



**HAL**  
open science

# Phénotypes et trajectoires des patients BPCO, données issues de la cohorte PALOMB

El Hassane Ouaalaya

► **To cite this version:**

El Hassane Ouaalaya. Phénotypes et trajectoires des patients BPCO, données issues de la cohorte PALOMB. Médecine humaine et pathologie. Université de Bordeaux, 2020. Français. NNT : 2020BORD0295 . tel-03180906

**HAL Id: tel-03180906**

**<https://theses.hal.science/tel-03180906>**

Submitted on 25 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Thèse présentée pour obtenir le grade de

# **DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX**

Ecole doctorale Sociétés, Politique et Santé Publique  
Spécialité : Santé publique Option : Biostatistique  
Thèse préparée dans le cadre du Projet PALOMB

Par **El Hassane OUAALAYA**

## **Phénotypes et trajectoires des patients BPCO, données issues de la cohorte PALOMB**

**Phenotypes and trajectories of COPD patients, PALOMB cohort Data**

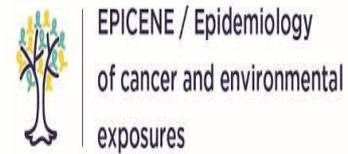
Présentée et soutenue publiquement  
Le 16 12 2020

### **Membres du Jury**

Monsieur Le Professeur Mathieu MOLIMARD, PU-PH  
Monsieur Le Professeur Nicolas ROCHE, PU-PH  
Monsieur Le Professeur Christophe PISON, PU-PH  
Madame Le Professeur Isabelle BALDI, PU-PH  
Monsieur Le Professeur Pierre-Régis Burgel, PU-PH  
Madame Le Professeur Chantal RAHERISON-SEMJEN, PU-PH  
Le Conseil scientifique du Projet PALOMB

Président du Jury  
Rapporteur  
Rapporteur  
Examinatrice  
Examinateur  
Directrice de thèse  
Invités

## Laboratoire d'accueil



Université De Bordeaux

ISPED (Institut de santé publique, d'épidémiologie et de développement)

INSERM UMR1219, BPH (Bordeaux Population Health Research Center)

Équipe EPICENE (Épidémiologie des cancers et expositions environnementales)

146 rue Léo Saignat

33076 BORDEAUX Cedex

Tél : 05 57 57 12 34

Fax : 05 57 57 47 33

<http://www.bordeaux-population-health.center>

## Financements

Ce travail a été possible grâce au soutien financier de la Fondation Bordeaux Université dans le cadre d'un mécénat. Initiée en 2013, le projet Palomb s'est concrétisé grâce au soutien de Novartis et d'ISIS Medical. Après plus de 2 000 patients inclus et la diffusion de premiers résultats très concluants, la volonté de poursuivre le projet a été unanime sur les années 2017, 2018, 2019 et 2020. GSK, Boehringer Ingelheim et Chiesi ont alors rejoint Isis Médical et Novartis qui ont eux-mêmes renouvelé leur soutien ; un grand merci à eux pour leur engagement aux côtés des pneumologues et autres acteurs engagés en faveur d'une meilleure prise en charge des patients BPCO.

Suite aux dons qu'ISIS Aquitaine, GSK, Novartis, Boehringer Ingelheim et Chiesi ont réalisés au bénéfice de la Fondation Bordeaux Université avec pour charge afférente de soutenir le développement du projet PALOMB et conformément à la décision du comité de pilotage du projet, une convention a été mise en place afin d'établir les modalités financière et comptable du soutien au projet, effectué sous forme d'une subvention à l'université, afin de permettre à l'équipe EPICENE, de poursuivre la bonne réalisation du Projet. Ma thèse a été financée durant les trois années suite à cette convention de reversement.



# Valorisation scientifique

## Articles en premier auteur dans le cadre du travail de thèse :

### ➤ **Publié**

- [1] Chantal Raherison, **El-Hassane Ouaalaya**, Alain Bernady, Julien Casteigt, Cecilia Nocent-Eijnani, Laurent Falque, Frédéric Le Guillou, Laurent Nguyen, Annaig Ozier and Mathieu Molimard. Comorbidities and COPD severity in a clinic-based cohort. BMC Pulm Med. 2018 Jul 16; 18(1) :117. DOI: <https://doi.org/10.1186/s12890-018-0684-7>
- [2] **El Hassane Ouaalaya**, Laurent Falque, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Ejnaini, Frédéric Le Guillou, Mathieu Molimard, Maéva Zysman, Chantal Raherison-Semjen. Susceptibility to frequent exacerbation in COPD patients: Impact of the exacerbations history, vaccinations and comorbidities? Respir Med. 2020 May 11 ; 169 : 106018. DOI: <https://doi.org/10.1016/j.rmed.2020.106018>
- [3] **El Hassane Ouaalaya**, Laurent Falque, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Ejnaini, Frédéric Le Guillou, Mathieu Molimard, Chantal Raherison-Semjen. The determinants of dyspnoea evaluated by the mMRC scale: The French Palomb cohort. Respir Med Res. 2020 Nov 22; 79: 100803. DOI: <https://doi.org/10.1016/j.resmer.2020.100803>

### ➤ **En cours de publication**

- [1] **El Hassane Ouaalaya**, Maéva Zysman, Laurent Falque, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Ejnaini, Frédéric Le Guillou, Mathieu Molimard, Chantal Raherison-Semjen. Main determinants associated with mortality in COPD patients. Plos One.

## Communications orales dans le cadre de la thèse :

- [1] **El-Hassane Ouaalaya**, Laurent Falque, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Ejnaini, Frédéric Le Guillou, Mathieu Molimard, Chantal Raherison-Semjen ; Quels sont les déterminants de la dyspnée chez les patients BPCO de la cohorte PALOMB ? [Présentation orale aux journées de l'EDSP2 ; PUSG Bordeaux Bastide ; 16 mai 2018. Sélection via résumé]
- [2] Chantal Raherison-Semjen, **El-Hassane Ouaalaya**, Alain Bernady, Julien Casteigt, Cécilia Nocent-Eijnani, Laurent Falque, Frédéric Le Guillou, Laurent Nguyen, Annaig Ozier and Mathieu Molimard. Caractéristiques des femmes atteintes de BPCO : données de la cohorte PALOMB. Revue des Maladies Respiratoires. Volume 36, Supplement, January 2019, Page A68. DOI: <https://doi.org/10.1016/j.rmr.2018.10.129> [CPLF - Marseille 2019 (23e édition du Congrès de Pneumologie de Langue Française) : sélection via résumé]
- [3] Claire Bon, **El-Hassane Ouaalaya**, Chantal Raherison-Semjen. Identifier les facteurs associés à la survie chez les patients BPCO sous VNI au long cours avec un recul à 5 ans. Revue des Maladies Respiratoires. Volume 36, Supplement, January 2019, Pages A71-A72. DOI : <https://doi.org/10.1016/j.rmr.2018.10.136> [CPLF - Marseille 2019 (23e édition du Congrès de Pneumologie de Langue Française) : sélection via résumé]
- [4] **E. H. Ouaalaya**, A. Bernady, J. Casteigt, J. M. Dupis, C. Nocent-Eijnaini, M. Sabatini, L. Falque, F. Le Guillou, L. Nguyen, A. Ozier, M. Molimard, C. Raherison-Semjen. Le cancer et les comorbidités cardiovasculaires sont associés à la mortalité chez les patients BPCO. Revue des Maladies Respiratoires Actualités. Volume 12, numéro 1, janvier 2020, pages 41-42. DOI :

<https://doi.org/10.1016/j.rmra.2019.11.070> [CPLF - Paris 2020 (24e édition du Congrès de Pneumologie de Langue Française) : sélection via résumé]

- [5] **E. H. Ouaalaya**, A. Bernady, J. Casteigt, J. M. Dupis, C. Nocent-Eijnaini, M. Sabatini, L. Falque, F. Le Guillou, L. Nguyen, A. Ozier, M. Molimard, C. Raheison-Semjen. Susceptibilité aux exacerbations fréquentes chez les patients BPCO : impact des vaccinations et des comorbidités ? Revue des Maladies Respiratoires Actualités. Volume 12, numéro 1, janvier 2020, page 24. DOI: <https://doi.org/10.1016/j.rmra.2019.11.036> [CPLF - Paris 2020 (24e édition du Congrès de Pneumologie de Langue Française) : sélection via résumé]

### **Posters dans le cadre de la thèse :**

- [1] **E.H. Ouaalaya**, L. Falque, J.M. Dupis, M. Sabatini, A. Bernady, L. Nguyen, A. Ozier, C. Nocent-Eijnaini, F. Le Guillou, M. Molimard, C. Raheison-semjen. Cancer and Cardiovascular Comorbidities Are Associated with COPD Mortality: The French PALOMB Cohort. American Thoracic Society 2020 International Conference -Philadelphia- USA. Conference Paper, May 15-20, 2020. DOI: [https://doi.org/10.1164/ajrccm-conference.2020.201.1\\_MeetingAbstracts.A7162](https://doi.org/10.1164/ajrccm-conference.2020.201.1_MeetingAbstracts.A7162)
- [2] **El-hassane Ouaalaya**, Maéva Zysman, Émilie Berteaud, Laurent Falque, Laurent Nguyen, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Annaig Ozier, Cécilia Nocent-Eijnaini, Christophe Roy, Frédéric Le Guillou, Emmanuel Monge, Mohammed Aliati, Anne Prudhomme, Laura Petrov, Marie Line Quinquenel, Mohammed Staali, Frédéric Pilard, Marc Sapène, Julien Casteigt, Jean Moinard, Yannick Daoudi, Esther Iglesias, Élodie Blanchard, Julie Macey, Rémi Veillon, Xavier Demant, Claire Bon, Leo Grassion, Mathieu Molimard, Chantal Raheison-Semjen. COPD phenotypes and 5-years mortality risk prediction: PALOMB Cohort. European Respiratory Journal 2020; [The virtual ERS International Congress 7-9 September, 2020]
- [3] Nicolas Roche, Bernard Aguilaniu, Pierre-Régis Burgel, Maéva Zysman, David Hess, **El-Hassane Ouaalaya**, Thi-Chien Tran, Chantal Raheison. Influenza and pneumococcal vaccination in patients with COPD. European Respiratory Journal 2020 ; [The virtual ERS International Congress 7-9 September, 2020]
- [4] **El-hassane Ouaalaya**, Laurent Falque, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Eijnaini, Frédéric Le Guillou, Mathieu Molimard, Chantal Raheison-Semjen. Differences between men and women with COPD: The French Palomb Cohort. European Respiratory Journal (November 21, 2019) vol. 54 no. suppl 63; DOI: <https://doi.org/10.1183/13993003.congress-2019.PA4424> ERS International Congress - Madrid- Spain.
- [5] **El-hassane Ouaalaya**, Laurent Falque, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Eijnaini, Frédéric Le Guillou, Mathieu Molimard, Chantal Raheison-Semjen. Characterization of the “frequent exacerbator phenotype” over a period of 3 years: Palomb cohort. European Respiratory Journal (November 21, 2019) vol. 54 no. suppl 63; DOI: <https://doi.org/10.1183/13993003.congress-2019.PA4432> ERS International Congress -Madrid- Spain.
- [6] **E.H. Ouaalaya**, L. Falque, J.M. Dupis, M. Sabatini, C. Nocent-Eijnaini, A. Bernady, L. Nguyen, A. Ozier, F. Le Guillou, M. Molimard, C. Raheison-Semjen. The Resting Pulmonary Hyperinflation: A Determinant of the MMRC Dyspnea Scale in COPD Patients in a Real-Life Cohort? American Thoracic Society 2019 International Conference -Dallas- USA. Conference Paper, May 21, 2019. DOI: [https://doi.org/10.1164/ajrccm-conference.2019.199.1\\_MeetingAbstracts.A4888](https://doi.org/10.1164/ajrccm-conference.2019.199.1_MeetingAbstracts.A4888)

- [7] C. Raheison, **E. Ouaalaya**, A. Bernady, J. Casteigt, J. M. Dupis, C. Nocent-Ejnaini, M. Sabatini, L. Falque, F. Le Guillou, L. Nguyen, A. Ozier, M. Molimard. Quels sont les déterminants de la dyspnée chez les patients BPCO de la cohorte PALOMB ? Revue des Maladies Respiratoires. Volume 35, Supplement, January 2018, Pages A10-A11. DOI: <https://doi.org/10.1016/j.rmr.2017.10.024> [Présentation orale au CPLF - Lyon 2018 (22e édition du Congrès de Pneumologie de Langue Française) : sélection via résumé]
- [8] **El-hassane Ouaalaya**, Emilie Berteaud, Alain Bernardy, Jean-Michel Dupis, Marielle Sabatini, Laurent Falque, Frédéric Le Guillou, Cecilia Nocent-Ejnaini, Laurent Nguyen, Annaig Ozier, Julien Casteigt, Mathieu Molimard, Chantal Raheison-Semjen. Determinants of dyspnea in COPD patients: The French PALOMB cohort. European Respiratory Journal, November 19, 2018. vol. 52 no. Suppl 62; DOI: <https://doi.org/10.1183/13993003.congress-2018.PA757> ERS International Congress -Paris- France.
- [9] C. Raheison-Semjen, **H. Ouaalaya**, A. Bernady, J. Casteigt, C. Nocent-Eijnani, L. Falque, F. Le-Guillou, L. Nguyen, J. Moinard, A. Ozier, C. Roy, E. Berteaud, M. Molimard, F. Pellet. Profils de comorbidités et sévérité de la BPCO : données de la cohorte PALOMB. Revue des Maladies Respiratoires ; Volume 34, Supplement, January 2017, Pages A4-A5. DOI: <https://doi.org/10.1016/j.rmr.2016.10.014> [CPLF - Marseille 2017 (21e édition du Congrès de Pneumologie de Langue Française) : sélection via résumé]
- [10] **El-Hassane Ouaalaya**, Laurent Falque, Marielle Sabatini, Jean-Michel Dupis, Alain Bernady, Laurent Nguyen, Cecilia Nocent-Eijnaini, Frederic Le Guillou, Christophe Roy, Elodie Blanchard, Jean Moinard, Julien Casteigt, Emilie Berteaud, Mathieu Molimard, François Pellet, Chantal Raheison-Semjen, Annaig Ozier. COPD severity using GOLD 2017 in real life: The French Palomb Cohort. European Respiratory Journal. December 6, 2017, vol. 50 no. suppl 61 ; DOI: <https://doi.org/10.1183/1393003.congress-2017.PA1210> ERS International Congress -Milan- Italy.
- [11] Chantal Raheison-Semjen, **El-Hassane Ouaalaya**, Laurent Falque, Alain Bernady, Jean-Michel Dupis, Marielle Sabatini, Laurent Nguyen, Annaig Ozier, Cecilia Nocent-Eijnaini, Frederic Le Guillou, Christophe Roy, Elodie Blanchard, Helene Jungmann, Julien Casteigt, Emilie Berteaud, François Pellet. Comorbidities and COPD severity: The French Palomb cohort. European Respiratory Journal (November 8, 2016) vol. 48 no. suppl 60; DOI: <https://doi.org/10.1183/13993003.congress-2016.PA1129> ERS International Congress - London- United Kingdom.

# Remerciements

## **A ma Directrice de Thèse, madame le Professeur Chantal RAHERISON,**

Merci madame d'avoir accepté de m'encadrer tout le long de mon doctorat et d'avoir dirigé ma thèse avec beaucoup d'efforts et de patience. Vos qualités pédagogiques remarquables m'ont permis de profiter de vos connaissances et ont contribué à l'avancement de mon travail en ne négligeant ni vos conseils avisés et ni vos critiques constructives. Merci pour la pertinence de vos remarques et pour la justesse de vos corrections qui sont pour moi un exemple de rigueur. Merci pour votre disponibilité tout au long de ce travail. J'espère avoir été à la hauteur de vos attentes. Merci pour votre accueil au sein de l'équipe EPICENE (Épidémiologie des cancers et expositions environnementales) il y a déjà quatre ans, et pour la qualité de votre formation. Merci pour votre écoute et vos conseils lors de mes doutes sur mes travaux de recherche. Ces quatre années à Bordeaux, au sein de l'équipe m'ont permis d'apprendre tellement sur la pneumologie, l'épidémiologie et les bio statistiques. Vous m'avez orienté et conseillé dans mon travail de thèse, Veuillez recevoir l'expression de mes profonds remerciements.

## **A Monsieur le Professeur Mathieu MOLIMARD,**

Pr. Mathieu Molimard, président du jury de ma thèse, a accepté de me consacrer son temps en examinant le manuscrit. J'en suis honoré et je lui exprime toute ma profonde reconnaissance.

## **A Messieurs les Professeurs Nicolas ROCHE et Christophe PISON,**

Vous m'avez immédiatement fait l'honneur d'accepter de prendre part à ce jury, merci pour votre réactivité à mon égard. Merci d'avoir accepté d'être les rapporteurs de cette thèse et de relire ce travail, veuillez recevoir mes sentiments les meilleurs.

## **A Madame le Professeur Isabelle BALDI et Monsieur le Professeur Pierre-Régis Burgel,**

Vous m'avez fait l'honneur d'accepter d'examiner et de juger ce travail, veuillez recevoir mes plus sincères remerciements et soyez assuré de toute ma reconnaissance.

**A toute l'équipe PALOMB : A Madame Emilie Berteaud (attachée de recherche clinique à l'ISPED),** merci pour votre contribution et vos aides si précieuses. **Aux membres du conseil scientifique du projet PALOMB,** merci pour votre contribution dans ce travail et pour votre disponibilité.

**A Madame Christel DANTAS (Assistante administrative et pédagogique de l'équipe EPICENE) et Madame Ludivine CHRISTOPHE (Chargée de gestion des agents du centre Inserm BPH U1219 – ISPED),** merci pour vos aides précieuses et merci de m'avoir aidé continuellement à accomplir les démarches administratives.

