

CoviSas

Impact de la pandémie de Covid-19 sur les
patients atteints d'apnée du sommeil

Conflits d'intérêt

Cette intervention est faite en toute indépendance vis-à-vis de l'organisateur de la manifestation.

Je n'ai pas de conflit d'intérêts en lien avec le sujet traité.

Le projet CoviSas est porté par la chaire Santé de l'institut MIAI et l'entreprise Sêmeia



L'ambition de la chaire Santé de l'institut MIAI est d'implémenter des algorithmes d'intelligence artificielle pour développer la « Médecine de Trajectoire » avec un focus sur le syndrome d'apnée du sommeil (SAS)



Sêmeia est une entreprise proposant des outils de télésuivi des patients à destinations des professionnels de santé s'appuyant sur des modèles d'IA pour anticiper le parcours de patient et optimiser les interventions

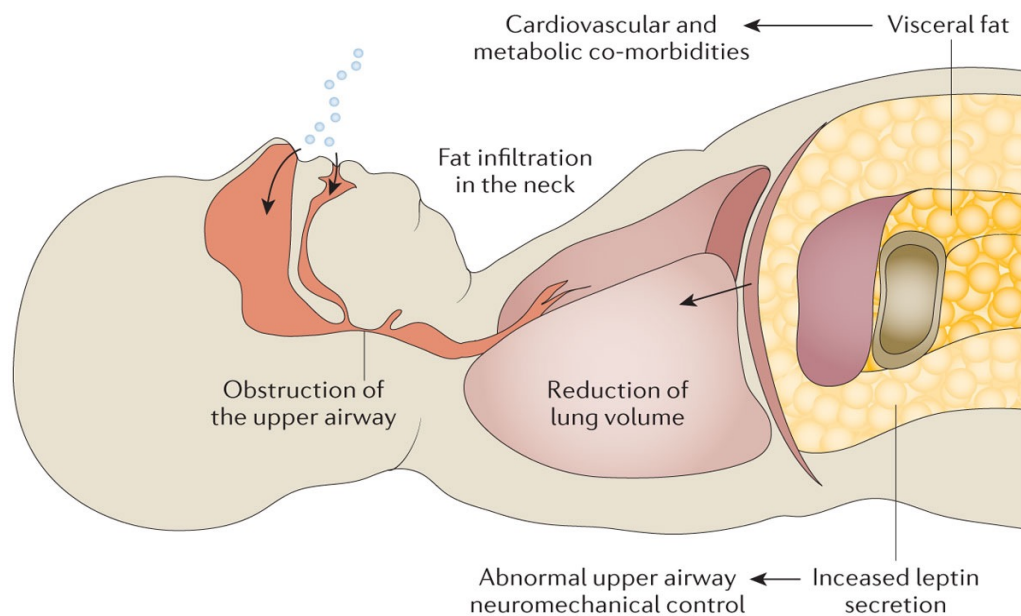
L'objectif du projet est d'identifier le sur-risque Covid des patients souffrant d'apnée du sommeil

Les comorbidités cardiovasculaires associées à un sur-risque de gravité ou de décès chez les patients Covid **sont fréquemment associées à un SAS**

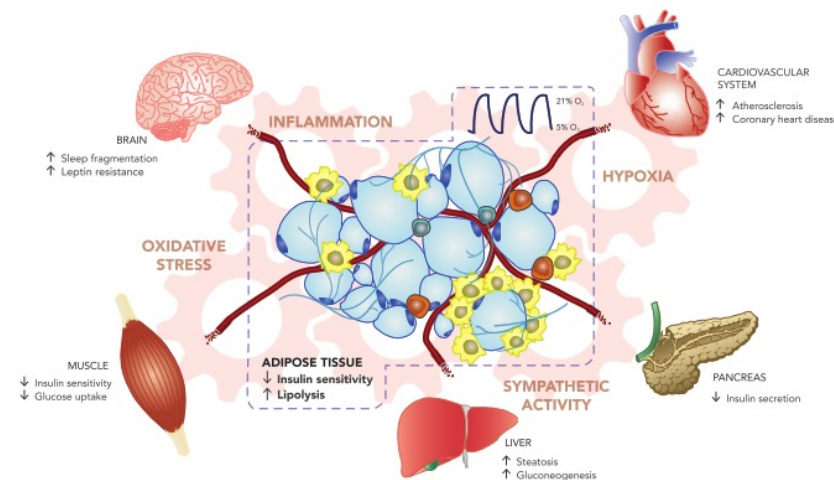
Notre principale question est d'identifier le **rôle spécifique du SAS** dans la survenue des formes graves du Covid-19 ainsi que **les principales caractéristiques discriminantes** de ces patients

