorical predictors handled appropriately?

(-) L'information n'est pas présente dans l'article. On ne sait pas quel embedding a été choisi pour coder les features catégoriques. Après analyse du code, les auteurs ont utilisé un dummy encoding. Par contre, un effet indésirable est que toutes les features catégorielles du même type par exemple le type pays : "france", "allemagne", "Liban", ne sont pas a priori présente simultanément. L'article n'utilise qu'une sous sélection de pays, par exemple les petits pays auront tendance à être évincés, car l'article utilise seulement une sélection de features automatiques et non contrôlées, ce qui est potentiellement embêtant en fonction de l'utilisation prévue du modèle.

2.4.3 Were all enrolled participants included in the analysis?

(-) L'article dit qu'ils ont créé un dataset équilibré, donc a priori, 99% des individus non morts n'ont pas été inclus dans l'analyse.

2.4.4 Were participants with missing data handled appropriately?

(?) "data imputation techniques including mean/median/mode value replacement and KNN technique were used to handle missing values" : la phrase entre guillemets constitue les informations dont nous disposons.
Mais l'utilisation de KNN pour l'imputation de données en grande dimension est délicate. En effet, l'utilisation de KNN demande à se poser la question de la métrique utilisée, généralement euclidienne. Hors, ici, toutes les features n'ont pas la même importance et utiliser une métrique euclidienne donne la

même importance à toutes les features. De plus, **le code qui s'occupe de l'imputation de features (missing data) est introuvable**.

2.4.5 Was selection of predictors based on univariable analysis avoided?

(-) Non, l'article utilise différentes méthodes de features sélection, dont une méthode univariée. Mais l'article annonce par la suite que la sélection de feature est entièrement validée par cross validation. Mais on ne retrouve aucune telle analyse dans le code. La seule chose qui est faite est l'utilisation de la pipeline de Sklearn avec des modèles qui utilisent une pénalité L1 sur les features utilisées. Mais les features sélectionnées par Sklearn n'ont jamais été analysées et encore moins "match the original feature selection" ce qui eût été dans tous les cas invraisemblable étant donné le nombre de features utilisées.

2.4.6 Were complexities in the data (e.g., censoring, competing risks, sampling of control participants) accounted for appropriately?

(-) D'après l'article : "We considered patients that were discharged from hospital or patients in stable situation with no more symptoms as recovered patients.", donc a priori il n'y a pas de risque de donnée censurée. Mais il s'avère qu'étant donnée le traitement des patients sévèrement malades 2.3.1 et la figure 4, le problème de la **donnée censurée** n'est pas traité par l'article.
De plus, rien n'est précisé à propos des **competing risks**, or on sait que les patients gravement malades de la Covid-19 sont souvent très âgés avec des comorbidités. Donc les risques concurrents sont donc très grands.

2.4.7 Were relevant model performance measures evaluated appropriately?

(?) Pas vraiment. Le modèle utilise des métriques qui ne sont pas vraiment informatives pour les datasets non équilibrés tels que l'accuracy. On aurait préféré l'utilisation du F1-score, ou du recall, étant donné que le nombre de morts est très inférieur au nombre de recovered.
De plus, les courbes ROC peuvent parfois être trompeuses dans certaines applications très déséquilibrées. Une courbe ROC peut par exemple sembler assez bonne tout en classant de manière erronée la plupart ou la totalité de la classe minoritaire.
Et on ne sait pas a priori dans l'article si les courbes ROC et les autres métriques ont été calculées à partir du dataset équilibré ou du dataset représentatif. **Après examen du code, aucune vérification n'a été menée sur le dataset représentatif**.

2.4.8 Were model overfitting, underfitting, and optimism in model performance accounted for?

(-) Non. On n'a pas d'information sur l'overfitting et l'underfitting. A priori, vu le grand nombre de features utilisées, le risque d'overfitting est assez grand. Par exemple, la régression logistique fait presque aussi bien que le réseau de neurone ce qui est très dommageable pour le réseau de neurone et laisse présager des comportements inattendus et inappropriés.
Les modèles ont été testés non pas dans le dataset représentatif mais seulement dans le dataset équilibré. Cela peut éventuellement causer des problèmes par exemple avec le modèle en régression logistique qui n'a pas d'ordonnée à l'origine (le fit intercept = False) et donc qui est "calibré" pour fonctionner uniquement en environnement équilibré.
De plus le fait que la régression logistique ait une AUC de 0.92 qui est aussi proche de l'AUC du réseau de neurone qui est de 0.93, (mais qui est **non interprétable** et choisi par cross validation et sélection de méta paramètre) montre à quel point l'article donne de l'importance à l'accuracy, métrique pourtant peu pertinente.
Enfin, l'article n'utilise pas de validation externe.

2.4.9 Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?

(?) La figure 6 dans l'article original est totalement inutile et noie le poisson alors qu'on aurait pu allouer cette place à **l'analyse des features pour la régression logistique, qui est malheureusement absente** et qui nuit énormément à l'interprétabilité du modèle et de la data.

2.5 Quelle est votre conclusion globale sur le risque de biais ?

Conclusion globale

Participants	Predictors	Outcome	Analysis
?	-	?	-

References

2021. URL `https://en.wikipedia.org/wiki/COVID-19_pandemic_on_Diamond_Princess`. [Online; accessed 27/03/2021].

Kenji Mizumoto, Katsushi Kagaya, Alexander Zarebski, and Gerardo Chowell. Estimating the asymptomatic proportion of coronavirus disease 2019 (covid-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020. *Eurosurveillance*, 25(10):2000180, 2020.

Mohammad Pourhomayoun and Mahdi Shakibi. Predicting mortality risk in patients with covid-19 using artificial intelligence to help medical decision-making. *MedRxiv*, 2020.

Robert F Wolff, Karel GM Moons, Richard D Riley, Penny F Whiting, Marie Westwood, Gary S Collins, Johannes B Reitsma, Jos Kleijnen, and Sue Mallett. Probast: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of internal medicine*, 170(1):51–58, 2019.