**A Monsieur Adrien LE LEON (chargé de projet - Fondation Bordeaux Université),** merci pour votre soutien professionnel et votre accompagnement tout au long du projet PALOMB.

**Aux membres de l'ISPED et l'équipe EPICENE,**

Merci particulièrement à notre petit groupe de doctorants avec qui j'ai passé de très bons moments tous les jours au laboratoire. Je remercie aussi tous les membres de l'ISPED, de l'équipe EPICENE et en particulier ceux qui ont généreusement donné de leur temps pour m'aider. Merci à mes amis et camarades qui m'ont soutenu durant nos études.

Un merci bonus à toutes les personnes bienveillantes qui ont fait mon quotidien durant ces années en thèse.

# Table des matières

<b>INTRODUCTION GÉNÉRALE</b> .....	<b>1</b>
<b>ÉTAT DES CONNAISSANCES</b> .....	<b>4</b>
I. Définitions et généralités.....	5
I.1. Rappels sur la BPCO.....	5
I.1.1. Définition de la BPCO.....	5
I.1.2. Le diagnostic de BPCO.....	6
I.1.3. Les manifestations cliniques de la BPCO.....	7
I.1.4. GOLD, évaluation de la sévérité de la BPCO.....	9
I.1.5. Facteurs de risque.....	11
I.2. Comorbidités dans la BPCO.....	11
I.2.1. Comorbidités cardiovasculaires.....	12
I.2.2. Dénutrition et l'obésité.....	12
I.2.3. Les troubles du sommeil.....	13
I.2.4. Diabète et syndrome métabolique.....	13
I.2.5. Anxiété et dépression.....	14
I.2.6. Cancer broncho-pulmonaires.....	14
I.3. Phénotypes dans la BPCO.....	14
II.Prise en charge de la BPCO.....	15
II.1. Traitement médicamenteux.....	16
II.2. Vaccination antigrippale et antipneumococcique.....	17
II.3. Sevrage Tabagique.....	17
II.4. Réhabilitation respiratoire.....	18
III.Les cohortes de la BPCO.....	18
III.1. Les cohortes historiques de la BPCO.....	18
III.2. Les cohortes récentes de la BPCO.....	20
III.3. Les cohortes françaises.....	23
III.3.1. Cohorte EXACO.....	23
III.3.2. Cohorte initiatives BPCO.....	24
III.3.3. Cohorte COLIBRI.....	24
IV.Identification des phénotypes de la BPCO dans des études épidémiologiques.....	24
IV.1. Des classifications traditionnelles aux approches statistiques non-supervisées.....	24
IV.2. Clustering.....	26
IV.2.1. Machine learning.....	26
IV.2.2. Classification automatique.....	27
IV.2.3. Partitionnement de données.....	27
IV.3. K-means.....	30
IV.4. K-médoïdes.....	31

IV.4.1 L'algorithme PAM .....	32
IV.4.2 L'algorithme CLARA.....	32
IV.4.3 L'algorithme CLARANS.....	33
V. Pénalisation – Comment choisir parmi un grand nombre de variables.....	33
V.1. La méthode du LASSO - Least Absolute Shrinkage and Selection Operator .....	33
V.1.1. L'idée .....	34
V.1.2. L'algorithme.....	36
<b>MATÉRIEL ET MÉTHODES.....</b>	<b>42</b>
I. Présentation générale du projet PALOMB.....	43
I.1. La cohorte PALOMB.....	43
I.2. Objectifs pour les professionnels de santé .....	43
I.3. Objectif d'évaluation professionnelle.....	43
I.4. Objectif de santé publique .....	44
II.1. Suivi longitudinal.....	44
II.2. Aspects éthiques.....	44
II.3. Evaluation de l'état de santé des patients BPCO de la cohorte PALOMB.....	45
II.3.1. Les données cliniques.....	45
II.3.2. Les explorations fonctionnelles respiratoires (EFR).....	46
II.3.3. Les vingt-cinq comorbidités .....	47
II.3.4. Les médicaments .....	47
III. Exploitation des données.....	48
III.1. Gestion des données.....	48
III.2. Analyse statistique.....	49
III.2.1 Descriptif des patients BPCO de la cohorte PALOMB .....	49
III.2.2 Identification de Profils de comorbidités - Analyse en cluster .....	49
III.2.3 Identification des déterminants de la dyspnée par une méthode de régression pénalisée .....	50
III.2.4 Identification de profils d'exacerbations - classification non-supervisée.....	51
III.2.5 L'analyse en cluster la plus optimale (supervisé vs non-supervisé).....	52
III.2.6 L'analyse de survie – Principaux déterminants associés à la mortalité.....	53
<b>RÉSULTATS.....</b>	<b>55</b>
I. Descriptif des patients de la cohorte PALOMB (n=1391).....	56
I.1. Population d'étude.....	56
I.2. Les principales caractéristiques cliniquement pertinentes.....	58
I.3. Les prévalences des différentes comorbidités .....	59
I.4. Prise en charge des patients BPCO.....	60
II. Les profils de comorbidités identifiés par l'analyse en clusters (n=1584).....	64
III. Les déterminants de la dyspnée par une régression pénalisée et l'analyse de sensibilité (n=1973).....	76

IV. Les phénotypes et les déterminants d'exacerbations identifiés par une classification non-supervisée (n=932).....	86
V. L'analyse de mortalité et les phénotypes cliniques des patients BPCO identifiés par l'analyse en cluster la plus optimale (n=2653 et n=2968).....	97
V.1. Les caractéristiques initiales des patients inclus dans cette étude.....	98
V.2. Les déterminants de la mortalité.....	99
V.3. L'analyse de déclin du VEMS : Le sous-groupe.....	102
V.4. L'analyse CART (Arbre de classification et de régression).....	102
V.5. Les phénotypes cliniques des patients BPCO identifiés par l'analyse en cluster la plus optimale (n=2968).....	104
.....	
V.5.1 Partitionnement en k-Médoïdes – Choix du nombre optimal de clusters .....	106
V.5.2 Résultats de la classification PAM .....	108
V.5.3 Validation de la classification PAM (vs. HCPC et k-means).....	113
<b>DISCUSSION GÉNÉRALE ET CONCLUSION.....</b>	<b>114</b>
Points forts et points faibles de la méthodologie .....	115
PALOMB : Les caractéristiques sont conformes aux connaissances.....	115
L'identification des profils des comorbidités.....	116
L'identification des profils d'exacerbations.....	118
Les déterminants de la dyspnée (échelle mMRC).....	119
L'analyse de mortalité et l'identification des phénotypes cliniques.....	121
Conclusion.....	126
Apports principaux de cette thèse.....	126
Apports d'un point de vue clinique.....	126
Apports d'un point de vue de modélisation statistique.....	128
Perspectives.....	128
<b>BIBLIOGRAPHIE.....</b>	<b>129</b>
<b>ANNEXES.....</b>	<b>142</b>

# Liste des tableaux

<b>Tableau 1.</b> Classification GOLD de la sévérité de l'obstruction bronchique dans la BPCO selon le volume expiratoire maximal au cours de la 1ère seconde rapporté à sa valeur théorique (VEMS/Th). .....	7
<b>Tableau 2.</b> Évaluation de la dyspnée selon le mMRC. ....	8
<b>Tableau 3 :</b> Etudes étrangères en vraie vie. ....	22
<b>Tableau 4 :</b> Algorithm PAM.....	32
<b>Tableau 5.</b> Catégorisation des comorbidités. ....	47
<b>Tableau 6.</b> Caractéristiques cliniques et démographiques de la population totale.....	57
<b>Tableau 7.</b> Caractéristiques du tabagisme en fonction du stade de sévérité. ....	58
<b>Tableau 8.</b> Répartition des comorbidités selon le stade de sévérité. ....	60
<b>Tableau 9.</b> Aide au sevrage tabagique selon le conseil minimal.....	61
<b>Tableau 10.</b> Traitement BD (CDA et LDA) selon le stade de sévérité.....	62
<b>Tableau 11.</b> Traitement corticoïdes selon le stade de sévérité. ....	63
<b>Tableau 12.</b> Association des caractéristiques de 2653 patients BPCO à l'inclusion selon le statut vital. ....	98
<b>Tableau 13.</b> Les déterminants de la mortalité ajustés sur l'âge chez 2653 patients BPCO.....	101
<b>Tableau 14.</b> Les déterminants de la mortalité ajustés sur l'indice BOD chez 2653 patients BPCO.....	102
<b>Tableau 15.</b> Description des variables cliniques vis-à-vis de la Classification GOLD 2007. ....	105
<b>Tableau 16.</b> Distribution des comorbidités vis-à-vis de la Classification GOLD 2007. ....	106
<b>Tableau 17.</b> Description des variables cliniques vis-à-vis des groupes obtenus selon PAM.....	111
<b>Tableau 18.</b> Description des comorbidités vis-à-vis des groupes obtenus selon PAM.....	112
<b>Tableau 19.</b> Score de la meilleure méthode de classification par rapport à l'indice de choix de validation interne. ....	113

# Liste des figures

<b>Figure 1.</b> La BPCO va devenir la 3 <sup>ème</sup> cause de mortalité dans le monde.....	2
<b>Figure 2.</b> Mécanisme de l'obstruction bronchique.....	6
<b>Figure 3.</b> Illustration anatomique du remodelage et des altérations au niveau des voies aériennes causées par la réponse inflammatoire chez les patients BPCO (à droite) comparativement à des sujets sains (à gauche).....	6
<b>Figure 4.</b> Apparition tardive des symptômes de BPCO.....	8
<b>Figure 5.</b> Evaluation GOLD 2011 de la BPCO. ....	9
<b>Figure 6.</b> Classification de la BPCO : Outil d'évaluation ABCD affiné - GOLD 2017.....	10
<b>Figure 7.</b> Traitement en fonction du stade de sévérité de la BPCO. ....	16
<b>Figure 8.</b> Exemple de hiérarchie sous forme standard de dendrogramme. La section horizontale indique à quel niveau la hiérarchie les subdivisions ne comptent plus, ce qui produit dans cet exemple trois partitions.....	29
<b>Figure 9.</b> Cas où A se trouve dans la zone qui annule le coefficient $\beta_2$ .....	35
<b>Figure 10.</b> Calcul de la fonction $\beta LARS$ .....	38
<b>Figure 11.</b> L'algorithme 1 - LARS-Lasso.....	40
<b>Figure 12.</b> L'algorithme 2 (LARS-Lasso Amélioré).....	41
<b>Figure 13.</b> Diagramme d'inclusion des patients BPCO de la cohorte PALOMB participant aux différentes parties de l'exploitation statistique réalisée dans le cadre de ce travail de thèse.....	57
<b>Figure 14.</b> Dyspnée en fonction du stade de sévérité de la BPCO.....	58
<b>Figure 15.</b> Nombre d'exacerbations par an selon le stade de sévérité. ....	59
<b>Figure 16.</b> Traitement non médicamenteux en fonction du stade de sévérité.....	61
<b>Figure 17.</b> L'analyse de Kaplan-Meier pour évaluer la mortalité toutes causes confondues à 5 ans dans la cohorte PALOMB BPCO : <b>(A)</b> selon les outils d'évaluation ABCD de GOLD. <b>(B)</b> selon les cinq classes identifiées par l'algorithme CART.....	100
<b>Figure 18.</b> Prévalence des comorbidités chez les patients BPCO vivants et décédés.....	101
<b>Figure 19.</b> L'algorithme développé par CART pour la classification des patients BPCO. Application aux cohortes franco-belge, 3CIA et PALOMB.....	103
<b>Figure 20.</b> Choix du nombre de classes - Gap Statistic .....	107
<b>Figure 21.</b> Choix du nombre de classes - Statistique de Silhouette .....	107
<b>Figure 22.</b> Choix du nombre de classes - Méthode du coude.....	108
<b>Figure 23.</b> Distribution des patients par rapport au PAM.....	108
<b>Figure 24.</b> Fréquence des classes GOLD 2007 dans chaque cluster PAM.....	109
<b>Figure 25.</b> Fréquence des classes GOLD 2007 dans chaque cluster de PAM.....	110



# Liste des abréviations

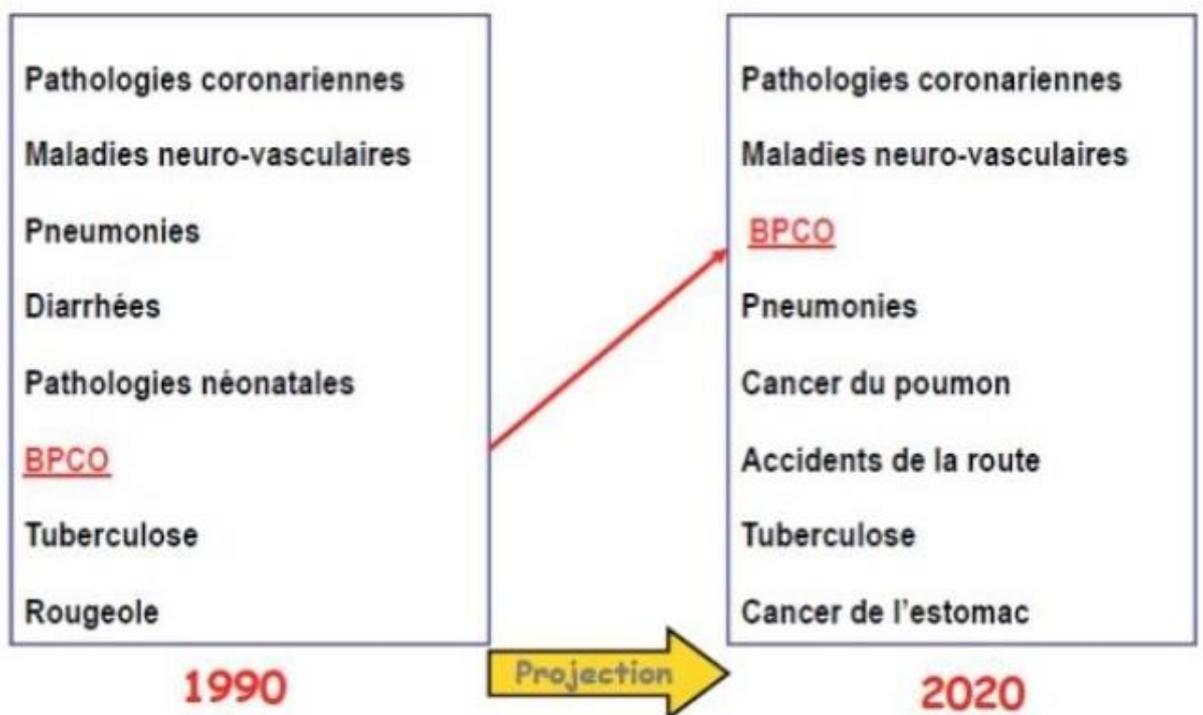
AJRCCM	American journal of respiratory and critical care medicine
AMI	Artériopathie des membres inférieurs
ATS	American Thoracic Society
AVC	Accident Vasculaire Cérébral
BMI	Body Mass Index
BODE	Body mass index, obstruction, dyspnea, exercice
BPCO	Broncho-pneumopathie chronique obstructive
CAT	COPD assessment test
CCTIRS	Advisory Committee for Data Procession in Health Research
CNIL	Commission nationale de l'informatique et des libertés National
COPD	Chronic Obstructive Pulmonary Disease
CPT	Capacité pulmonaire totale
CRF	Capacité résiduelle fonctionnelle
CSI	Corticostéroïdes inhalés
CVF	Capacité vitale forcée
DDB	Dilatation des bronches
EELV	End-expiratory lung volume
EFR	Explorations fonctionnelles respiratoires
ERS	European Respiratory Society
FEV <sub>1</sub>	Forced Expiratory Volume in 1 second
FRC	Functional residual Capacity
FVC	Forced vital capacity
GOLD	Global initiative for chronic obstructive lung disease
HTA	Hypertension artérielle
HTAP	Hypertension artérielle pulmonaire
IC	Intervalle de confiance
IC	Inspiratory Capacity
ICC	Insuffisance Cardiaque Chronique
ICS	Inhaled corticosteroid
IRC	Insuffisance respiratoire chronique
LABA	Long-acting $\beta$ -agonist
LAMA	Long-acting muscarinic antagonist
mMRC	Modified Medical Research Council
MRC	Medical Research Council
OMS	Organisation mondiale de la santé
PA	Paquets-années
PFT	Pulmonary function test
RNIPP	Répertoire national d'Identification des personnes physiques
ROC	Receiver operating characteristic

RV	Residual Volume
SABA	short-acting $\beta$ -agonist
SAMA	Short acting muscarinic antagonist
SAOS	Syndrome d'apnées obstructives du sommeil
TDM	Tomodensitométrie
TLC	Total Lung Capacity
TVO	Trouble ventilatoire obstructif
VC	Vital Capacity
VEMS	Volume expiré maximal en une seconde
VNI	Ventilation non invasive
VO <sub>2</sub> max	Volume d'oxygène maximum
VR	Volume résiduel

# **INTRODUCTION GÉNÉRALE**

La BPCO est une maladie respiratoire chronique et lentement progressive, caractérisée par une diminution non complètement réversible des débits aériens (une inflammation chronique peu réversible des voies aériennes), principalement liée au tabagisme actif mais d'autres facteurs de risque sont également incriminés tels que l'exposition professionnelle et la pollution (les aérocontaminants). Cet état pathologique comprend la forme obstructive de la bronchite chronique et de l'emphysème.

Dans le monde, la BPCO pose un problème de santé publique, sa prévalence est sous-estimée car elle est sous diagnostiquée. Selon l'OMS, 210 millions de personnes sont atteintes de BPCO. La BPCO représentait la 6<sup>ème</sup> cause de mortalité en 1990 (Voir **figure 1**), et sera la 3<sup>ème</sup> cause de mortalité en 2020 [1]. Elle concerne des sujets de plus en plus jeunes, et de plus en plus de femmes. Elle est responsable d'un handicap respiratoire, est souvent diagnostiquée tardivement. Elle est devenue une cause de transplantation pulmonaire chez les moins de 60 ans. En France, Il existe peu de données épidémiologiques sur la BPCO.