Notre hypothèse est qu'une combinaison de maladies cardio-métaboliques préexistantes incluant le SAOS est déjà associée à une production élevée de cytokines à l'état basal qui en présence de l'infection COVID-19 initiera et amplifiera l'orage cytokinique entraînant à son tour une insuffisance respiratoire aiguë avec une défaillance multi organes

Le SAS : une maladie associées à de nombreuses comorbidités...



Nature Reviews | **Disease Primers**



Lévy, P, Kohler M, McNicholas W, Barbé F, McEvoy D, Somers VK, Lavie L and Pépin JL
Nat Rev Dis Primers 2015

Plus d'1 million de patients diagnostiqués et traités en France

Le sur-risque des patients SAS est significatif selon la littérature

Redline S

AJRCCM 2020;1462-1463

Outcome	Reference	Model 1 (Unadjusted)	Model 2 (Adjusted for Demographics)	Model 3 (Model 2 Plus WHO BMI Class)	Model 4 (Model 3 Plus Comorbidities)
Death	Alive	1.79 (1.31–2.45)*	1.53 (1.08–2.15) [†]	1.39 (0.97–1.98) [‡]	1.16 (0.8–1.68)
Death, mechanical ventilation, or ICU admission	All others	1.64 (1.32–2.04)*	1.29 (1.03–1.62) [†]	1.11 (0.87–1.40)	1.04 (0.81–1.34)
Inpatient admission	Outpatient	1.55 (1.27–1.89)*	1.18 (0.95–1.45)	1.01 (0.81–1.26)	0.91 (0.72–1.14)

Definition of abbreviations: BMI = body mass index; COVID-19 = coronavirus disease; WHO = World Health Organization.

Results are presented as odds ratios with 95% confidence intervals. Model 1 = unadjusted (base); model 2 = adjusted for race and ethnicity, age, sex (demographics); model 3 = model 2 plus BMI (demographics + BMI); model 4 = model 3 plus asthma, chronic bronchitis, emphysema, essential hypertension, BMI, and type 2 diabetes (fully adjusted).

* $P \leq 0.001$ (Wald's test).

[†] $P \leq 0.05$.

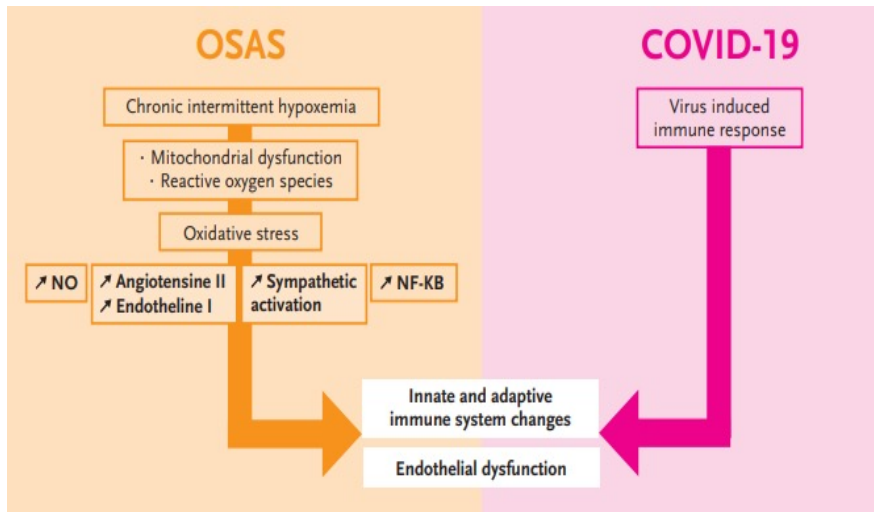
[‡] $P = 0.07$.

