**creation\_table\_ATC:**

*Entrée :*

*- fichiers : 2023\_03\_23\_ESADA\_CLEAN\_EXPORT\_V0\_TO\_5\_WITH\_SCORES.xlsx et codes\_ATC.xlsx*

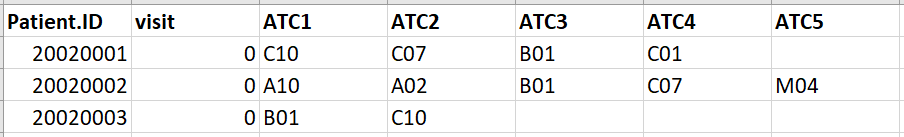
*Note : au lieu d’ouvrir le fichier excel directement, j’ouvre l’environnement bd\_esada\_initial.RData qui correspond à l’excel pour que ce soit plus rapide*

*Sortie :*

*- fichier : ATC\_patients.xlsx*

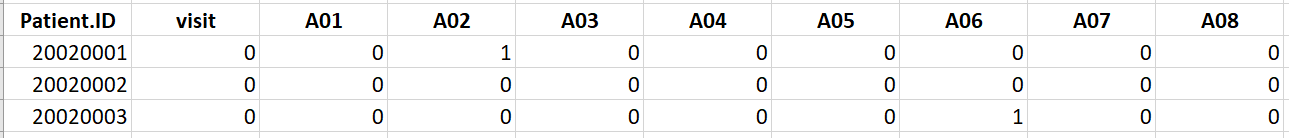
*- tableaux avec pour chaque type de médicament le pourcentage de patients qui en ont, et la description du type de médicament (en anglais ou en français).*

Dans la base de données initiale on a pour chaque patient les noms des médicaments qu’il prend dans le format ci-dessous :



Pour chaque visite de chaque patient, la liste des médicaments qu’il prend est écrit dans les cases ‘ATC’, avec un médicament par case.

Le but de ce ficher est de créer à la place une variable binaire pour chaque type de médicament pour indiquer si le patient le prend ou non, dans le format ci-dessous :



Le fichier de sortie comprend 4 onglets :

- first\_level\_data : pour chaque patient pour chaque visite, une variable binaire pour chaque groupe principal de médicament (A, B, C, …. V)

- first\_level\_classification : Correspondance entre la lettre de la catégorie et la description de la catégorie en français et en anglais

- second\_level\_data : pour chaque patient pour chaque visite, une variable binaire pour chaque sous-groupe de médicament

- second\_level\_classification : correspondance entre le code et sa description en anglais et en français

**CRP\_baseline.R :**

*Entrée :*

*- fichiers : 2023\_03\_23\_ESADA\_CLEAN\_EXPORT\_V0\_TO\_5\_WITH\_SCORES.xlsx et ATC\_patients.xlsx*

*Sortie :*

*- fichier : final\_dataframe\_baseline.xlsx*

*- tableaux et images : tous les tableaux de description et graphs de l’analyse. Les tableaux peuvent être exportés en word ou en latex*

La variable ‘Cause.Of.Fatality’ ne se charge pas correctement dans R car il y a trop de NA au début, ce qui empêche à R de déterminer le type de la variable. J’ai donc rajouté un onglet dans le ficher excel de la base de données avec l’ID du patient, le numéro de visite et la variable cause of fatality pour l’importer à part du reste des données, en précisant le type de la variable.

Ce fichier permet de réaliser l’étude de la CRP à baseline. Il contient les différentes étapes :

- Sélection de la population

- création de nouvelles variables

- description des variables avant data-management

- data-management

- création d’un dataframe final : dataframe avec uniquement les variables d’intérêt, et des labels pour que les variables soient renommées dans les tableaux. Export de ce dataframe (fichier final\_dataframe\_baseline.xlsx)

- description après data-management : comparaison des variables avant et après imputation, et description générale par groupe de sévérité d’apnée du sommeil avec des tests de comparaison entre les groupes

- graphiques

- analyse univariée : impact de chaque variable séparément sur la CRP, utilisé pour la sélection des variables

- Création d’un DAG = graphique de représentation des liens connus entre les variables

- Calcul des poids : calcul avec la méthode IPW et les variables sélectionnées avec l’analyse univariée. Ensuite, étude des poids obtenus pour vérifier si les hypothèses sont vérifiées avec la description des poids par groupe, la description des données après l’application des poids au dataframe et un graphique représentant les SMD (différences moyennes standardisées)

- Modèle : modèle mixte avec toutes les variables sélectionnées et application des poids. Etude des résidus et des effets aléatoires

- Bootstrap : recalcul des intervalles de confiance du modèle avec 1000 bootstrap, puis graph de ces coefficients.

**CRP\_baseline\_subgroup :**

*Entrée :*

*- fichier : final\_dataframe\_baseline.xlsx*

*Sortie :*

*- tableaux et graphs des analyses*

Ce fichier a pour but d’automatiser l’analyse précédente et de la réaliser pour différents sous-groupes.

L’analyse ne peut pas être totalement automatisée car on ne peut pas implémenter dynamiquement la formule utilisée pour le calcul des poids. Ainsi, pour chaque analyse en sous-groupe que l’on veut faire il faut créer un dataframe approprié (en sélectionnant que les hommes par exemple), et une fonction qui permet de créer les poids associés.

On lance ensuite la fonction ‘subgroup\_analysis’, en mettant en argument le dataframe, la fonction pour calculer les poids, et les variables à enlever du modèle s’il y en a (par exemple pour une analyse de sous-groupe sur le sexe il faut enlever la variable sexe du modèle car dans le dataframe il n’y a que des hommes ou que des femmes).

Les analyses en sous-groupes ont été effectuées pour le sexe, COPD (bronchopneumopathie chronique obstructive) et le type de diagnostic (Polygraphie/Polysomnographie)

**CRP\_delta :**

*Entrée : - fichiers : 2023\_03\_23\_ESADA\_CLEAN\_EXPORT\_V0\_TO\_5\_WITH\_SCORES.xlsx*

*Sortie : - tableaux et graphs des analyses*

Comme pour l’étude à baseline, la variable Cause.Of.Fatality doit être chargée différemment du reste de la base de données dans R pour spécifier son type. Deux autres variables sont dans le même cas : Use.h.Day et PAP.Treatment.start. Ces deux variables ont donc aussi été rajoutées dans des onglets du fichier excel 2023\_03\_23\_ESADA\_CLEAN\_EXPORT\_V0\_TO\_5\_WITH\_SCORES.xlsx.

Ce fichier commence par la sélection de patients, créations de nouvelles variables et data-management. Deux méthodes différentes sont ensuite implémentées pour étudier l’impact du traitement PPC sur les niveaux de CRP des patients :

1) dataframe avec une ligne par patient : Création d’une variable delta CRP (CRP visite 1 – CRP visite 0), delta\_weight (poids visite 1 – poids visite 0). Toutes les autres variables baseline sont prises, et on rajoute l’observance à la PPC et le temps de traitement (informations de la visite 1). Le modèle est un modèle mixte qui explique la variable delta\_CRP en fonction des autres, avec le site et l’année de la visite en effet aléatoire

2) dataframe avec deux lignes par patients : on conserve deux lignes par patients, une pour la visite baseline et une pour la visite follow-up. Le modèle utilisé est aussi un modèle mixte mais qui explique la CRP en fonction des autres variables, avec le site, l’année de la visite ET l’ID de patient en effet aléatoire.