**Figure 1.** La BPCO va devenir la 3<sup>ème</sup> cause de mortalité dans le monde.

L'exploration fonctionnelle respiratoire (EFR) permet de poser le diagnostic et d'évaluer la sévérité de la maladie par la mise en évidence d'un trouble ventilatoire obstructif (TVO) non ou peu réversible sous bronchodilatateurs. Il est aujourd'hui admis que la BPCO est une véritable maladie générale, à point de départ respiratoire. Des comorbidités sont fréquemment associées à la BPCO dont elles vont compliquer l'évolution naturelle. Ces derniers temps, l'accent est mis sur l'importance des comorbidités associées à la BPCO, telles que l'ensemble des pathologies cardiovasculaires, le cancer du poumon, le diabète, le syndrome métabolique, la dépression, l'anxiété et l'ostéoporose. Ces comorbidités sont un nouvel enjeu médical et thérapeutique dans la BPCO ; elles constituent un élément incontournable par leur fréquence élevée et leur impact considérable sur la qualité de vie et le pronostic vital des patients. L'évolution de la maladie est marquée aussi par des épisodes d'aggravation cliniques appelés : exacerbations. Celles-ci jouent un rôle dans la morbi-mortalité et dans l'altération de la qualité de vie des patients.

Dès la fin des années 1960, Burrows et Fletcher reconnaissaient l'hétérogénéité de la BPCO et reprenaient des concepts développés au milieu des années 1950 par Dornhorst, suggérant l'existence de sous-groupes de patients ayant des caractéristiques cliniques différentes. L'identification de sous-groupes de patients atteints de BPCO et partageant des caractéristiques cliniques et/ou paracliniques communes et/ou une réponse thérapeutique différente est à l'origine du concept de phénotypes dans cette pathologie. De nombreuses études de cohortes observationnelles sont en cours pour tenter d'identifier de nouveaux phénotypes de patients BPCO. À ce jour, seuls 4 phénotypes ont été identifiés : le déficit en  $\alpha$ 1-antitrypsine (AAT), le phénotype « emphysème » des sommets associé à une faible capacité à l'exercice chez les patients ayant une obstruction bronchique sévère, le phénotype « exacerbateur fréquent » et le phénotype de « la bronchite chronique ». D'autres candidats phénotypes ont généralement été suspectés par l'observation de caractéristiques cliniques et/ou morphologiques particulières. Les données actuelles restent insuffisantes pour affirmer ou infirmer l'importance de la majorité de ces candidats phénotypes.

La classification par clusters est une technique exploratoire qui permet d'avoir une vision globale et résumée les informations contenues dans de grandes bases de données. Il s'agit d'un groupe d'analyses statistiques complexes permettant d'identifier des structures d'associations de données au sein d'un groupe de données. L'analyse de clusters a pour objectif « d'identifier des groupes homogènes d'individus ». L'identification de groupes aux caractéristiques similaires peut permettre d'adapter la prise en charge médicale selon les risques encourus. Dans le cadre de la BPCO, ces analyses pourraient permettre d'identifier, à partir de bases de données constituées sur des cohortes observationnelles de patients, des phénotypes non repérés par la simple observation clinique. Plusieurs équipes dans le monde ont adopté ces stratégies et les résultats de ces études, encore préliminaires, ont déjà apporté certaines informations nouvelles.

Dans ce contexte, le présent travail s'inscrit dans le cadre du suivi prospectif de la cohorte PALOMB initiée par : l'association des pneumologues privés du sud-ouest (APPSO), les pneumologues Hospitalo universitaires du service des maladies respiratoires du CHU de Bordeaux et l'équipe EPICENE - Inserm U1219- Bordeaux Population Health Research Center à l'université de Bordeaux. Depuis janvier 2014, ces pneumologues hospitaliers et libéraux se mobilisent ensemble pour développer une plateforme sur la prise en charge de la BPCO en Aquitaine et en Charentes. Palomb a été mis en place avec un recueil standardisé des variables concernant les caractéristiques démographiques, la sévérité de la BPCO, le score pronostique, les comorbidités et la prise en charge thérapeutique. Il s'agit d'une étude de cohorte avec un suivi des patients. L'objectif de cette thèse est d'explorer l'histoire naturelle de la BPCO, d'analyser les trajectoires des patients BPCO de la cohorte PALOMB, et d'identifier les phénotypes de ces patients dans le but d'améliorer la prise en charge médicamenteuse.

Dans la première partie de ce manuscrit, il sera exposé l'état des connaissances sur la BPCO, en s'attardant sur les phénotypes/profils des patients BPCO rapportés dans la littérature, puis les travaux épidémiologiques portant sur les trajectoires des patients BPCO seront détaillés. Dans un second temps, la méthodologie utilisée dans ce travail sera présentée : la cohorte PALOMB, le suivi des patients (les mesures fonctionnelles, la fréquence des exacerbations, l'échelle de la dyspnée mMRC et les traitements) et les modalités de l'exploitation statistique de ces données, en particulier : L'analyse de sensibilité, la régression régularisée (Lasso) et les modèles de classification des données. Le troisième volet sera consacré aux résultats et à leur discussion. Il offrira d'abord un descriptif détaillé des patients BPCO de la cohorte PALOMB, notamment de leur santé respiratoire, de leurs symptômes, de leurs comorbidités et des traitements, puis abordera l'identification et la caractérisation de profils de trajectoires de symptômes dyspnée/exacerbations, avant de se terminer par l'analyse de la mortalité chez ces patients au cours du suivi à 5 ans. Ce manuscrit s'achèvera sur une discussion générale dans laquelle seront soulignés les points forts et les limites de ce travail, seront mis en évidence les principaux apports de cette thèse, et seront envisagées quelques perspectives pour la suite.

## **ÉTAT DES CONNAISSANCES**

Dans ce premier chapitre, il sera fait un bilan des connaissances actuelles sur la BPCO, et sur les différentes classifications de la sévérité de la BPCO. Dans un premier temps, un rappel sur la définition fonctionnelle et clinique sera effectué, et les données épidémiologiques à l'échelle mondiale et française seront exposées. Ensuite, les différents phénotypes identifiés dans des études épidémiologiques par des méthodes de classification non supervisées, seront décrits. Enfin, dans une troisième partie, Les objectifs généraux de la prise en charge de la BPCO seront présentés. Les principaux résultats recensés dans des travaux épidémiologiques, portant sur l'association entre les facteurs de risque et la sévérité de la BPCO, viendront conclure cet état des connaissances.

## I. Définitions et généralités

### I.1. Rappels sur la BPCO

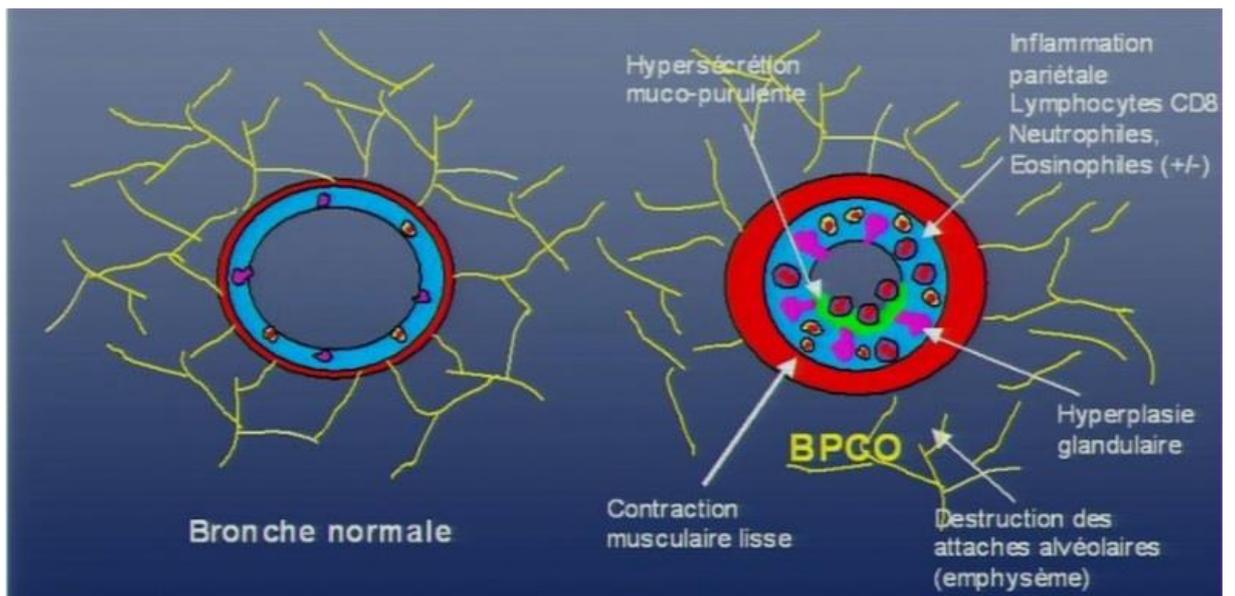
#### I.1.1. Définition de la BPCO

La broncho-pneumopathie chronique obstructive (BPCO) se définit comme une maladie chronique et lentement progressive, caractérisée par une diminution non complètement réversible des débits aériens. Les professionnels de santé doivent penser à la BPCO y compris chez l'adulte jeune et poser la question des facteurs de risque (le tabagisme associé ou non à l'inhalation de cannabis et l'exposition professionnelle ou domestique à des toxiques ou des irritants : silice, poussières de charbon, poussières végétales et de moisissures). La définition est internationale et fondée sur celle des sociétés savantes de pneumologie des Etats-Unis, du Canada, de Grande-Bretagne et de France.

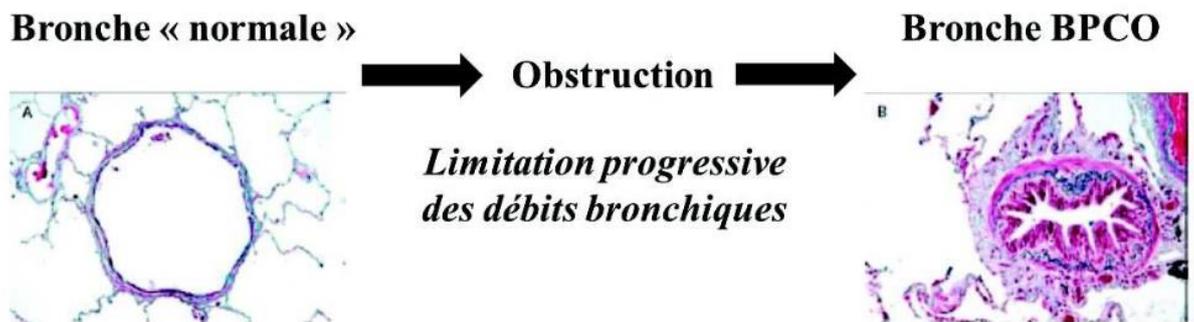
Initialement, les bronches sont soumises, par l'exposition tabagique ou à d'autres toxiques, à une réaction inflammatoire chronique pariétale avec une infiltration par des lymphocytes T et des polynucléaires neutrophiles. Cette inflammation entraîne une hyperplasie glandulaire associée, responsable d'une hypersécrétion de mucus formant un bouchon muco-inflammatoire et causant une obstruction bronchique. Si l'exposition au tabagisme est arrêtée précocement, l'essentiel de ces lésions peut régresser.

Si le tabagisme est poursuivi, des lésions parenchymateuses vont se développer avec une contraction et une hypertrophie des muscles lisses, toucher les voies aériennes distales, puis entraîner progressivement une destruction des attaches alvéolaires et une perte des fibres élastiques. Les lésions sont alors irréversibles.

En lien avec ces lésions parenchymateuses, on retrouve fréquemment des lésions d'emphysème qui a une définition anatomique car il correspond à la destruction des espaces aériens distaux au-delà des bronchioles terminales, sans fibrose, à l'origine d'une hyperinflation pulmonaire. On en distingue deux types selon la localisation : l'emphysème centro-lobulaire et pan-lobulaire. À un stade tardif, une insuffisance respiratoire chronique peut apparaître. Elle est définie par une hypoxémie chronique diurne au repos, avec une  $PaO_2 < 70$  mmHg retrouvée à deux mesures de gaz du sang à 3 semaines d'intervalle. Des exacerbations de BPCO peuvent survenir et sont définies par un événement aigu avec une aggravation des symptômes respiratoires imposant une modification temporaire de traitement. Chez les patients BPCO, la muqueuse et la paroi sont épaissies de façon irrégulières contribuant à la diminution de la lumière bronchique et donc à l'obstruction bronchique (**Voir figure 2 et figure 3**).



**Figure 2.** Mécanisme de l'obstruction bronchique.



**Figure 3.** Illustration anatomique du remodelage et des altérations au niveau des voies aériennes causées par la réponse inflammatoire chez les patients BPCO (à droite) comparativement à des sujets sains (à gauche).

### 1.1.2. Le diagnostic de BPCO

Que ce soit pour les hommes ou les femmes, même s'ils apparaissent à des âges et des intensités différentes, les premiers signes cliniques évocateurs de la maladie sont similaires et « insidieux ». En effet, les premiers symptômes tels que la toux ou encore les expectorations liées à l'irritation des bronches sont bien souvent négligées par les patients. Ainsi, au fil des années l'obstruction bronchique va se mettre en place entraînant l'apparition du principal signe clinique de la BPCO : la dyspnée.

D'un point de vue médical, l'évaluation des symptômes, notamment de la dyspnée (e.g. l'échelle du Medical Research Council (MRC), l'échelle modifié du Medical Research Council (mMRC)

ou son adaptation Française, l'échelle de Sadoul), combinée à l'analyse de l'exposition aux différents facteurs de risques constituent une première étape vers le diagnostic final de la BPCO.

Selon la définition des sociétés savantes (ATS : American Thoracic Society et ERS : European Respiratory Society) et la Global Initiative for Obstructive Lung Disease (GOLD), la composante respiratoire de la BPCO est caractérisée par « **une limitation persistante et progressive des débits bronchiques associée à une réponse inflammatoire anormale des voies aériennes et du poumon à des particules ou des gaz nocifs** » [2].

Cette limitation des débits bronchiques est mise en évidence à l'aide d'une **épreuve fonctionnelle respiratoire** : la spirométrie. Lors de cet examen, l'existence d'un déficit ventilatoire obstructif est défini lorsque le rapport entre le Volume Expiratoire Maximal au cours de la 1ère seconde (VEMS) et la capacité vitale (CV) est inférieur à 70%. Ce rapport VEMS/CV est aussi appelé le **rapport de Tiffeneau**. Le diagnostic final de la BPCO est posé seulement après un test de réversibilité (inhalation d'un bronchodilatateur bêta-2 adrénergique) et un **rapport de Tiffeneau** toujours inférieur à 70%. La sévérité de l'atteinte pulmonaire est répartie en 4 stades définis selon la GOLD [1] en fonction du VEMS exprimé en pourcentage de la valeur prédite. Les 4 stades sont présentés dans le Tableau 1 ci-dessous :

**Tableau 1.** Classification GOLD de la sévérité de l'obstruction bronchique dans la BPCO selon le volume expiratoire maximal au cours de la 1ère seconde rapporté à sa valeur théorique (VEMS/Th).

Stades	Obstruction Bronchique	EFR
<b>Obstruction : VEMS/CV &lt; 70%</b>		
<b>I</b>	<b>Légère</b>	VEMS/Th $\geq$ 80%
<b>II</b>	<b>Modérée</b>	50% $\leq$ VEMS/Th < 80%
<b>III</b>	<b>Sévère</b>	30% $\leq$ VEMS/Th < 50%
<b>IV</b>	<b>Très sévère</b>	VEMS/Th < 30%

Le diagnostic de BPCO repose sur la mise en évidence d'un trouble ventilatoire obstructif par une exploration fonctionnelle respiratoire avec mesure des débits et des volumes (spirométrie) : rapport VEMS/CVF < 70% après administration d'un bronchodilatateur. La réalisation des épreuves fonctionnelles respiratoires doit répondre aux critères de qualité d'exécution recommandés.

### I.1.3. Les manifestations cliniques de la BPCO

Les manifestations cliniques sont variables. Le début peut être insidieux avec apparition d'une toux et d'expectorations, symptômes souvent banalisés et attribués à la « bronchite chronique » du fumeur par les patients eux-mêmes. Chez les patients emphyémateux, on observera plutôt une dyspnée isolée. Enfin, dans un nombre de cas non négligeable, le diagnostic de BPCO peut être porté à l'occasion d'une exacerbation aiguë dite inaugurale de la maladie. L'histoire naturelle de la BPCO est émaillée d'épisodes d'aggravation de la maladie appelés exacerbations. La SPLF définit une exacerbation par une majoration des symptômes respiratoires au-delà des variations quotidiennes (en pratique, d'une durée  $\geq$  48h ou justifiant une modification thérapeutique). Les critères couramment utilisés sont l'augmentation de la dyspnée, de la toux, du volume de l'expectoration ou la modification

de l'expectoration (aspect purulent). L'exacerbation peut mettre en jeu le pronostic vital, on parle alors d'exacerbation sévère ou de décompensation. L'évolution chronique de cette pathologie peut conduire à une insuffisance respiratoire chronique.