Etude sur une base observationnelle de 4700 patients

L'utilisation des données massives du SNDS peut aider à identifier les mécanismes sous-jacents de ce sur-risque

Breville G

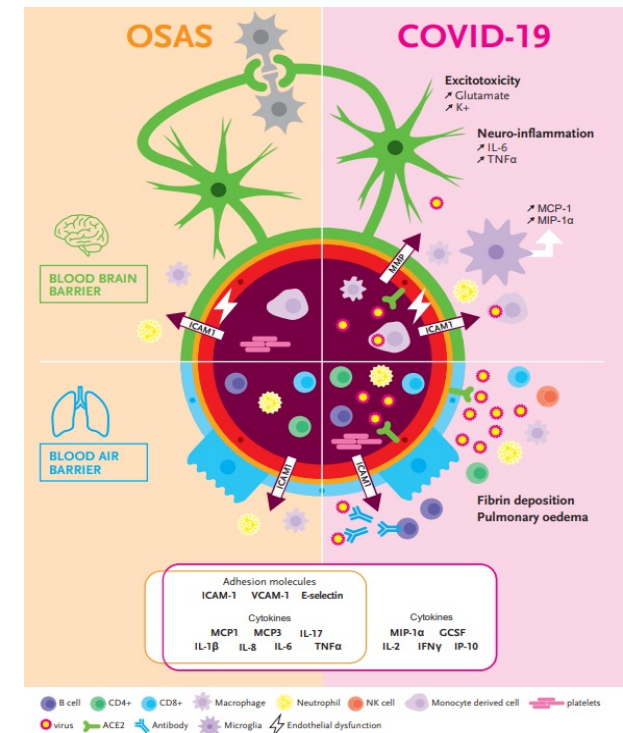
Chest 2021; in press



Système national
des données de
Santé



66 M patients



L'utilisation des données massives du SNDS peut aider à identifier les mécanismes sous-jacents de ce sur-risque

Données de l'Assurance Maladie
Données de ville
(base SNIIRAM)

Données des hôpitaux
(base PMSI)

Données statistiques relatives aux
causes de décès (Cepi-DC).



Données patients disponibles

- Âge, mois et année de naissance
- Genre
- Diagnostic ALD
- Diagnostic PMSI
- Comorbidités
- Date et cause de décès
- CMU, ACS



Données de parcours disponibles

- Toutes les données de remboursements avec dates de réalisation/dispensation et spécialités prescriptrices
- Consultations de médecins
- Médicaments remboursés
- Actes médicaux
- Tests de biologie
- Dispositifs médicaux
- Actes paramédicaux
- Hospitalisations et durées de séjour
- GHM/GHS
- Consultations externes

Avantages

- ✓ Données sans biais
- ✓ Données exhaustives
- ✓ Données fiables

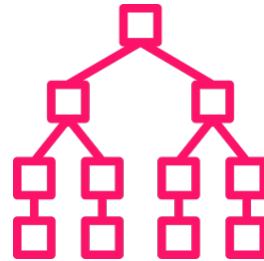
Inconvénients

- ✓ Données administratives sans but clinique
- ✓ Données d'intérêt manquantes (IMC)

Trois principaux enjeux méthodologiques



Identifier la population cible à partir de données médico-administratives



Gérer des données « sparse »
[beaucoup de modalités /
peu de fréquence]



Analyser efficacement un grand nombre de trajectoires différentes

Une étude préalable approfondie a été menée pour effectuer le ciblage des patients

L'identification des **patients SAS** a été réalisée à partir :

- ✓ Des actes de diagnostic (actes de polysomnographie)
- ✓ Des consommations de soins (orthèse d'avancée mandibulaire, dispositif de pression positive continue)
- ✓ Des actes thérapeutiques (pharyngoplastie, amygdalectomie...)

L'identification des **patients Covid hospitalisés** a été réalisée à partir :

- ✓ Des séjours hospitaliers avec un diagnostic Covid +

Les variables d'intérêt ont aussi fait l'objet d'algorithmes de construction

La **sévérité des patients Covid+** a été construite à partir :

- ✓ Des passages en soins intensifs et/ ou réanimation lors du séjour Covid+

Les **comorbidités** ont été détectées grâce :

- ✓ Aux diagnostics posés lors des séjours hospitaliers sur les 6 dernières années
- ✓ Aux diagnostics liés aux affections longue durée des patients
- ✓ Aux consommations de soins (traitements ou actes traceurs) sur les 6 dernières années

Le traitement des comorbidités présente une difficulté méthodologique majeure

Il existe de **nombreuses modalités différentes** de pathologie (14 400 codes CIM-10) et la plupart d'entre elles ne concerne qu'un **nombre restreint de patients**

Pour les pathologies peu fréquentes, il est très difficile d'obtenir **la puissance statistique suffisante** pour identifier un impact « toutes choses égales par ailleurs »

L'utilisation des nomenclatures est peu satisfaisante

Une première approche du problème est d'utiliser les **regroupements existants** de la nomenclature



Ces regroupements ne prennent en compte qu'une seule **dimension** de la pathologie

Ex : la greffe de rein est dans le regroupement « Greffe d'organe »



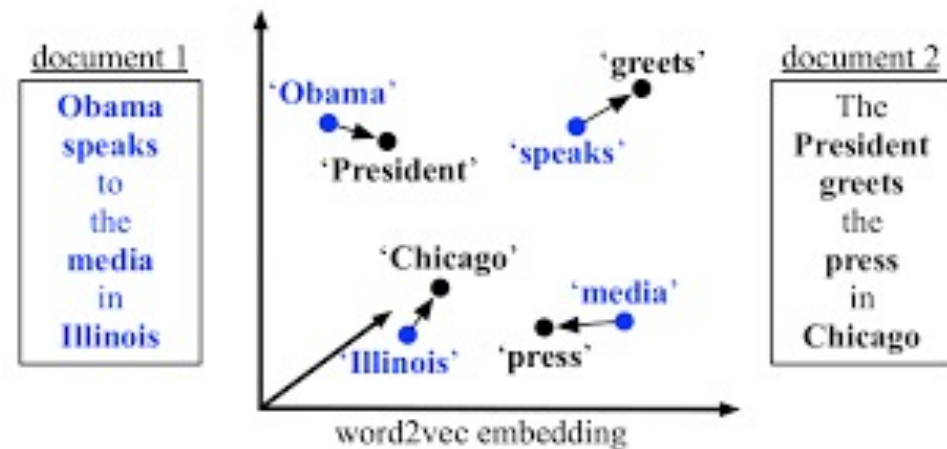
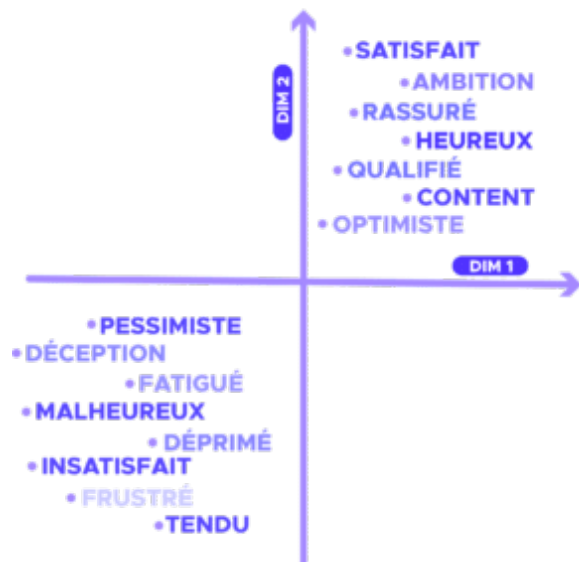
Il n'y a pas de **concept de similarité** entre les pathologies

Ex : le diabète est aussi « proche » de l'hypertension que du cancer de la prostate

Dans le cadre du projet CoviSas, il est fait appel aux techniques « d'embedding »

L'embedding est une technique de réduction de dimension de variables catégorielles en s'appuyant sur les similarités entre des concepts

Il a été initialement développé pour **l'analyse du langage**



Ces techniques seront transposées au domaine médical

Nous avons choisi les algorithmes « GloVe » et « Gram »

Algorithme « GloVe »

Cet algorithme utilisera les co-occurrences de pathologies pour estimer des « distances » de similarité entre celles-ci

Algorithme « GRAM »