Le rôle principal du professionnel de santé est de penser à la BPCO chez tout fumeur, sans qu'il n'y ait pour autant de symptômes ou de signes physiques car ceux-ci apparaissent à une phase tardive [3]. (Voir figure 3)

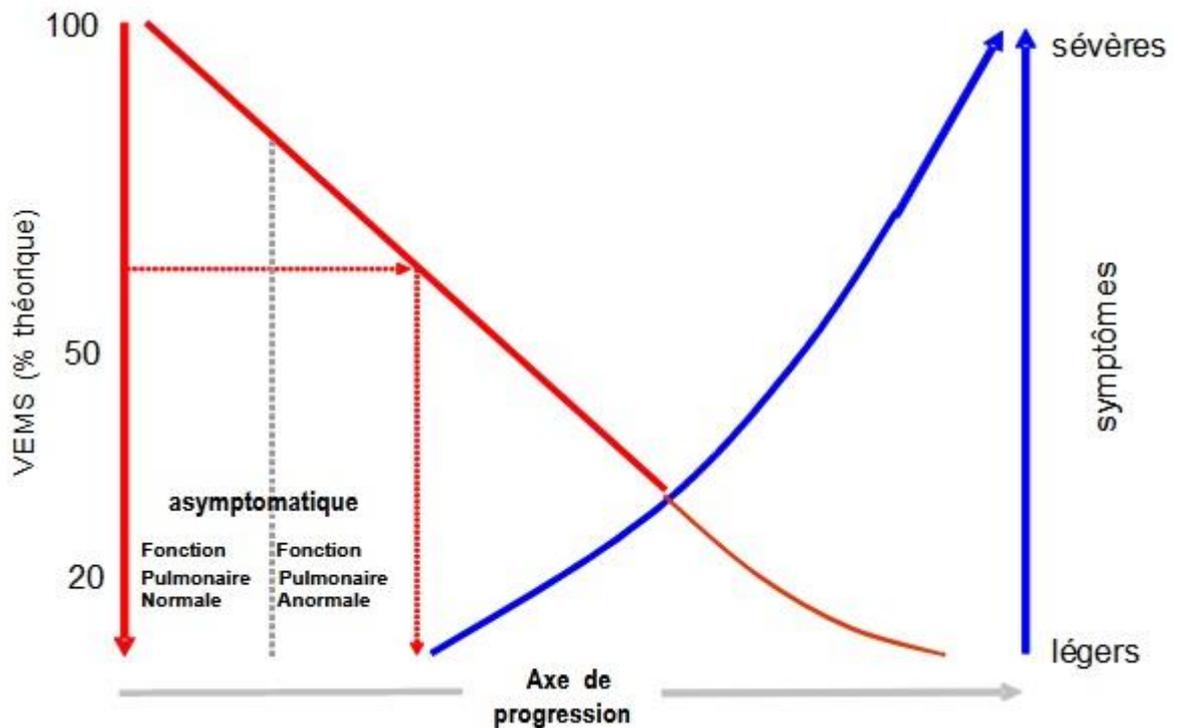


Figure 4. Apparition tardive des symptômes de BPCO.

Les principaux symptômes à rechercher sont ceux de l'auto-questionnaire réalisé par l'HAS [4] à savoir la dyspnée et la toux plus ou moins associées à des expectorations. La dyspnée, qui apparaît tardivement, doit être recherchée à l'interrogatoire chez tout patient ayant une intoxication tabagique. Initialement, elle ne surviendra qu'à l'effort important avant d'apparaître pour des efforts de moins en moins soutenus. On peut l'évaluer par l'échelle du Medical Research Council modifiée (mMRC) [5, 6]. (Voir Table 2)

Tableau 2. Évaluation de la dyspnée selon le mMRC.

<b>Stade 0</b>	Dyspnée pour des efforts soutenus (montée de 2 étages)
<b>Stade 1</b>	Dyspnée lors de la marche rapide ou en pente
<b>Stade 2</b>	Dyspnée à la marche sur terrain plat en suivant quelqu'un de son âge
<b>Stade 3</b>	Dyspnée obligeant à s'arrêter pour reprendre son souffle après quelques minutes ou une centaine de mètres sur terrain plat
<b>Stade 4</b>	Dyspnée au moindre effort

La toux et les expectorations doivent être recherchées. Elles signent la bronchite chronique dont la définition est clinique, correspondant à une hypersécrétion bronchique qui se manifeste par une toux productive quotidienne ou quasi-quotidienne durant au moins trois mois par an lors de deux années successives. L'examen clinique peut être quasi-normal [7] au stade précoce de la BPCO, avec simplement à l'auscultation des râles bronchiques témoignant de l'inflammation et de l'hypersécrétion associée.

Lorsque l'obstruction est importante, on peut retrouver un allongement de l'expiration avec les lèvres pincées, une diminution du murmure vésiculaire en cas d'emphysème et/ou un thorax en tonneau témoignant de la distension thoracique. À un stade tardif, on peut retrouver cliniquement un signe de Hoover (pincement à la base du thorax lors de l'inspiration marquant une distension thoracique sévère), une cyanose témoignant de l'hypoxie, des signes de tirage avec mise en action des muscles respiratoires accessoires, des signes d'insuffisance cardiaque droite et/ou une amyotrophie et des signes de dénutrition. Certains de ces signes cliniques peuvent être retrouvés en cas d'exacerbation de BPCO, dont le diagnostic est posé devant une majoration de la dyspnée, de la toux et/ou de l'expectoration. L'exacerbation peut être d'intensité légère, modérée ou sévère (détresse respiratoire aiguë). Son étiologie principale est infectieuse (virale ou bactérienne).

### I.1.4. GOLD, évaluation de la sévérité de la BPCO

Selon la classification de GOLD 2001-2010, la BPCO est classée en 4 stades de sévérité en fonction de la valeur du VEMS à la spirométrie : stade I (BPCO légère : VEMS/CV < 70% et VEMS > 80% prédit) ; stade II (BPCO modérée : VEMS/CV < 70% et VEMS compris entre 50 et 80% prédit) ; stade III (BPCO sévère : VEMS/CV < 70% et VEMS entre 30 et 50% prédit) ; stade IV (BPCO très sévère : VEMS/CV < 70% et (VEMS < 30% prédit). (Voir Tableau 1)

Depuis 2011, une nouvelle classification (ABCD) a été proposée prenant en compte le risque d'exacerbation, l'intensité des symptômes ressentis par le patient (quantifiés au moyen du COPD Assessment Test : CAT et de l'échelle modifiée « Medical Research Council Dyspnea Scale » : mMRC) et le degré de sévérité de l'atteinte fonctionnelle respiratoire (stades GOLD I-IV). (Voir Figure 5)

Spirométrie	VEMS ≤ 50%	C	D	≥ 2 exacerbations par an	Exacerbations
	VEMS > 50%	A	B	< 2 exacerbations par an	
		CAT < 10 et/ou mMRC ≤ 1	CAT ≥ 10 et/ou mMRC > 1		
		Symptômes			

Figure 5. Evaluation GOLD 2011 de la BPCO.

La valeur prédite est définie par des abaques, prenant en compte l'âge, le sexe, la taille et l'origine ethnique. Cette classification est à visée essentiellement didactique et ne résume pas à elle seule la sévérité de la maladie. Il faut tenir compte également des symptômes cliniques et notamment de la dyspnée qui devra être évaluée à chaque consultation. Différentes échelles sont disponibles, notamment celle du Medical Research Council (MRC), dite aussi de Sadoul qui l'a traduite en français ou l'échelle modifiée du MRC (mMRC). Bien que l'intensité des symptômes soit mal corrélée avec les stades de sévérité de la classification spirométrique de la BPCO, les symptômes cliniques et leur intensité sont pris en compte dans le choix des investigations diagnostiques et des traitements. À tous les stades, une toux et une expectoration chronique peuvent être présentes. À tous les stades, l'intensité de la dyspnée est fréquemment sous-évaluée. La fréquence des consultations varie selon la sévérité de la maladie et d'autres facteurs intercurrents.

Les points saillants des dernières recommandations GOLD 2017 pour la BPCO sont les suivants. L'objectivation d'un syndrome obstructif à la spirométrie est nécessaire pour retenir le diagnostic de BPCO et en prescrire un traitement. [8] En effet, la clinique est trompeuse et on observe un taux important de sur-diagnostic (44 % des patients traités pour une BPCO n'en n'ont en fait pas selon une étude en Suisse [9]). Le sous-diagnostic est également important, de sorte que tout symptôme respiratoire chronique justifie une spirométrie, a fortiori chez les fumeurs. GOLD propose de classier les BPCO dans un stade spirométrique (I à IV selon le VEMS, comme précédemment), mais également dans un groupe (A – D) selon les symptômes et les exacerbations (**Voir Figure 6**).

Stade selon la limitation du débit expiratoire		Groupe selon les symptômes et les exacerbations					
	VEMS (% du prédit)	<b>Exacerbations</b>					
Stade 1	≥ 80 %	≥ 2/an ou 1 hospitalisation	<table border="1"> <tr> <td>C</td> <td>D</td> </tr> <tr> <td>A</td> <td>B</td> </tr> </table>	C	D	A	B
C	D						
A	B						
Stade 2	50-79 %	0-1/an Pas d'hospitalisation					
Stade 3	30-49 %		<b>Symptômes</b>				
Stade 4	< 30 %		<table> <tr> <td>Score dyspnée 0-1 (NYHA ou mMRC)</td> <td>Score dyspnée ≥ 2 (NYHA ou mMRC)</td> </tr> <tr> <td>Score CAT &lt; 10</td> <td>Score CAT ≥ 10</td> </tr> </table>	Score dyspnée 0-1 (NYHA ou mMRC)	Score dyspnée ≥ 2 (NYHA ou mMRC)	Score CAT < 10	Score CAT ≥ 10
Score dyspnée 0-1 (NYHA ou mMRC)	Score dyspnée ≥ 2 (NYHA ou mMRC)						
Score CAT < 10	Score CAT ≥ 10						

**Figure 6.** Classification de la BPCO : Outil d'évaluation ABCD affiné - GOLD 2017

### I.1.5. Facteurs de risque

Comme son nom l'indique, la BPCO est avant tout une maladie pulmonaire et les bronches constituent le premier organe atteint. Même si les causes précises ne sont pas totalement connues, les données récentes de la littérature permettent d'avancer le fait que cette maladie résulte de l'interaction entre des facteurs de risque irritant les bronches et des prédispositions génétiques.

En première ligne, le tabagisme constitue le principal facteur de risque puisqu'il serait en cause chez 80 à 90% des patients [10]. De par sa composition en agents irritants et en radicaux libres de l'oxygène, le tabac va directement altérer la paroi bronchique. Ces risques sont aussi mis en évidence chez les fumeurs passifs [11]. De plus, le tabac est non seulement un facteur causal mais aussi un facteur aggravant la BPCO avec notamment des relations étroites entre la quantité et l'ancienneté des consommations et la sévérité de l'atteinte pulmonaire [12]. Comme précisé en introduction, la prévalence de la BPCO ne cesse d'augmenter. Historiquement la BPCO était considérée comme une maladie essentiellement masculine mais depuis une trentaine d'années, il est observé une augmentation de la prévalence et du taux de mortalité chez les femmes en lien avec une augmentation de la consommation de tabac [13]. Le sevrage tabagique représente d'ailleurs la thérapeutique la plus efficace pour prévenir et limiter la progression de l'atteinte pulmonaire de cette maladie [14].

Cependant, le tabac n'est pas le seul responsable puisque d'autres facteurs environnementaux [14] ont été identifiés tels que : La combustion de biomasse utilisée notamment dans les pays en voie de développement ; les polluants d'origine professionnelle comme l'amiante, le charbon, les solvants et la pollution atmosphérique (gaz d'échappement, fumées d'usine).

Par ailleurs, il est bien admis qu'il existe une importante variabilité dans la susceptibilité aux irritants bronchiques, notamment au tabac, puisque l'on considère que seulement 25% à 50% des fumeurs vont développer une BPCO [15]. Ces données suggèrent plus particulièrement l'existence de prédispositions génétiques [16]. A l'heure actuelle, seul un déficit en  $\alpha$ -1 antitrypsine a été clairement identifié comme facteur génétique [17] mais ce déficit est rare car il concerne seulement 1% à 3% des cas de BPCO. Les recherches dans ce domaine n'en sont qu'à leur début car même si certains gènes tels que le STAT1, HHIP ou FAM13A ont été identifiés, leurs implications dans le développement de la BPCO restent à prouver.

Pour expliquer cette variabilité de réponse aux facteurs étiologiques de la BPCO, le genre semble être aussi un paramètre à prendre en compte. De façon intéressante, une étude [18] a montré que les femmes BPCO que ce soit d'origine Africaine ou Américaine, présentent une sévérité, une précocité de l'atteinte pulmonaire et une mortalité plus importante que les hommes à consommation tabagique égale.

## I.2. Comorbidités dans la BPCO

Les pathologies chroniques dont la BPCO, surviennent rarement de façon isolée. Dans une étude américaine, portant sur près de 6000 patients consécutifs, recrutés dans une consultation de médecine interne, 60% à 80% des patients ayant une pathologie chronique avaient au moins une autre comorbidité [19]. La BPCO ne surviendrait de façon isolée que dans 30% à 40% des cas et les principales comorbidités retrouvées chez les patients BPCO sont les pathologies cardio-vasculaires, les cancers (en particulier le cancer bronchique), la dénutrition, l'ostéoporose, les troubles cognitifs, la dépression, l'anxiété et l'anémie. Des études récentes indiquent que la présence et le cumul de comorbidités chez les patients atteints de BPCO ont un impact important sur les symptômes, le pronostic vital et les coûts de traitement [20].

En France, près de 3 millions de personnes sont atteintes de la BPCO, responsable d'un handicap majeur, à l'origine d'un important coût de santé (100000 hospitalisations, 370 millions

d'euros). Une des caractéristiques de cette affection est la place que prennent les comorbidités associées. L'augmentation du coût global de santé est liée à leur nombre : les patients ayant plus de 2 comorbidités représentent 26% de la population, mais plus de la moitié des coûts de traitements [19]. La présence de l'une ou plusieurs de ces comorbidités en altère le pronostic.

### **I.2.1. Comorbidités cardiovasculaires**

Les maladies cardiovasculaires impactent sur la survie du patient atteint de la BPCO. Au cours des exacerbations de la BPCO, une maladie thromboembolique est retrouvée dans 3% à 29% des cas [21, 22] et augmente la mortalité à 1 an de 30% [23]. La prévalence de l'HTP (l'hypertension pulmonaire) au cours de la BPCO se situe entre 5% et 40% [24, 25]. Sa présence s'associe à une aggravation de la dyspnée et des échanges gazeux, à un risque d'insuffisance cardiaque (IC) droite et d'œdème des membres inférieurs et enfin à une surmortalité. Seule l'oxygénothérapie est bénéfique sur l'HTP des BPCO. De plus, Il existe un lien étroit entre coronaropathie et la BPCO : une baisse de 10% du VEMS s'accompagne d'une augmentation du risque de mortalité de cause cardiovasculaire de 28%, indépendamment des autres facteurs de risque cardiovasculaire dont le tabac [26]. L'insuffisance cardiaque (IC) et la BPCO coexistent fréquemment et induisent les mêmes symptômes, conduisant à une sous-estimation et à une prise en charge thérapeutique insuffisante. La prévalence de l'IC est de 4% à 16% des patients avec une BPCO stable et de 46% des BPCO en exacerbation [27]. L'IC et la BPCO sont l'un pour l'autre un facteur pronostique péjoratif, en terme de survie. Il est important de souligner que la prise de bêtabloquants cardio sélectifs n'est pas contre-indiquée, et, qu'à l'inverse, elle pourrait améliorer la survie des patients BPCO en cas d'IC [23].

### **I.2.2. Dénutrition et l'obésité**

Dans la BPCO, on parle de dénutrition lorsque l'IMC est inférieur à 20 kg/m<sup>2</sup>. Cette dénutrition se fait au détriment de la masse maigre (muscles). Le score de BODE, un index composite d'évaluation du pronostic, tient compte du périmètre de marche de 6 min, de l'importance de la dyspnée et de l'IMC en plus du VEMS. Elle est complexe : apports inadéquats du fait d'une dyspnée induite par l'exercice que représente le repas mais aussi par l'altération de la régulation de la leptine entraînant une diminution de la prise alimentaire [28] ; augmentation de la dépense énergétique, notamment due au travail respiratoire, au tabagisme et à certains médicaments (théophylline et bêta 2 agonistes). Au stade d'insuffisance respiratoire chronique, 17% des patients [29] ont un IMC < 20 kg/m<sup>2</sup>. Dans une cohorte danoise [30], il était retrouvé une dénutrition plus marquée chez les femmes que chez les hommes, à tous les stades de sévérité de la BPCO. Plusieurs études de cohorte européennes ont montré un lien entre dénutrition et mortalité [31-34], d'où l'importance de sa prise en charge. Après dépistage de la dénutrition, les moyens à disposition qui s'offrent au clinicien sont : La pratique de l'exercice physique dans le cadre d'une réhabilitation respiratoire ; la supplémentation hypercalorique, hyperprotidique et l'adjonction d'acide gras polyinsaturés [35]. Cette séparation des moyens est sémantique, les moyens de lutte contre la dénutrition doivent être intégrés au sein d'un programme de réhabilitation, on parle alors de réhabilitation nutritionnelle [36].