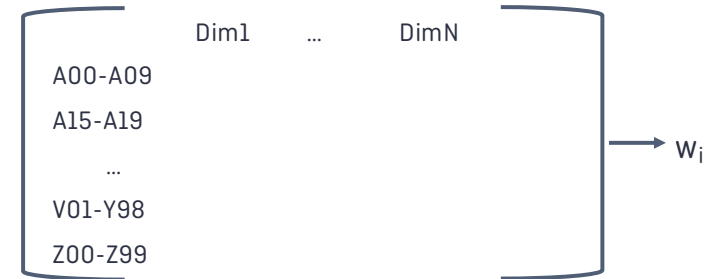
L'algorithme « Gram » nous permettra de consolider la représentation de chaque pathologie en s'appuyant sur la classification déjà existante (CIM-10)

In fine, nous obtenons, pour chaque pathologies, une projection (une donnée numérique) sur une dizaine d'axes, nous permettant ainsi d'avoir des données numériques en entrée du modèle

Cette projection est adaptée au contexte propre des données et pourra être « visualisée »

GloVe – modalités d'implémentation

L'objectif est d'avoir une matrice de projection de l'ensemble des 14 000 codes CIM-10 sur un nombre fixe de dimensions



Etape 1 : calcul de probabilité de co-occurrence des pathologies

Etape 2 : initialisation aléatoire de la matrice de projection

Etape 3 : Entraînement dans un Multi-layer Perceptron

La fonction de perte à optimiser est la distance entre le produit scalaire de la représentation de deux codes CIM-10 et le logarithme de la probabilité de co-occurrence

$$J = \sum_{i,j=1}^V f(X_{ij}) (w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij})^2$$

Afin de prendre en compte également le nombre d'occurrences, un poids spécifique est affecté à chaque ratio de co-occurrence

$$f(x) = \begin{cases} (x/x_{\max})^{100} & \text{if } x < x_{\max} \\ 1 & \text{otherwise} \end{cases}$$

GRAM – modalités d'implémentation

L'utilisation d'un modèle Glove nous donne une matrice de projection pour chacun des codes CIM-10 existant

L'algorithme GRAM doit nous permettre d'intégrer le concept de hiérarchie existante

➔ Au lieu d'avoir un poids pour C50 (cancer du sein), C50-C58 (Cancers seins et organes génitaux de la femme) et C00-D48 (Tumeurs), le poids de C50 dans chaque dimension **sera une combinaison des poids** de C50, C50-C58 et C00-D48

Plus le code sera rare, **plus le poids du code « parent » sera important**

L'estimation optimale de ces poids sera faite en même temps que le modèle à travers un Multi-Layer Perceptron