Avec un poids trop élevé (surpoids ou obésité), on a besoin de plus d'énergie pour respirer et se déplacer, ce qui occasionne également plus d'essoufflement. De plus, l'obésité limite la fonction

respiratoire. En cas d'emphysème, on observe une perte de la masse maigre (muscle) et un amaigrissement. Bien que ce soit moins apparent chez le bronchitique chronique, qui est souvent en surpoids, voire obèse, la masse musculaire est aussi affectée chez ces sujets. Il y a plusieurs causes à ceci : la respiration, plus laborieuse, exige une plus grande consommation d'énergie et ce, de manière substantielle. De plus, la BPCO sous toutes ses formes est à l'origine d'une inflammation chronique à bas bruit. Enfin, l'atteinte respiratoire limite l'effort, ce qui s'associe à une perte de masse musculaire. Dès lors, le suivi de l'état nutritionnel avec un ajustement des besoins, mais aussi un contrôle de la masse musculaire (mesure du diamètre des cuisses et des biceps), s'avèrent indispensables pour le pronostic du patient. L'essoufflement causé par la maladie peut conduire à la sédentarité : le patient bouge moins pour minimiser sa sensation d'inconfort. Or, il risque ainsi de se déconditionner. Et moins il fait d'effort, plus il sera essoufflé et vite fatigué. Pour mettre fin à ce cercle vicieux, des programmes de réhabilitation respiratoire sont proposés sur plusieurs mois. Ils consistent à redévelopper la masse musculaire ainsi que la tolérance à l'effort dans le but d'améliorer les performances au quotidien. Ce traitement est sans doute le plus efficace pour lutter contre l'invalidité que peut causer la maladie. Il semble exister un paradoxe de l'obésité (ou tout au moins du surpoids) : il n'existe pas de majoration de la symptomatologie fonctionnelle respiratoire, mais un effet protecteur sur la mortalité. Ces données sont à nuancer pour les obésités morbides (IMC>40kg/m<sup>2</sup>) [37]. Pour les patients obèses et a fortiori porteur d'une BPCO, il convient de rechercher un syndrome d'apnées obstructives du sommeil (SAOS) et un syndrome obésité/hypoventilation par la réalisation d'une gazométrie artérielle.

### **I.2.3. Les troubles du sommeil**

L'altération chronique de la qualité du sommeil (difficultés d'endormissement, difficultés de maintien du sommeil, réveils précoces et/ou cauchemars) est retrouvée chez 36% à 50% des patients BPCO [38,39]. Elle serait plus fréquente chez les patients BPCO sévères. Les facteurs de risque en sont le tabagisme persistant, les symptômes respiratoires, le sexe féminin, le syndrome anxiodépressif, l'âge croissant, l'obésité : Les causes sont multiples : la distension et l'augmentation du travail respiratoire ; les symptômes (dyspnée, toux, expectoration) ; l'anxiété et la dépression réactionnelles, les douleurs ; l'existence d'un SAOS ; l'existence d'un syndrome des jambes sans repos. Le Syndrome d'apnées obstructives du sommeil (SAOS) doit être évoqué sur les signes cliniques habituels nocturnes (ronflements, pauses respiratoires notées par le conjoint, mauvais sommeil, etc.) et diurnes (sommeil non réparateur, céphalées au réveil, etc.). Les conséquences de cette association sont importantes. Il est diagnostiqué par un enregistrement de la ventilation pendant le sommeil. La pression positive continue (PPC), traitement de référence, est efficace et améliore le pronostic.

### **I.2.4. Diabète et syndrome métabolique**

Le diabète, dit sucré, se définit par un défaut de sécrétion, d'effet de l'insuline. Les critères associent un syndrome polyuropolydypsique, une glycémie supérieure à 11,1 mmoles/L ou une glycémie à jeun supérieure à 7 mmoles/L ou une glycémie deux heures postprandial supérieure à 11,1 mmoles/L avec une HbA1c supérieure ou égale à 6,5%. Sa prévalence varie selon les études : entre 10,3%, et 18,7% [40, 41]. Des études de cohorte ont montré qu'une BPCO modérée à sévère augmente le risque de diabète avec un risque relatif à 1,4 et 1,5 respectivement. Le diabète altère le pronostic de la BPCO. Le tabagisme, cause d'inflammation, peut expliquer certaines données : les fumeurs ont deux fois plus de risque de devenir diabétiques que les non-fumeurs. La prévalence du syndrome métabolique varie selon les études de 22,5% [42] à 47,5% tous niveaux de BPCO confondus [43]. Les exacerbations sont plus fréquentes en présence d'un syndrome métabolique.

## I.2.5. Anxiété et dépression

Dans la BPCO, l'intensité de la souffrance respiratoire et sa permanence, vont pouvoir entraîner un épuisement psychologique [44]. Les troubles anxieux, sous-diagnostiqués doivent être recherchés car ils influent sur l'évolution de la maladie [45]. La particularité de la BPCO tient à la fréquence de ces troubles, à l'intrication anxiété-dyspnée et à la fréquence plus importante de la dépression chez des patients BPCO par rapport à d'autres pathologies chroniques. La dyspnée est, en effet, le symptôme angoissant par excellence. La réadaptation à l'exercice vise à hausser le seuil de la dyspnée. Une étude suggérerait que la première hospitalisation arriverait plus tôt dans l'histoire de la maladie chez un patient BPCO présentant une anxiété et/ou une dépression. Une des hypothèses pourrait être une perception plus intense et plus précoce de la dyspnée, chez ces patients [46].

La dépression affecte entre 20% et 60% des patients BPCO selon les études, les stades de la BPCO et les échelles [47]. Une étude [48] montre que les patients BPCO dépressifs meurent plus après une exacerbation que ceux qui n'ont pas de dépression. Parallèlement, la présence d'un trouble anxieux et dépressif augmente le taux d'exacerbations et prolonge la durée d'hospitalisation dans l'année qui suit. Une autre étude confirme ces données en montrant l'augmentation de la mortalité et des réadmissions chez les patients BPCO anxieux et/ou dépressifs, dans les 30 jours suivant une hospitalisation [49].

## I.2.6. Cancer broncho-pulmonaires

Le lien entre la BPCO et le cancer broncho-pulmonaires a été démontré. La prévalence de la BPCO parmi les patients atteints du cancer broncho-pulmonaires varie de 40% à 70% [50]. Une méta-analyse basée sur des études prospectives [51] montre qu'un trouble ventilatoire obstructif constitue un facteur prédictif indépendant de cancer bronchique, et cela, dès le stade I de la BPCO. On estime qu'une BPCO associée à un cancer broncho-pulmonaires diminue de 50% la survie à 3 ans. Le degré de sévérité de la BPCO a un impact sur les traitements du cancer broncho-pulmonaires lorsque celui-ci est diagnostiqué à un stade localisé ou localement avancé. Dans ce cas, le choix du traitement (chirurgie de résection pulmonaire, radiothérapie, radiofréquence) dépend en grande partie de la fonction respiratoire. Cependant, l'émergence de technique chirurgicale moins invasive par vidéothoroscopie et de la radiothérapie thoracique stéréotaxique permet dorénavant des traitements potentiellement curateurs chez des patients ayant des fonctions respiratoires très altérées.

La BPCO est un problème de santé publique qui se traduit par une morbidité, un handicap et une mortalité élevés. Les comorbidités sont très répandues chez les patients atteints de BPCO, en raison du vieillissement, des facteurs de risque communs et de facteurs génétiques. Nous sommes donc confrontés à la question de la multimorbidité et à la difficulté de l'approche pratique et pertinente de la gestion de ces comorbidités. De plus en plus, l'intérêt thérapeutique d'une prise en charge globale des comorbidités par une équipe multidisciplinaire [52].

## I.3. Phénotypes dans la BPCO

Classiquement, un phénotype est l'ensemble des caractéristiques observables d'un individu, résultant de l'interaction entre les facteurs génétiques et environnementaux. Un consensus d'expert a proposé de limiter la définition de phénotypes de BPCO aux caractéristiques cliniques et/ou paracliniques permettant d'identifier des patients ayant une histoire naturelle et/ou une réponse thérapeutique différente. La classification historique, fondée sur la sévérité de l'obstruction bronchique estimée sur le VEMS, ne rends donc que partiellement compte de la sévérité de la BPCO. Néanmoins, la BPCO est désormais considérée comme une maladie complexe à l'origine de multiples manifestations cliniques directement impliquées dans le pronostic des patients d'où la notion récente de phénotypes de patients BPCO. Parmi ces phénotypes, on retrouve différents profils de patients tels que des patients dits « exacerbateurs fréquents », ou encore des patients avec une prédominance d'emphysème, ou plus une symptomatologie de bronchite chronique [53-55].

L'étude ECLIPSE [53] a porté sur plus de 2000 patients BPCO suivis sur une période de 3 ans. Cette étude a montré que pour un même stade d'obstruction bronchique (c'est-à-dire un même niveau de VEMS), les patients avaient des symptômes très différents (intensité de la dyspnée, fréquence des exacerbations), une capacité à l'exercice différente et un retentissement différent sur la qualité de vie. Ainsi, cette étude a mis en évidence un sous-groupe de patients dits « exacerbateurs fréquents », indépendant du stade de sévérité de la maladie, qui présentent plus de deux exacerbations par an tout au long du suivi malgré la prise en charge thérapeutique. Sur la première année de suivi, dans le stade modéré, les exacerbateurs fréquents représentaient 22 % des patients. À l'inverse du phénotype « exacerbateurs fréquents », il semble exister un phénotype « non exacerbateurs » de patients qui ne présentent aucune exacerbation tout au long du suivi. Chez les patients présentant une BPCO très sévère, 29% des patients ne présentaient aucune exacerbation. Cette étude a également souligné le fait que le meilleur facteur prédictif d'exacerbations dans la première année de suivi était un antécédent d'exacerbations récurrentes.

La cohorte « Initiatives BPCO », [54] a évalué 322 patients à l'aide d'un outil de recueil standardisé permettant l'analyse de 8 variables différentes : l'âge, le tabagisme, le VEMS (%), l'index de masse corporel, le nombre d'exacerbations par an, la dyspnée selon le score mMRC et la qualité de vie selon le questionnaire de Saint Georges. L'analyse statistique en clusters des différents phénotypes cliniques a permis de classer les patients en 4 groupes distincts qui ne recouvraient pas la classification GOLD.

Plusieurs travaux se sont également intéressés aux différences phénotypiques concernant la présentation scannographique de la BPCO. Mille deux patients de la cohorte COPD Gene study ont été analysés [55]. Les auteurs décrivaient deux présentations radiologiques distinctes : l'une à prédominance « emphysémateuse », l'autre à prédominance « bronchique ». Ces deux types de présentation étaient corrélés à la sévérité d'un certain nombre de manifestations cliniques ou fonctionnelles telles que le score de BODE, indépendamment de la sévérité de l'obstruction bronchique.

## II. Prise en charge de la BPCO

Les objectifs généraux de la prise en charge de la BPCO sont [56]: Limiter le handicap en soulageant les symptômes et en améliorant la tolérance à l'exercice et la qualité de vie ; Réduire la fréquence et la sévérité des exacerbations ; Ralentir la dégradation progressive de la fonction respiratoire ; Diminuer le risque de complications (évolution vers l'insuffisance respiratoire chronique, l'hypertension artérielle pulmonaire) ; Limiter l'impact des comorbidités avec un minimum d'effets indésirables des traitements ; Réduire la mortalité. Le traitement de la BPCO repose sur le stade de sévérité de la maladie (**Voir Figure 7**).

STADE I LÉGER	STADE II MODÉRÉ	STADE III SÉVÈRE	STADE IV TRÈS SÉVÈRE
VEMS/CV < 70 % VEMS ≥ 80 % de la valeur prédite	VEMS/CV < 70 % 50 % ≤ VEMS < 80 % de la valeur prédite	VEMS/CV < 70 % 30 % ≤ VEMS < 50 % de la valeur prédite	VEMS/CV < 70 % VEMS < 30 % de la valeur prédite ou VEMS < 50 % de la valeur prédite avec insuffisance respiratoire chronique
<b>Traitement systématique</b>			
<ul style="list-style-type: none"> <li>• Arrêt du tabac : évaluation du statut tabagique, des freins et de la motivation, à renouveler régulièrement, médicaments recommandés†, thérapies cognitivo-comportementales</li> <li>• Prévention d'une exposition respiratoire aux polluants</li> <li>• Vaccination antigrippale tous les ans ‡ Chez les patients insuffisants respiratoires chroniques, vaccination antipneumococcique tous les 5 ans</li> <li>• Information/éducation thérapeutique du patient (ETP)</li> </ul>			
<b>Traitement selon les symptômes</b>			
Bronchodilatateur de courte durée d'action (CA) si besoin : β-2 agoniste CA ou anticholinergique CA			
Bronchodilatateur de longue durée d'action (LA) § : β-2 agoniste LA ou anticholinergique LA <u>Réhabilitation respiratoire</u>			
Glucocorticostéroïdes inhalés sous forme d'association fixe si exacerbations répétées et symptômes significatifs			
Oxygénothérapie de longue durée si IRC			

**Figure 7.** Traitement en fonction du stade de sévérité de la BPCO.

\*: hors exacerbations/décompensations, † : médicaments recommandés : substituts nicotiniques en 1re intention, varénicline en 2e intention, ‡ : remboursé par la Sécurité sociale chez les patients BPCO, § : si la réponse n'est pas satisfaisante, il est préférable de changer de classe plutôt que d'augmenter les doses.

## II.1. Traitement médicamenteux

Les principaux objectifs du traitement pharmacologique des patients atteints de BPCO sont la prévention et le contrôle des symptômes, la réduction de la fréquence et de la sévérité des exacerbations, l'amélioration de la qualité de vie, et enfin l'amélioration de la tolérance à l'exercice. Deux études menées sur de larges effectifs confirment l'efficacité symptomatique des bronchodilatateurs et des associations corticoïdes inhalés et β2-agonistes de longue durée d'action sur la qualité de vie et la fréquence des exacerbations [57, 58].

Alors que l'arrêt du tabac a clairement montré son efficacité sur le déclin du VEMS chez les patients atteints de BPCO, ces 2 études [57, 58] ne tranchent pas sur l'efficacité des traitements médicamenteux sur le déclin à long terme du VEMS. Le tiotropium réduit de façon significative le déclin du VEMS à long terme chez les patients BPCO stade II et chez les patients sans traitement de fond

associant un  $\beta$ 2-agoniste de longue durée d'action et un corticoïde inhalé (étude UPLIFT). Toutefois, l'amplitude de cette réduction est faible, la pente de déclin passant de 47 à 43 ml/an. Par ailleurs, la réduction de la mortalité sous traitement pharmacologique reste également controversée.

L'efficacité du traitement par bronchodilatateurs inhalés, associés ou non à des corticoïdes inhalés peut être observée sur la dyspnée et le débit de pointe dans les 15 jours qui suivent le début du traitement [59, 60].

Le mode d'action des bronchodilatateurs  $\beta$ 2-agonistes est le relâchement des cellules musculaires lisses bronchiques par la stimulation des récepteurs  $\beta$ 2-adrénrgiques. Celui des anticholinergiques est le blocage des récepteurs M3. Il n'existe pas de relation étroite entre l'augmentation des débits expiratoires (VEMS, DEP) et l'amélioration de la symptomatologie [61] ; celle-ci est plus [62] vraisemblablement liée à une réduction de la distension qui est corrélée à la réduction de la dyspnée au repos et à l'amélioration de la tolérance à l'exercice. Ainsi le bénéfice clinique rapporté par le malade prime sur l'éventuelle amélioration des débits expiratoires.

## II.2. Vaccination antigrippale et antipneumococcique

Malgré la recommandation pratiquement unanime selon laquelle les patients atteints de BPCO devraient recevoir une vaccination antigrippale annuelle, très peu d'essais contrôlés randomisés ont évalué les effets de la vaccination chez ces patients. Dans une revue Cochrane de 2010, 6 études portant sur des patients atteints de BPCO ont été analysées. En effet, les essais randomisés apportaient certaines preuves indiquant que le vaccin antigrippal inactivé diminuait effectivement les « poussées » de BPCO, en particulier lorsqu'elles étaient associées au virus de la grippe.

Le vaccin antigrippal peut réduire le risque de grippe grave et le décès chez les patients BPCO d'environ 50 % [56]. Les vaccins contenant du virus vivant inactivé ou mort sont recommandés car ils sont plus efficaces chez les patients BPCO les plus âgés. La souche virale est réétudiée chaque année pour une efficacité appropriée et les patients doivent être revaccinés chaque année. La vaccination antigrippale est particulièrement efficace chez les personnes âgées [63] et divise par quatre le risque de grippe au cours de la BPCO [64].

Le vaccin anti-pneumocoque est recommandé chez les patients BPCO de 65 ans et plus. De surcroît, ce vaccin a montré une diminution de l'incidence des pneumopathies communautaires chez les patients BPCO de moins de 65 ans avec un VEMS < 40 % de la théorique [56].

## II.3. Sevrage Tabagique

L'arrêt du tabagisme est la principale mesure susceptible d'interrompre la progression de l'obstruction bronchique et de retarder l'apparition de l'insuffisance respiratoire quel que soit le stade de la BPCO. Dans leur étude, Fletcher et Peto ont décrit l'histoire naturelle du déclin du VEMS à partir de l'étude des fonctions respiratoires mesurées de façon longitudinale tous les 6 mois sur une période de 8 ans chez des travailleurs anglais (en majorité des hommes). Cette étude montrait la relation entre le déclin du VEMS et l'âge chez les fumeurs dits « sensibles au tabac » ; en effet, la constatation d'une baisse significative du VEMS à 40 ans était un facteur prédictif puissant de la décroissance rapide du VEMS dans les années suivantes. Cette étude suggérait que l'arrêt du tabac pouvait ralentir le déclin du VEMS [65]. Cette donnée a également été démontrée de façon claire par les résultats de la Lung

Health Study [66]. Dans cette étude, Anthonisen et coll. ont suivi 5887 patients atteints de BPCO modérée sur une période de 5ans. Les sujets chez qui un sevrage tabagique complet était obtenu avaient un ralentissement net du déclin du VEMS par rapport à ceux chez qui le sevrage tabagique n'avait pu être obtenu.

Ainsi, selon les recommandations de l'HAS, un dépistage de la consommation du tabac auprès de tous les patients est préconisé ainsi qu'un conseil d'arrêt systématique. Il est recommandé d'utiliser le test de dépendance à la cigarette de Fagerström. L'aide au sevrage tabagique repose sur le conseil minimal qui doit être proposé à tout fumeur et renouvelé régulièrement, les 3 seules thérapeutiques médicamenteuses validées et recommandées pour l'arrêt du tabac chez un sujet dépendant du tabac, sont les substituts nicotiniques, la varénicline et le bupropion.

## II.4. Réhabilitation respiratoire

La réhabilitation respiratoire est l'ensemble des soins personnalisés, dispensés au patient atteint d'une maladie respiratoire chronique, par une équipe transdisciplinaire ayant pour objectif de réduire les symptômes, (dyspnée, fatigue musculaire), d'optimiser les conditions physiques (intolérance à l'effort), et psychosociales (handicap, qualité de vie), et de diminuer les coûts de santé.

Elle permet également une amélioration de la survie [67]. Les composants de la réhabilitation sont prescrits et coordonnés par le pneumologue en fonction des besoins et des souhaits du patient, de la sévérité de la maladie, et de l'environnement familial, social, et médical. Ces composants sont basés sur le réentraînement à l'exercice individualisé, l'éducation thérapeutique répondant aux besoins éducatifs spécifiques, la kinésithérapie respiratoire, ainsi que le sevrage tabagique, et des prises en charge nutritionnelle et psychosociale adaptées à chaque situation clinique.

La réhabilitation respiratoire peut être proposée aux patients BPCO qui présentent une dyspnée ou une intolérance à l'exercice, une diminution de leurs activités quotidiennes, en rapport avec leur maladie ou l'altération de leur état de santé, malgré une prise en charge optimale de leur maladie (sevrage tabagique, traitement bronchodilatateur, vaccinations). L'indication est donc essentiellement basée sur la dyspnée dans la vie quotidienne et l'intolérance à l'effort en sachant que la dyspnée est multifactorielle (respiratoire, cardiovasculaire, musculaire, nutritionnelle, psychologique) consécutive à la maladie et/ou à l'inactivité et/ou la dégradation du statut fonctionnel, aggravée par les comorbidités. La réhabilitation est indiquée chez les patients en état stable ou au décours d'une exacerbation. Elle peut être proposée dès les stades 2 et est indiquée dans les cas les plus sévères [68].

Il n'y a pas de critère objectif déterminant le besoin de réhabilitation respiratoire, puisque ce sont les symptômes, l'incapacité, le handicap et la limitation fonctionnelle qui déterminent le besoin de réhabilitation et non pas la sévérité de l'altération des paramètres physiologiques. Ainsi, il n'existe pas de critère fonctionnel spécifique [69]. Les contre-indications à la réhabilitation sont représentées essentiellement par les contre-indications cardio-vasculaires à l'exercice, qui doivent être systématiquement recherchées. Les résultats de la réhabilitation en termes de qualité de vie et de capacité d'effort sont supérieurs à ceux des médicaments. Ce traitement est l'un des plus efficaces dont nous disposons au cours de la BPCO [70].

## III. Les cohortes de la BPCO

### III.1. Les cohortes historiques de la BPCO

De nombreuses études, dans différents pays se sont intéressées à la BPCO en vraie vie. Tout d'abord, Charles Badham va décrire un relativement fréquent type de pathologie dont il va créer le nom sous forme du terme la bronchite. Cette maladie pulmonaire sera ensuite répertoriée par René Laennec, un médecin français. Cependant, Charles Badham va être le premier à différencier la bronchite de la pleurésie et de la pneumonie à travers les essais qu'il a écrits en 1808 puis en 1814. Dès les années 1950, un nouvel intérêt fut apporté à la bronchite et celui-ci va alors largement se développer. En fait, on va tout d'abord parler de cette maladie comme étant la « maladie britannique » parce qu'elle est la cause commune de décès et d'invalidité en Grande-Bretagne. Plus tard, le terme de bronchite chronique et celui d'emphysème seront souvent associés. Ainsi les travaux de NEVILLE C. OSWALD permirent une approche spécifique de cette maladie en recherchant précisément les facteurs pathogènes susceptibles de provoquer une telle pathologie. On va en étudier les aspects bactériologiques. Puis investir dans la thérapie médicamenteuse permettant de lutter le plus efficacement possible contre cette bronchite. Il est alors apparu que certains patients étaient en fait souvent des fumeurs et l'on va tenter alors de dresser une description précise de l'aspect clinique de cette pathologie avec ses facteurs étiologiques. Ainsi en fin des années 1950 il va être tenté d'établir ce qui va mieux définir la bronchite chronique. Dans l'état actuel des connaissances, les causes de la bronchite ne sont pas évidentes. L'âge de l'apparition des symptômes est extrêmement variable. Puis dans les années 1960, c'est alors un suivi d'abord sur les 5 années passées puis sur 10 ans qui va concerner 327 fonctionnaires atteints de bronchite (301 hommes et 26 femmes). Plus de la moitié (54%) des hommes sont décédés au cours de la période de 10 ans, environ 57% des décès étant attribués à des causes respiratoires et 8% associés à des carcinomes des bronches. La mortalité due à ces maladies était plus élevée que dans la population générale, mais les autres causes de décès avaient un profil normal. Les connaissances sur cette pathologie s'affinent et l'accent est mis sur la recherche d'un traitement efficace. Il s'agit principalement de mettre en place une prise en charge standard de ces patients et de tenter l'approche avec divers traitement impliquant l'utilisation d'antibiotiques.

Les travaux se poursuivent dans les années 1970, et une étude épidémiologique sur les prospectives des premiers stades du développement de cette pathologie baptisée « la bronchopneumopathie chronique obstructive » est alors réalisée chez des travailleurs de Londres. En 1980, une recherche systématique est entreprise sur la relation entre les performances musculaires d'un individu et la spirométrie chez les fumeurs et les ex-fumeurs, ceci en rapport avec le développement d'une bronchite chronique. En 1983, c'est plus spécifiquement la recherche de la pertinence chez les adultes de l'obstruction du flux d'air, mais pas de l'hypersécrétion de mucus en relation avec une potentielle mortalité attribuable au développement d'une maladie pulmonaire chronique. Cette étude compile les résultats sur 20 ans d'observations prospectives dans ce domaine de la BPCO. En 1984, cette étude fait l'analyse de l'oxygénothérapie à domicile des patients atteints de bronchite chronique. Les résultats hémodynamiques cliniques, fonctionnels et pulmonaires figurent en détail et portent sur une analyse précise de 17 cas. En 1985, c'est le diagnostic de l'hypertension artérielle pulmonaire qui est important lorsque l'on considère à la fois le pronostic et l'approche thérapeutique. Notamment, cela apparaît important en matière d'obstruction des voies respiratoires chroniques (BPCO), et donc l'intérêt pour les méthodes de diagnostic non invasives des HTAP est largement développé dans ce travail pour des applications bénéfiques dans le diagnostic de la BPCO. En 1989, un bilan fut alors possible sur les résultats obtenus dans de nombreuses études qui ont montré que l'oxygénothérapie à long terme chez les patients hypoxémiques présentant une obstruction chronique du flux d'air (BPCO) est capable d'améliorer le pronostic et de diminuer le risque de décompensation cardio-respiratoire. C'est ainsi qu'à partir des années 1992, il fut possible de mieux évaluer les avantages et les incompatibilités à pratiquer un traitement spécifique chez le patient BPCO. C'est le cas de cette étude soigneusement menée qui vise à tester l'efficacité de l'oxymétrie nocturne. Pour cela, il sera fait appel à l'utilisation d'une analyse mathématique du signal d'oxymétrie, afin d'en évaluer l'efficacité. En 1993, une analyse précise porte sur l'étude de l'emphysème pulmonaire. Cette étude informe sur la quantification par tomographie et corrélations avec des tests de la fonction

respiratoire. Durant l'année 1996 apparait une première étude qui propose sur les effets de la réhabilitation respiratoire ambulatoire sur la tolérance à l'effort et la qualité de vie des patients atteints de pneumopathie chronique obstructive. Puis en 1998, cette nouvelle analyse fait un bilan sur les patients atteints de bronchopneumopathie chronique obstructive et suivis par une oxygénothérapie à domicile. Cette étude permet d'établir des paramètres cliniques, ainsi que de répertorier l'état nutritionnel et la capacité ambulatoire de ces individus.

Parmi les cohortes historiques de la BPCO : Le Conseil de la recherche médicale (Medical Research Council (MRC)), une organisation britannique dont la mission est « d'améliorer la santé humaine à travers la recherche médicale à l'échelle internationale ». Il a pour objectif d'améliorer la santé des personnes au Royaume-Uni - et dans le monde entier - en soutenant une excellente science et en les meilleurs scientifiques. Ensuite nous retrouvons la fameuse cohorte nommée « Lung Health Study », un essai clinique randomisé multicentrique parrainé par la Division des maladies pulmonaires du National Heart, Lung, and Blood Institute. L'hypothèse testée est que sur une période de 5 ans, un programme d'intervention complet peut réduire à la fois le taux de déclin de la fonction pulmonaire et les taux de morbidité et de mortalité respiratoires chez les fumeurs d'âge moyen présentant une obstruction légère à modérée des voies respiratoires. En 1975, une grande étude prospective sur la population cardiovasculaire de 20 000 hommes et femmes a été lancée par le Dr Peter Schnohr et le Dr Gorm Jensen avec le statisticien Jørgen Nyboe et le professeur A. Tybjærg Hansen comme mentor. L'étude a été nommée « Østerbrounderisørgelsen », ou en anglais, « The Copenhagen City Heart Study ». Le but initial de l'étude était de se concentrer sur la prévention des maladies coronariennes et des AVC. Au cours des années, de nombreux autres aspects ont été ajoutés étude : maladies pulmonaires, insuffisance cardiaque, arythmie, alcool, arthrose, maladies oculaires, allergie, épilepsie, démence, stress, épuisement vital, réseau social, apnée du sommeil, vieillissement et génétique.

## III.2. Les cohortes récentes de la BPCO

Entre janvier 1997 et juin 2002, un total de 859 patients ambulatoires présentant un large éventail de sévérité de la BPCO ont été recrutés dans des cliniques aux États-Unis, en Espagne et au Venezuela. Il s'agit de la cohorte BODE. En 2004, Celli et coll. ont publié l'index BODE, sous forme d'un score pronostique capable d'estimer la mortalité à 52 mois d'un patient avec BPCO, en prenant en compte l'aspect pluridimensionnel de la maladie : B pour Body mass index, indice de masse corporelle (IMC) ; O pour obstruction, donc VEMS % du prédit ; D pour dyspnée selon l'échelle du MRC ; E pour exercice capacity, soit la distance parcourue en six minutes). Cette publication a représenté une avancée importante dans l'appréhension de la complexité des éléments pronostiques de la BPCO et a marqué le début d'une activité prolifique dans ce domaine. L'index BODE a réuni d'emblée un certain nombre d'atouts : il avait été réalisé dans différents centres dans le monde et validé, avec un nombre relativement élevé de sujets. Son pouvoir de discrimination était satisfaisant. Son talon d'Achille était le test de la capacité d'effort par la mesure de la distance effectuée en marchant pendant six minutes.

L'Espagne est un pays où le fardeau de la population de la BPCO et d'autres maladies chroniques associées au tabagisme devrait augmenter dans les années à venir, bien que les récentes réductions du tabagisme global justifient une réévaluation de ces prévisions. Les données épidémiologiques disponibles en Espagne ont été récemment examinées. L'Estudio epidemiológico de EPOC (IBERPOC), une enquête historique de prévalence menée en 1997, a signalé que 9,1% de la population générale espagnole âgée de 40 à 69 ans souffraient de la BPCO. L'IBERPOC a été la première grande enquête de population à utiliser la spirométrie post-bronchodilatatrice pour déterminer la prévalence de la BPCO, et ses résultats sur les sous-diagnostic et les sous-traitements ont également

été identifiés dans des enquêtes internationales plus récentes. L'étude épidémiologique de la BPCO en Espagne (EPI-SCAN) est une nouvelle évaluation plus récente de la répartition de la population de la BPCO en Espagne

ECLIPSE est une étude prospective longitudinale non interventionnelle de 3 ans en cours, menée dans 46 centres dans 12 pays. Après une visite de base, les sujets doivent être suivis à un total de sept visites à : 3 mois, 6 mois et tous les 6 mois par la suite pendant 3 ans. En plus des visites d'étude, les patients atteints de BPCO doivent être téléphonés chaque mois entre les visites à la clinique afin d'évaluer les taux d'exacerbation. Tous les sujets doivent continuer à recevoir leurs médicaments normaux prescrits tout au long de l'étude ; aucun médicament n'est interdit pendant l'étude. L'étude ECLIPSE a les objectifs spécifiques suivants : Utilisation de questionnaires, de spirométrie, de tests d'effort et de tomodensitométrie (TDM) pour la définition des sous-types de BPCO cliniquement pertinents chez les patients atteints de BPCO de stade II-IV de l'Initiative mondiale pour la maladie pulmonaire obstructive chronique (GOLD) ; Identification et définition des paramètres qui prédisent la progression de la maladie sur 3 ans dans les sous-types de MPOC cliniquement pertinents chez les personnes atteintes de BPCO de stade GOLD II-IV.

« COPDGene study » est une étude observationnelle multicentrique conçue pour identifier les facteurs génétiques associés à la BPCO. Elle caractérisera également les phénotypes de tomodensitométrie thoracique chez les sujets atteints de BPCO, y compris l'évaluation de l'emphysème, du piégeage de gaz et de l'épaississement de la paroi des voies respiratoires. Enfin, les sous-types de BPCO basés sur ces phénotypes seront utilisés dans une étude complète à l'échelle du génome pour identifier les gènes de susceptibilité à la BPCO. Cette étude recrutera 10 000 fumeurs avec et sans BPCO à travers les stades GOLD. Les sujets blancs et afro-américains non hispaniques sont inclus dans la cohorte. Un scanner thoracique inspiratoire et expiratoire sera obtenu pour tous les participants.

SPIROMICS est une étude observationnelle multicentrique de la maladie pulmonaire obstructive chronique (BPCO) conçue pour guider le développement futur de thérapies contre la BPCO en fournissant des critères solides pour classer les participants atteints de BPCO dans les groupes les plus susceptibles de bénéficier d'une thérapie donnée au cours d'une clinique essai, améliorant ainsi les chances de succès; et identifier les biomarqueurs / phénotypes qui peuvent être utilisés comme résultats intermédiaires pour prédire de manière fiable les bénéfices cliniques au cours des essais thérapeutiques, réduisant ainsi les coûts. L'objectif est de recruter 3 200 participants dans quatre stades : BPCO sévère, BPCO légère / modérée, fumeurs sans obstruction des voies respiratoires et témoins non-fumeurs. Les participants subissent une visite de base comprenant des mesures morphométriques, une spirométrie, une marche de six minutes, un scanner thoracique inspiratoire et expiratoire et un ensemble de questionnaires standardisés. Les échantillons biologiques, y compris le plasma, le sérum, l'ADN, l'urine et les expectorations induites, sont collectés et stockés. Il y a trois examens de suivi annuels, avec des appels téléphoniques trimestriels pour évaluer les exacerbations, les hospitalisations et la mortalité. Une bronchoscopie est pratiquée dans un sous-ensemble de participants et un sous-ensemble de patients atteints de BPCO sera évalué pendant les exacerbations. Une évaluation des exacerbations et de la mortalité sera entreprise. SPIROMICS est conçu pour que des sous-études et des études auxiliaires testant des hypothèses supplémentaires puissent être ajoutées.

**Tableau 3** : Etudes étrangères en vraie vie.

Etude	n	Critères d'inclusion	Objectif principal, résultats principaux
How far is real life from COPD therapy guidelines? An Italian observational study	4094	≥40 ans, fumeurs ou ex-fumeurs, VEMS/CVF < 70%, en état stable. Etude observationnelle multicentrique.	Evaluation et comparaison des pratiques cliniques par rapport aux recommandations GOLD. -Bronchodilatateurs : 73.5% -ICS monothérapie : 15.2% -Association ICS + LABA : 66.8% -corticoïdes : 3.7%
Management of COPD in the UK primary-care setting: an analysis of real-life prescribing patterns.	-24957 patients avec dg de BPCO. -13557 patients GOLD II	Dg de BPCO avec spirométrie de moins de 5 ans, âge > 35 ans. Etude rétrospective sur 5 ans	Evaluation de la PEC de la BPCO dans les soins primaires, analyse en sous-groupe des patients GOLD II. Patients GOLD II: 17.2% non traités, >50% des patients ont des ICS seuls ou en association: ICS seuls: 6.5%, ICS+LABA: 26.7%, ICS+LABA+LAMA: 19.9%, Tabac actif: 30%.
General practitioners' adherence to the COPD GOLD guidelines: baseline data of the Swiss COPD cohort study.	615	VEMS/CVF < 70% Etude observationnelle prospective	Evaluation et comparaison dg et PEC BPCO aux recommandations GOLD -44% des patients avec dg initial de BPCO n'avaient pas les critères spirométriques. -prescriptions conformes aux recommandations : dans 55% des cas pour stade I, 46% pour stade II, 65% pour stade III et 66% pour stade IV. -Réhabilitation pulmonaire peu prescrite -seuls 56% des patients stade II à IV avaient LAMA ou LABA.-ICS: 48% dans stade I, 63% stade II.
Guidelines adherence in management of stable COPD	450	Patients suivis pour une BPCO en consultation au CHU. VEMS/CV < 70%, en état stable. Etude Américaine rétrospective, unicentrique	Evaluation adhérence des médecins face aux recommandations par rapport ttt pharmacologique. 43.6% des patients ne recevaient pas ttt en concordance avec recommandations.
Quality of COPD care in hospital outpatient clinics in Denmark: the KOLIBRI study.	941	2 audits à partir des dossiers de tous les patients suivis pour une BPCO	Décrire la qualité des soins de la BPCO au Danemark et évaluer si un programme éducatif sur les Guidelines de la BPCO améliore la qualité des soins. Amélioration de la PEC entre les 2 audits

Management, morbidity and mortality of COPD during an 11-year period: an observational retrospective epidemiological register study in Sweden (PATHOS)	21361	-Dg de BPCO ou prescription de bronchodilatateur LDA -Etude rétrospective observationnelle.	Description de la population BPCO, caractéristiques de la maladie et de la prise en charge sur des données longitudinales sur 11 ans. -Dg plus jeune de la BPCO en 2009 par rapport à 1999 (66 ans vs 73 ans), augmentation incidence et prévalence de la BPCO, augmentation prescription LAMA et ICS et diminution nb exacerbation (3-> 1.3), tabagisme stable (57%).
Canadian practice Assessment in COPD: respiratory specialist physician perception vs patient reality.	931	VEMS/CV< 70% Ou BPCO suspectée (symptômes de bronchite chronique, + facteurs de risque)Etude prospective, en vraie vie.	Comparaison des prescriptions des pneumologues selon la perception des patients et les recommandations.32% de tabagisme actif, 81% de vaccination antigrippale, 70% vaccination pneumocoque, 24% des patients ont une réhabilitation respiratoire.