GRAM – schéma d'intégration au modèle

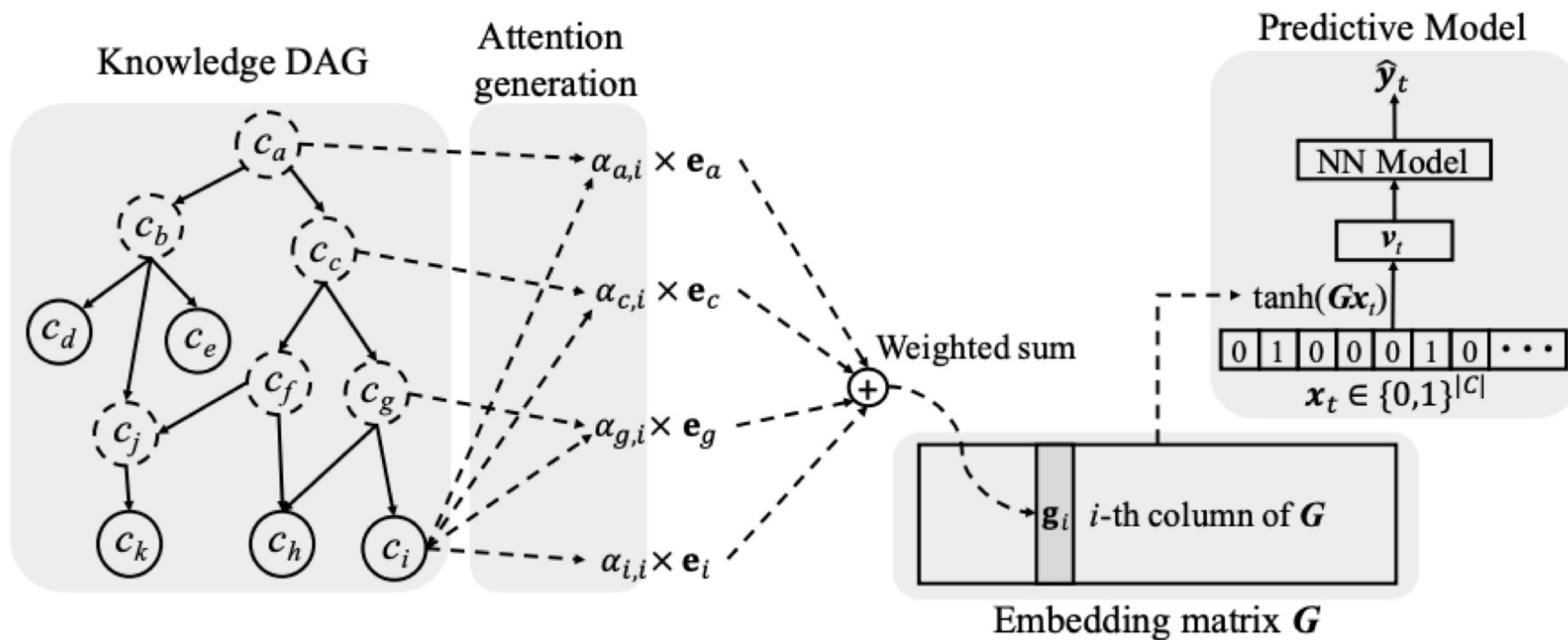
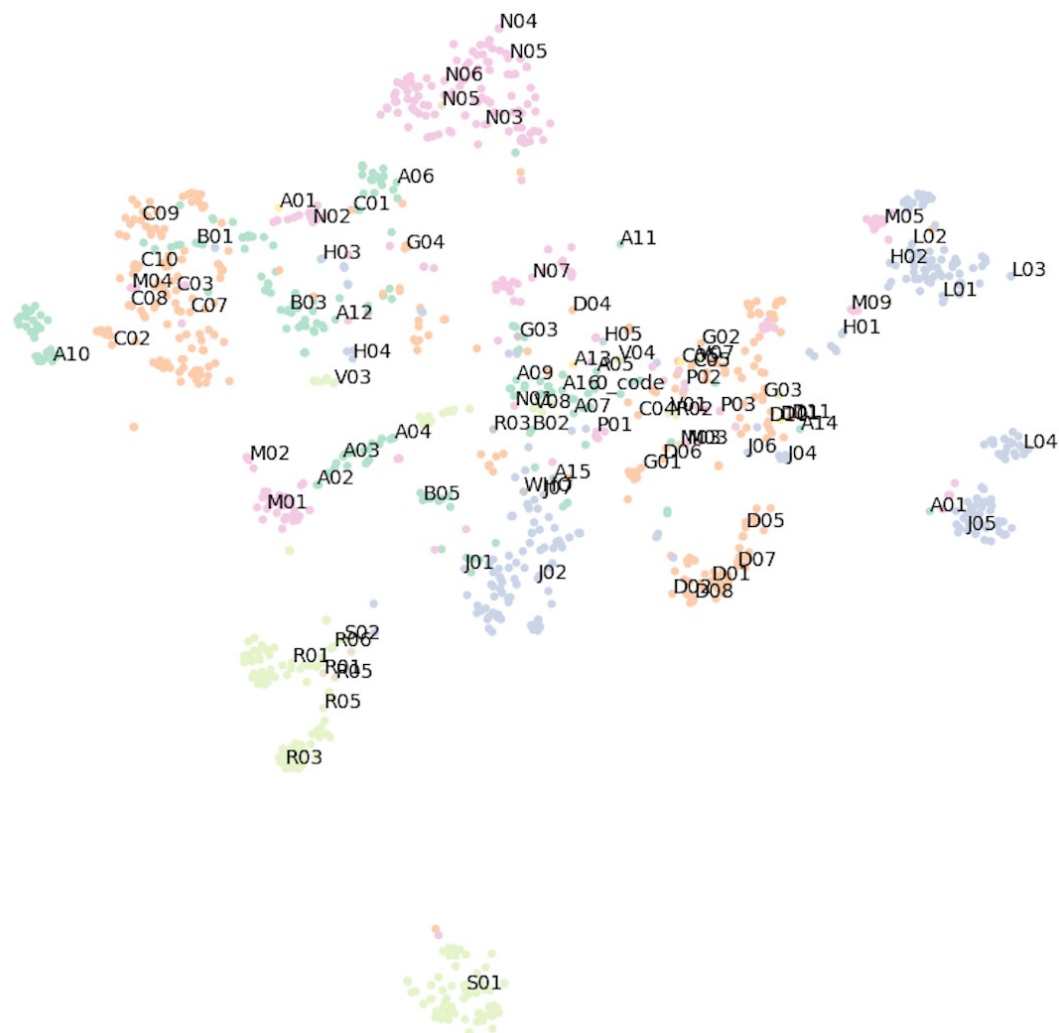


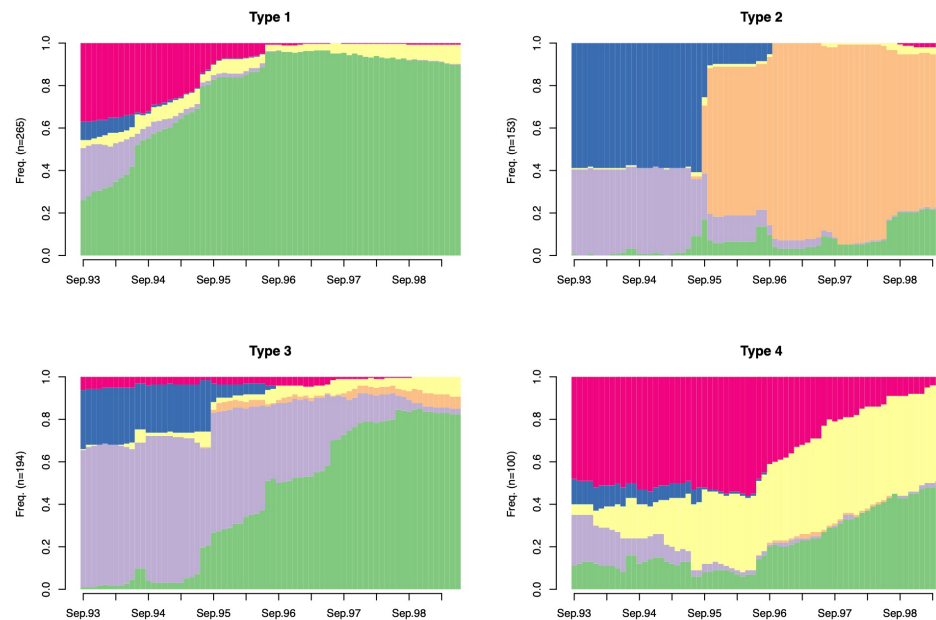
Illustration sur deux dimensions



Une autre difficulté est la gestion des « trajectoires »

La sévérité des patients hospitalisés Covid+ doit être estimée à partir d'un grand nombre de trajectoires possibles (hospitalisation conventionnelle / Surveillance continue / Soins Intensifs / Réanimation Décès)

→ Un regroupement des **différentes trajectoires en « clusters »** apparaît nécessaire pour estimer le sur-risque de sévérité des patients SAS



La technique de clustering utilisée sera « l'optimal matching »



L'objectif est d'obtenir une distance (mesure de similarité) entre deux trajectoires



Cette distance est obtenue en comptant le nombre d'ajouts, suppression ou modification pour faire correspondre deux trajectoires (un « coût » différent peut être attribuer selon le type de modification)



Une fois ces distances deux à deux obtenus, des algorithmes classiques de clustering seront utilisés

Perspectives

Les résultats de notre étude ne sont pas encore disponibles, les données sont en attente de réception

L'identification des différentes « dimensions » de pathologies pertinentes au regard de la sévérité et des trajectoires des patients Covid-19 doit être **confrontée à l'expertise des médecins expert** et leur permettre **de formuler des hypothèses** et des pistes de recherche pour mieux comprendre les mécanismes sous-jacents

L'IA est utilisée ici pour **attirer l'attention des experts** sur des corrélations / combinaisons qui n'auraient pas été identifiées à partir des cas cliniques