### III.3. Les cohortes françaises

Pour pallier au manque de données épidémiologiques concernant la BPCO en France, plusieurs études sont actuellement en cours afin de mieux caractériser les patients BPCO et de mieux les prendre en charge.

#### III.3.1. Cohorte EXACO

EXACO [71] est une étude épidémiologique observationnelle, nationale de type suivi de cohorte auprès de 1200 patients consultant en pneumologie. Le suivi des patients s'effectue sur 4 ans. Les buts de cette étude sont de décrire qualitativement et quantitativement les exacerbations de patients atteints de BPCO en fonction du stade de sévérité et de vérifier l'hypothèse d'un sous-groupe de patients « exacerbateurs fréquents » présentant un nombre d'exacerbation homogène dans le temps. Le suivi de cohorte des patients BPCO sur 4 ans permet de préciser année par année le nombre seuil d'exacerbations par an des patients appartenant à ce sous-groupe, la régularité du nombre d'épisodes année par année et de définir les variables déterminants ce profil de patients. Les facteurs associés à ce profil « exacerbateurs fréquents » et au déclin accéléré de la fonction respiratoire sont également évalués. Les données sont recueillies auprès des pneumologues et par entretien téléphonique. Les pneumologues libéraux ou hospitaliers, incluent des patients âgés de 40 ans ou plus, atteints de BPCO de stade II ou III (selon les critères de la SPLF 2003), fumeurs ou ex-fumeurs (> 15 PA),

en état stable, dont le VEMS après bronchodilatateur est inférieur ou égal à 80% de la théorique et le rapport VEMS/CV < 70%.

### **III.3.2. Cohorte initiatives BPCO**

Une vingtaine de pneumologue, hospitaliers ou libéraux, réunis au sein de « Initiatives BPCO » ont mis en place une cohorte de patients BPCO dont le diagnostic repose sur les EFR. Il s'agit d'une inclusion de patients « au fil de l'eau ». Le début de l'inclusion des patients remonte à août 2008. Les nombreuses informations recueillies prospectivement comprennent des données démographiques, cliniques, fonctionnelles et thérapeutiques. Ces données ont été regroupées au sein d'une base informatique [72].

### **III.3.3. Cohorte COLIBRI**

Il s'agit d'un observatoire basé sur une consultation-web pour les pneumologues hospitaliers et libéraux de l'Isère, de la Savoie, Haute-Savoie, de la Drôme et des Hautes-Alpes. Le but de cette étude est d'identifier des profils de patients et d'adapter leur prise en charge et de participer à la recherche en vraie vie sur l'évolution de la BPCO. La visite web se déroule selon la logique habituelle d'une consultation avec recueil de l'histoire de la maladie, des comorbidités et du traitement habituel, un examen clinique guidé, des explorations fonctionnelles, morphologiques ou biologiques, l'évaluation des déterminants du handicap et la prescription à l'issue de la consultation (traitement médicamenteux, soins ou examen). Le projet a été lancé officiellement en mars 2013 et en juillet 2014, 1374 patients étaient inclus (dont 40% par des pneumologues de ville) [73].

## **IV. Identification des phénotypes de la BPCO dans des études épidémiologiques**

La BPCO est une pathologie multifactorielle dont l'expression, influencée par l'interaction avec des facteurs génétiques, comportementaux et environnementaux, est hétérogène. De plus, cette maladie a tendance à coexister avec d'autres comorbidités chez un même individu, suggérant également des interactions entre elles et des mécanismes communs. L'histoire naturelle de la BPCO est complexe ; elles apparaissent parfois tôt, persistent ou non, de l'âge adulte jusqu'au vieillissement et elle a un degré de sévérité variable. Ainsi, il est admis qu'au sein de la BPCO, il existe plusieurs profils (parfois désignés sous le terme de « phénotypes »), c'est-à-dire des groupes au sein desquels les sujets sont semblables au regard des caractéristiques de leurs symptômes (origine, sévérité, évolution).

### **IV.1. Des classifications traditionnelles aux approches statistiques non-supervisées**

Il a été procédé à la distinction de différents phénotypes au sein de la BPCO afin d'adapter les traitements et les décisions cliniques, de fournir un pronostic fiable, de faciliter l'identification des facteurs de risques et d'établir des stratégies de prévention efficaces.

Traditionnellement, ces phénotypes sont définis a priori en se basant sur des critères jugés pertinents par des groupes d'experts cliniciens. Ainsi, différents phénotypes de la BPCO peuvent être décrits suivant la sévérité de celle-ci, en se basant sur les scores de sévérité existants. À ce jour, quelques phénotypes ont été identifiés : le déficit en  $\alpha$ 1-antitrypsine (AAT) [74], le phénotype "emphysème des sommets associé à une faible capacité à l'exercice chez les patients ayant une obstruction bronchique sévère" [75], et le phénotype "exacerbateur fréquent" [76]. Le déficit en AAT est historiquement le premier phénotype de BPCO identifié. Il touche environ 1 % des patients BPCO. Son diagnostic nécessite un dosage de l'AAT sérique ou plasmatique et un génotypage du gène SERPINA1 (ou protéase inhibitor : PI) [74]. Ce phénotype a été initialement identifié chez des patients jeunes, fumeurs, développant un emphysème panlobulaire prédominant aux bases pulmonaires et une obstruction bronchique sévère. Il est aujourd'hui bien établi que le déficit en AAT peut être diagnostiqué chez des patients ayant un emphysème diffus ou prédominant aux bases et chez des patients plus âgés ; le degré d'obstruction bronchique varie selon les patients [74]. Le bénéfice individuel du diagnostic de déficit en AAT est faible à ce jour, car les traitements substitutifs ont des effets inconstants. Ce phénomène reflète la physiopathologie complexe de cette pathologie, qui fait intervenir non seulement un déséquilibre de la balance protéase/antiprotéase mais aussi le rôle pro-inflammatoire des polymères d'AAT mutés [74]. L'avènement de la chirurgie de réduction de volume pulmonaire chez les patients atteints d'obstruction bronchique sévère associée à une distension thoracique par poumon emphysémateux a conduit à identifier un phénotype de patients BPCO pour lesquels cette intervention chirurgicale présente un bénéfice [75]. Les patients ayant un emphysème hétérogène prédominant sur les sommets et une faible capacité d'exercice tirent un bénéfice important de cette chirurgie qui, en plus de réduire le volume pulmonaire, diminue le nombre d'exacerbations, améliore la capacité d'exercice (de façon transitoire néanmoins) et apporte probablement un gain en termes de survie [74]. Cette intervention chirurgicale est néanmoins lourde, associée à une morbidité non négligeable, et n'est que rarement pratiquée de nos jours, limitant l'intérêt actuel pour ce phénotype. Des études sont en cours pour évaluer l'intérêt de procédures endoscopiques de réduction de volume pulmonaire, mais la pertinence de ce phénotype pour prédire la réponse à la réduction de volume endoscopique n'est pas démontrée. Les exacerbations sont d'une importance clinique majeure chez les patients atteints de BPCO et peuvent être prévenues par des thérapeutiques médicamenteuses (bronchodilatateurs de longue durée d'action éventuellement associés aux corticoïdes inhalés, vaccinations) et non médicamenteuses (réhabilitation respiratoire) [77]. L'étude de cohorte observationnelle ECLIPSE a confirmé l'existence d'un sous-groupe de patients BPCO ayant des exacerbations fréquentes (2 ou plus par an) de façon persistante sur les 3 années de l'étude : seuls 12 % avaient des exacerbations fréquentes, et 23 % n'ont eu aucune exacerbation dans les 3 années de l'étude (identifiant un phénotype non-exacerbateur) [76]. De façon intéressante, les patients ayant des exacerbations fréquentes malgré les traitements habituels pourraient bénéficier de l'adjonction d'un traitement par azithromycine au long cours [78] et peut-être, dans le futur, de nouvelles thérapeutiques ciblant spécifiquement les exacerbations. Des études sont en cours pour évaluer la prévalence de ce phénotype dans la population des patients BPCO en France [71].

L'avantage de ces phénotypes traditionnels dans le cadre d'études épidémiologiques est qu'ils peuvent être déterminés de manière relativement simple dans de larges échantillons de patients. Cependant, le choix du critère utilisé amène beaucoup de subjectivité dans l'analyse, et ne considérer qu'un seul critère est sans doute trop réducteur pour isoler finement de réelles entités pathologiques. Plus récemment, l'utilisation de méthodes de classification dites non-supervisées utilisant des

approches statistiques s'est répandue et a permis de caractériser des profils de manière plus objective. La principale force de ces approches est d'utiliser simultanément plusieurs paramètres relatifs aux maladies respiratoires (différents symptômes, à des niveaux distincts de sévérité, déclenchés par divers facteurs, sur différentes périodes) sans émettre d'hypothèses quant à leur importance pour définir les profils ; ces hypothèses étant « générées » par des méthodes statistiques, d'où le terme d'approche non-supervisée.

## IV.2. Clustering

Les algorithmes de classifications ont pour objectif principal de séparer des données d'un échantillon, les agréger de manière à constituer un ensemble de groupes homogènes, c'est à dire contenant chacun des éléments partageant certaines caractéristiques communes. Cet objectif est plus largement celui du clustering, ou partitionnement de données en français, qui est une problématique phare du machine learning. En premier lieu nous éclaircirons ce que désignent les termes clustering et machine learning et ce qu'il se fait dans ce domaine. Dans un second temps nous parlerons en détail de l'algorithme k-means et k-medoids deux méthodes connues et efficaces.

### IV.2.1. Machine learning

Compte tenu du nombre d'observations et de paramètres les dépendances et tendances émergeant des jeux de données sont souvent trop complexes pour être décrits et mis en évidence aisément, d'où la nécessité d'utiliser des algorithmes pour automatiser le traitement.

Le machine learning est issu de la branche de l'intelligence artificielle et est donc lié par essence à l'informatique. Il a pour but de résoudre ces problèmes difficiles dans l'optique de prendre des décisions stratégiques à partir d'informations stockées sous la forme de base de données. Ces algorithmes également rassemblés sous le terme « d'apprentissage automatique » ont la particularité de pouvoir apprendre, d'ajuster leur modèle et leur comportement en fonction des observations et ainsi d'évoluer d'eux-mêmes pendant le traitement des données. Cette souplesse permet d'obtenir des algorithmes opérationnels malgré la difficulté du problème, de relever le défi généré par des bases de données toujours plus volumineuses et complexes auxquelles on peut désormais se confronter à notre époque. Derrière le terme d'apprentissage automatique se cache de nombreuses branches qui se différencient selon le type de données à traiter, le choix de modélisation, l'objectif de l'étude et les moyens techniques à disposition.

Les données les plus classiques sont sous forme matricielle où chaque ligne représente une observation et chaque colonne représente une variable. Lorsque les variables sont de nature discrète (catégorielles) on parle de variables qualitatives. On utilise le terme « quantitatives » quand elles sont continuées et « mixte » lorsque ces deux dernières sont présentes.

Mais on peut par exemple être amené à étudier des données de type enregistrement en temps réel comme celles obtenues par des capteurs sonores et visuels ou encore issues de l'enregistrement de l'activité d'individus sur internet. On peut vouloir également étudier des données fonctionnelles, c'est à dire décrivant une courbe d'évolution temporelle d'une mesure comme des enregistrements de trajectoires des patients ou de séries temporelles issues de la finance par exemple. Enfin on peut

dénoter les données textuelles où chaque observation est un fichier texte qui doit être analysé au préalable ou encore les données sous forme de graphe.

Dans chacun de ces cas, les méthodes et modélisations employées pourront énormément varier. Parmi l'ensemble des familles d'algorithmes traitant de ce domaine, les plus classiques sont : Les réseaux de neurones (Deep learning si à couches multiples) ; Les machines à vecteur de support (SVM) ; Les algorithmes et programmation génétique (itérations et mutations des solutions) ; Les méthodes statistiques (Algorithmes de type EM par exemple) ; Les arbres de décision ; k plus proches voisins ; Boosting (fusion de classifieurs binaires). Les applications réelles sont variées. On retrouve notamment le machine learning en première place dans l'aide au diagnostic médical.

### **IV.2.2. Classification automatique**

La classification automatique est une des sous-branches du machine learning dont la vocation est d'extraire/retrouver automatiquement les différents groupes et tendances communes au sein d'un échantillon de données.

Afin de pouvoir comprendre et se faire une représentation du réel, les limites de la cognition humaine nous obligent à penser en terme de classes, à catégoriser les choses qui nous entourent afin de pouvoir naviguer dans cette complexité. Cette catégorisation notre cerveau l'exécute de manière remarquable dans toutes les situations de la vie quotidienne. Cependant nous sommes bien incapables d'effectuer la même tâche lorsqu'on considère des gigas d'informations stockées dans des ordinateurs, sous forme matricielle par exemple. Notre subjectivité n'étant vue que comme un obstacle à la découverte d'informations fondues au sein d'un échantillon, le but de la classification automatique est d'attribuer une classe à chaque individu à partir de procédés ne faisant intervenir que les données et très peu l'utilisateur. La subjectivité de l'expérimentateur ne doit être présente que dans le choix de la modélisation et des représentations que l'algorithme utilise.

On peut différencier deux grandes familles d'algorithmes de classification automatique par rapport à l'expertise, à la connaissance préalable nécessaire à la modélisation et résolution : les algorithmes supervisés et les algorithmes non-supervisés.

Dans le premier cas, supervisé, les classes et leurs caractéristiques sont connues au préalable. À partir d'une base de données d'apprentissage qui contient des observations dont certaines possèdent leur classe de renseignée, on cherche à obtenir la meilleure règle de classification possible, pour enfin appliquer cette règle et prédire la classe des observations dont la classe est manquante. On parle alors de classification.

À l'inverse les algorithmes non-supervisés utilisent des jeux de données dans lesquels on ne connaît initialement rien sur les classes en son sein. Il faut donc que ces algorithmes soient capables de déterminer leur nombre (bien qu'il puisse être laissé à l'utilisateur dans certains algorithmes. . .), leur forme et quelles observations leur appartiennent. On parle dans ce cas de partitionnement, d'agrégation de données ou encore clustering. C'est dans cette dernière catégorie que notre recherche s'inscrit.

### **IV.2.3. Partitionnement de données**

Apparu au milieu du 20<sup>ème</sup> siècle, le partitionnement de données appartient au domaine du machine learning et plus précisément à la classification automatique non supervisée. On se place donc dans le cadre où le nombre de groupes et les caractéristiques communes de ces groupes au sein de l'échantillon ne sont pas connus par l'expérimentateur. Le pari est fait qu'au sein du jeu de données d'intérêt, des concentrations de données similaires vont se démarquer des autres amas de données par leur dissimilarité. Ainsi on va chercher à catégoriser une collection d'objets de telle manière que des observations appartenant à un même groupe soient plus similaires entre elles qu'avec tout autre élément appartenant à d'autres groupes, et dissimilaires quand ils appartiennent à des groupes différents. En d'autres termes le clustering vise à diviser les observations en différents groupes, homogènes au sein du groupe et hétérogènes entre groupes. On peut souvent lire que pour obtenir un bon partitionnement, il est nécessaire d'à la fois minimiser l'inertie intra-classe (variance au sein d'un groupe) pour obtenir des clusters les plus homogènes possibles mais aussi maximiser l'inertie inter-classe (variance entre les groupes) afin d'obtenir des groupes distincts. Un partitionnement idéal est constitué de groupes bien séparés et compacts.

Cette homogénéité est très souvent évaluée selon une distance ou plus largement une mesure de dissimilarité dont le choix revient à l'expert et peut plus ou moins fortement influencer sur la qualité des partitions, en induisant un biais potentiel. Cette mesure de similarité peut se traduire tout simplement par une distance euclidienne si les données s'y prêtent (repère orthonormé) ou se baser sur une mesure propre au problème et définie au préalable par l'utilisateur.

La communauté travaillant dans ce domaine éprouve des difficultés à unifier ce relativement informel concept qu'est le clustering. Un grand nombre d'approches et de concepts différents sont à la racine des algorithmes de regroupement. Les possibilités sont larges, les compromis inévitables compte tenu de la difficulté du problème et les objectifs ainsi que la nature des données sont tellement variés que les algorithmes existants ne sont pas tous aisément comparables. Cependant on peut classer les méthodes existantes en catégories selon sur quel principe elles se reposent.

Certains types d'algorithmes de clustering peuvent reposer sur des hypothèses sur la distribution des individus à classer. Les groupes présents au sein de l'échantillon sont alors supposés suivre une loi connue, où seuls les paramètres de cette loi demeurent inconnus et sont à déterminer. On appelle ces dernières les méthodes paramétriques car définir un groupe revient à en définir ses paramètres (moyenne et variance pour une loi normale par exemple). Les chances pour un individu d'être affecté à un groupe donné pourront alors être calculées de manière probabiliste, donnant des probabilités d'appartenance à chaque groupe plutôt qu'une attribution exclusive, induisant ainsi un clustering qualifié de « doux ». Les algorithmes de type « Espérance-Maximisation » en sont de bons représentants.

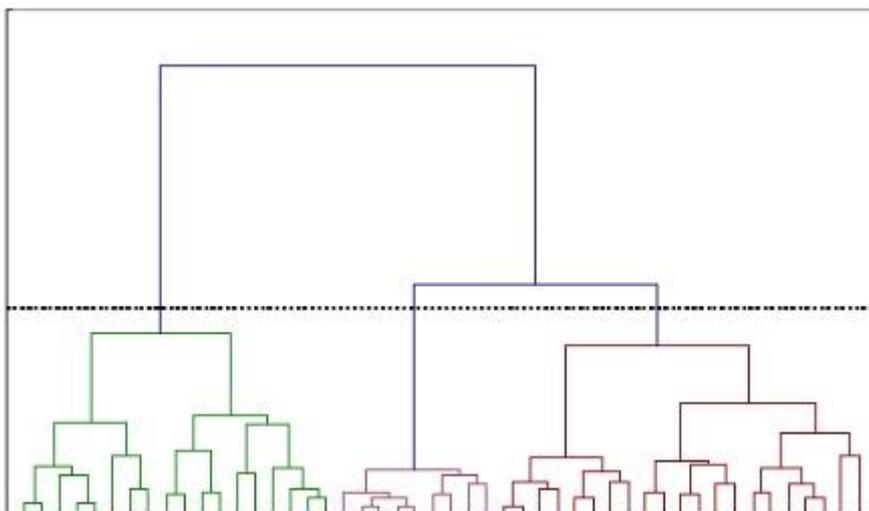
Le second type de méthodes, non-paramétriques, ne reposent sur aucune hypothèse vis-à-vis des distributions et donnent lieu à des méthodes variées se reposant sur des concepts diversifiés. Les résultats qu'ils produisent peuvent aussi être de plusieurs formes. Sous forme de partitions, où chaque élément possède une classe bien définie (« flat clustering » en anglais, opposé au clustering « doux » cité plus haut), ou sous une forme hiérarchique à laquelle on doit appliquer une étape supplémentaire pour en obtenir les partitions finales.

En fonction de l'approche choisie pour la modélisation (paramétrique / non-paramétrique) et le type de données étudiées, on se place dans une famille particulière d'algorithmes de clustering dont voici une liste non-exhaustive.

On peut tout d'abord compter les méthodes basées sur des centroïdes comme k-means [79], k-medoids [80], « vector quantization » [81] ou mean-shift [82] pour ne citer que les plus connus.

Comme leur nom l'indique, elles ont la particularité d'être toutes basées sur une notion de centre de masse, placés aléatoirement à l'initialisation, qui vont idéalement se déplacer graduellement vers les centres de masse réellement présents dans le jeu de données. Ces méthodes, qui sont non-paramétriques, supposent que les groupes sont concentriques, c'est à dire semblables à un amas gaussien.

Les méthodes dites à regroupement hiérarchiques [83] produisent des hiérarchies qui mettent en évidence la force des liens entre les différentes observations ce qui permet ensuite d'en extraire des partitions adéquates. Elles peuvent être ascendantes ou descendantes. Les ascendantes partent de l'état initial où chaque observation définit un cluster puis le nombre de clusters est ensuite réduit de manière itérative en fusionnant les groupes les plus proches jusqu'à arriver à un seul groupe. Ensuite revient le choix de l'expert (ou de l'algorithme lui-même) de choisir quelles sont les meilleures partitions parmi l'ensemble des possibles fournies par la hiérarchie. La représentation de la hiérarchie se fait très souvent sous la forme d'un dendrogramme (Voir **figure 8**). Les descendantes suivent le même principe mais partent d'un seul cluster auquel toutes les observations appartiennent initialement pour le subdiviser progressivement. Les méthodes appartenant à cette catégorie diffèrent à la fois sur le choix de la métrique (ou mesure de dissimilarité) et sur la stratégie d'agglomération ou d'élagage. L'un des meilleurs représentants actuels est HDBSCAN [84].



**Figure 8** . Exemple de hiérarchie sous forme standard de dendrogramme. La section horizontale indique à quel niveau la hiérarchie les subdivisions ne comptent plus, ce qui produit dans cet exemple trois partitions.

On peut distinguer les algorithmes de mixtures de lois [85] qui sont de nature probabiliste et paramétriques. Ces algorithmes supposent que les clusters suivent chacun une loi particulière dont l'objectif est de déterminer les paramètres. Dans le cas de mixtures gaussiennes par exemple, il s'agira de déterminer les paramètres moyennes  $\mu_i$  et variances  $\sigma_i$  de chaque groupe  $i$ . Dans cette catégorie on retrouve entre autres l'algorithme EM [86, 87], la « bottleneck information method » [88] et les algorithmes bayésiens basés sur les processus de Dirichlet [89]. Il existe la famille des algorithmes basés sur une estimation non-paramétrique de la densité qui considèrent les clusters comme étant des zones denses séparées par des régions de relativement faible densité. On retrouve dans cette catégorie une importante variété d'algorithmes tant les modélisations sont diverses. Les densités estimées, ou des mesures reflétant la densité locale, peuvent être obtenues de plusieurs façons : avec

des estimateurs à noyau, en découpant le domaine en une grille et en supposant chaque cellule uniforme, ou encore grâce aux k plus proches voisins. De plus les densités obtenues peuvent être exploitées de manières très différentes. Ces approches à densité ont comme particularité de souvent pouvoir détecter et reconstruire de manière effective les clusters de forme arbitraire.

### IV.3. K-means

Avec ses 60 ans d'ancienneté, k-moyennes, mais plus souvent appelé k-means [79], est l'un des plus connus et des plus anciens algorithmes de classification automatique non-supervisée. Cette heuristique est vouée à résoudre le problème de partitionnement modélisé de la manière suivante.

Soit  $(x_1, \dots, x_N)$  l'échantillon constitué de N points que nous voulons partitionner. Pour un paramètre k fixé par l'utilisateur, l'algorithme k-means cherche à partitionner les observations  $x_i$  en k ensembles  $C = \{C_1, \dots, C_k\}$  de manière optimale, au sens qu'il minimise la formule :

$$\mathop{\text{argmin}}_C \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \quad (1)$$

Où  $\mu_j$  correspond à la moyenne du  $j^{\text{ème}}$  cluster  $C_j$ .

Pour résoudre ce problème grâce à l'heuristique k-means on commence par choisir le nombre de groupes k et initialiser les moyennes correspondantes  $\{\mu_1^0, \dots, \mu_k^0\}$  de manière aléatoire. Puis la procédure suivante est répétée jusqu'à convergence, c'est-à-dire lorsque les itérations n'améliorent plus la solution. Cela consiste en deux étapes successives à savoir la réaffectation de chaque observation au cluster qui lui est le plus proche

$$C_j^{t+1} = \left\{ x_i : \|x_i - \mu_j^t\| \leq \|x_i - \mu_l^t\|, \forall l \in \{1, \dots, k\} \right\} \quad (2)$$

Suivi d'une mise à jour des moyennes en fonction des nouvelles affectations

$$\mu_j^{t+1} = \frac{1}{|C_j^t|} \sum_{x_i \in C_j^t} x_i \quad (3)$$

L'initialisation des moyennes joue un rôle important sur les résultats en sortie. Si elles sont mal initialisées, on risque de faire converger l'algorithme sur un minimum local potentiellement éloigné du

minimum global. La manière basique (méthode de Forgy) est d'initialiser les k moyennes initiales à k données d'entrée choisies aléatoirement. Des variantes de l'algorithme de base ont donc été proposés comme k-means++ [90] dont la stratégie d'initialisation est également de choisir aléatoirement mais en abaissant les chances de choisir un point proche des centres déjà initialisés par exemple.

Les moyennes peuvent également être remplacées par des médianes [91] ou encore par la somme de fonctions variées comme l'algorithme Fuzzy C-means [92] où l'appartenance n'est plus exclusive mais traduite par un degré d'appartenance. Les distances aux moyennes aussi peuvent être remplacées par une probabilité d'appartenance au groupe qui est supposé gaussien, ce qui donne lieu aux algorithmes de type Espérance-Maximisation EM. Cet algorithme et ses variantes sont très efficaces lorsque les données sont réparties en amas gaussiens mais leur efficacité s'effondre dès que les clusters sont de forme arbitraires, particulièrement pour des clusters non-concentriques. Aussi, le comportement de cet algorithme peut énormément varier selon le paramètre k qui est bien souvent difficile à déterminer compte tenu du cadre non-supervisé. Cependant la complexité de cet algorithme est exemplaire lorsqu'elle utilise les arbres k-d puisqu'elle se trouve ainsi égale à  $O(kTN)$  où T est le nombre d'itérations et N le nombre d'observations. Ainsi, bien implémentée, cette heuristique se trouve être une des plus rapides ce qui participe en partie à sa renommée et celle de ses variantes.

#### IV.4. K-médoïdes

Le partitionnement autour des médoïdes est une méthode de classification similaire à la méthode k-Means, en utilisant les médoïdes au lieu des moyennes. De plus, contrairement à l'algorithme k-Means où les centres peuvent ne pas être parmi les points des données, la méthode des médoïdes représente les classes par des points appartenant nécessairement aux données (médoïdes).

Un médoïde est le point d'un ensemble qui minimise la somme des distances entre lui et chacun des autres points de ce même ensemble.

$$\mathbf{M} = \underset{m}{\operatorname{argmin}} \sum_{i=1}^n d(x_i, m) \quad (4)$$

L'algorithme k-Medoids utilise une fonction objective qui définit la distance moyenne entre un objet et le médoïde. Comme avantages de cet algorithme par rapport à k-Means, on cite : Il s'adapte à n'importe quel type de données ; Il est insensible aux objets isolés. Le K-médoïde [93] est le point situé à l'intérieur d'un cluster qui peut représenter ce dernier au mieux. La représentation par les K-médoïdes ne dépend pas du type d'attribut. Le choix des médoïdes dépend de la concentration des points à l'intérieur d'un cluster, de ce fait, cette méthode est moins sensible à la présence de bruits.

$$E = \sum_{k=1}^K \sum_{x \in C_k} d(x, m_k) \quad (5)$$

Une fois les médoïdes choisis, les clusters sont définis comme des sous-ensembles de points proches de leurs médoïdes respectifs, et la fonction objectif est définie comme la distance moyenne ou une autre mesure de dissimilarité entre un point et son médoïde.

Nous choisissons la norme de Manhattan. Les distances n'apparaissant pas au carrée, cet algorithme est moins sensible aux outliers que le k-means. Cependant, sa programmation est plus délicate car dans le cas de la médiane spatiale il n'existe pas de formule analogue à celle de la moyenne. La première implémentation de l'algorithme est celle de Kaufman et Rousseeuw (1987), PAM pour Partitioning Around Medoids. Elle est assez coûteuse, mais donne de bons résultats et reste simple à écrire. Diverses heuristiques ou accélérations autour de cette implémentation ont été développés (voir Chu et al. 2002 par exemple), mais nous utilisons ici la version originale, qui se révèle suffisante.

Pour améliorer la qualité de clustering de L'algorithme k-Medoids, de nombreuses variantes s'y sont succédées, entre autres les algorithmes PAM, CLARA et CLARANS. Ces algorithmes adoptent le même principe de représentation par médoïde et utilisent la même fonction objectif donnée ci-dessus.

### IV.4.1 L'algorithme PAM

L'idée de cet algorithme consiste à commencer avec un ensemble de k médoïdes puis échanger le rôle entre un objet médoïde et un non-médoïde si cela permet de réduire la distance globale, ce qui revient à minimiser la fonction objective. L'inconvénient majeur de cet algorithme est son coût total de calcul, ceci le rend non adaptable pour une population importante d'objets.

**Tableau 4 :** Algorithm PAM

<p><b>Inputs :</b>          Ensemble de points <math>(x_i)_{1 \leq i \leq n}</math>          K nombre de classes, fixé.</p>
<p><b>Initialize :</b>          Sélection de K médoïdes <math>(m_1, \dots, m_k)</math> pris au hasard parmi les individus.          Chaque point est affecté au médoïde le plus proche au sens de la distance de Manhattan.</p>
<p><b>For k = 1 to K do</b> Sélectionner au hasard un point i non médoïde.          Calculer de nouveau le critère de dispersion si on considère ce point comme nouveau médoïde.          Si E diminue, le point i est gardé comme le nouveau médoïdes.          Répéter jusqu'à convergence.</p>

### IV.4.2 L'algorithme CLARA

L'algorithme CLARA (Clustering LARge Application) a été mise en œuvre par Kaufman et Rousseeuw [94] dans le but de réduire le coût de calcul de PAM. Cet algorithme travaille sur des échantillons. On prend une petite partie d'objets (échantillon). Ensuite, les k médoïdes sont déterminés en appliquant PAM sur cet échantillon. Si cet échantillon est choisi d'une manière aléatoire, alors il représente bien tous les objets, donc les médoïdes sont similaires à ceux qui sont créés à partir de tous les objets. L'algorithme est, généralement, exécuté sur plusieurs échantillons pour obtenir le meilleur résultat. Les auteurs ont indiqué, suite à leurs expérimentations, que la taille d'un échantillon de  $(40+2k)$ , k étant le nombre de classes, donne un bon résultat. Le principal inconvénient de l'algorithme CLARA, outre les difficultés d'échantillonnage, est que si un objet qui

serait le meilleur médoïde n'apparaît dans aucun échantillon, alors la meilleure solution ne pourra jamais être atteinte.

### IV.4.3 L'algorithme CLARANS

CLARANS (Clustering Large Application based on RANdomized Search) a été mis en place par Raymon Ng et Jiawei Han [95]. Les auteurs de cet algorithme ont utilisé un graphe pour résoudre le problème de recherche des  $k$  médoïdes. Chaque nœud dans le graphe est un ensemble de  $k$  objets médoïdes et possède  $k*(n-k)$  voisins. Un nœud voisin s'obtient en échangeant les rôles entre le médoïde et un non-médoïde.

Une fois le graphe construit, PAM parcourt tous les nœuds voisins à la recherche de la solution, ce qui rend le coût de traitement trop cher. CLARA, à son tour, ne parcourt qu'un sous-graphe de l'ensemble des voisins, il travaille donc sur une région localisée. CLARANS est une combinaison des deux concepts, il choisit aléatoirement un nombre de nœuds voisins (ce nombre est donné en paramètre) et se déplace d'un nœud au nœud adjacent qui minimise la fonction objective. L'algorithme s'arrête quand aucun nœud adjacent ne peut optimiser la fonction objective. Cet algorithme de clustering a tous les avantages de PAM et de CLARA avec une complexité linéaire par rapport au nombre d'objets. D'ailleurs, l'expérimentation des auteurs a montré qu'il fournit toujours les meilleurs résultats par rapport au PAM et CLARA dans tous les cas.

## V. Pénalisation – Comment choisir parmi un grand nombre de variables

Nous disposons d'un ensemble de variables explicatives  $(X_i)_{1 \leq i \leq p}$  pour expliquer une variable  $Y$  de manière linéaire et rien ne nous assure que toutes les variables interviennent dans l'explication. Nous avons donc à notre disposition un ensemble de variables potentiellement explicatives ou variables candidates. Notre objectif est de déterminer quelles sont les variables réellement explicatives. Il s'agit donc de choisir un modèle parmi les  $2^p$  possibles. Comment choisir le bon modèle ? Tous les étudier n'est tout simplement pas envisageable dès que  $p$  est grand, qui plus est, il faudrait savoir déterminer si un modèle est meilleur que les autres. La méthode du Lasso propose dans certains cas une solution à ce problème. Nous allons premièrement en expliquer le principe, puis en donner l'algorithme avant de voir quand est-ce qu'elle est efficace.

### V.1. La méthode du LASSO - Least Absolute Shrinkage and Selection Operator

### V.1.1. L'idée

Dans cette partie nous allons expliquer l'idée de la méthode. Nous cherchons à expliquer de manière linéaire une variable  $Y$ , par  $p$  variables potentiellement explicatives  $X_i$ . Pour cela nous faisons  $n$  observations. Nous modélisons la variable  $Y$  de la manière suivante :

$$Y^n = X^n \beta^n + \varepsilon^n \quad (6)$$

Tel que :  $\varepsilon^n = (\varepsilon_1, \dots, \varepsilon_n)^T$  est un vecteur de  $n$  variables aléatoires i.i.d. de moyenne 0 et de variance  $\sigma^2$  correspond au bruit lors des observations (qui peut contenir toutes les variables explicatives non prises en compte dans le modèle) ;  $Y^n \in \mathbf{R}^n$  correspond aux  $n$  observations de la variable  $Y$  et :

$$X^n = (X_{\cdot,1}^n, \dots, X_{\cdot,p}^n) = \left( (X_1^n, \dots, X_n^n)^T, \dots, (X_p^n, \dots, X_n^n)^T \right)^T \quad (7)$$

(7) est une matrice  $n \times p$ , où  $X_{\cdot,j}^n$  est la  $j^{eme}$  colonne, qui correspond au  $j^{eme}$  prédicteur  $X_j$  et  $X_{i,\cdot}^n$  la  $i^{eme}$  ligne qui correspond à la  $i^{eme}$  observation. le paramètre  $\beta$  (avec  $\beta^n \in \mathbf{R}^p$ ) est à estimer, il est indexé par  $n$  pour permettre à ses coefficients et à sa taille de varier lorsque  $n$  augmente ( $p$  peut a priori dépendre de  $n$ ).

Les variables  $X_i$  n'étant pas toutes pertinentes, l'objectif est d'éliminer les variables inutiles et uniquement celles-ci. L'idée du Lasso est donc non pas de faire une régression linéaire classique mais une régression régularisée qui rend nuls certains coefficients de l'estimation de  $\beta$ .

Cela consiste à estimer pour  $\lambda \in \bar{\mathbf{R}}_+$  :

$$\hat{\beta}^n(\lambda) = \arg \min_{\beta \in \mathbf{R}^p} \left( \frac{1}{2} \|Y^n - X^n \beta\|_2^2 + \lambda \|\beta\|_1 \right) \quad (8)$$

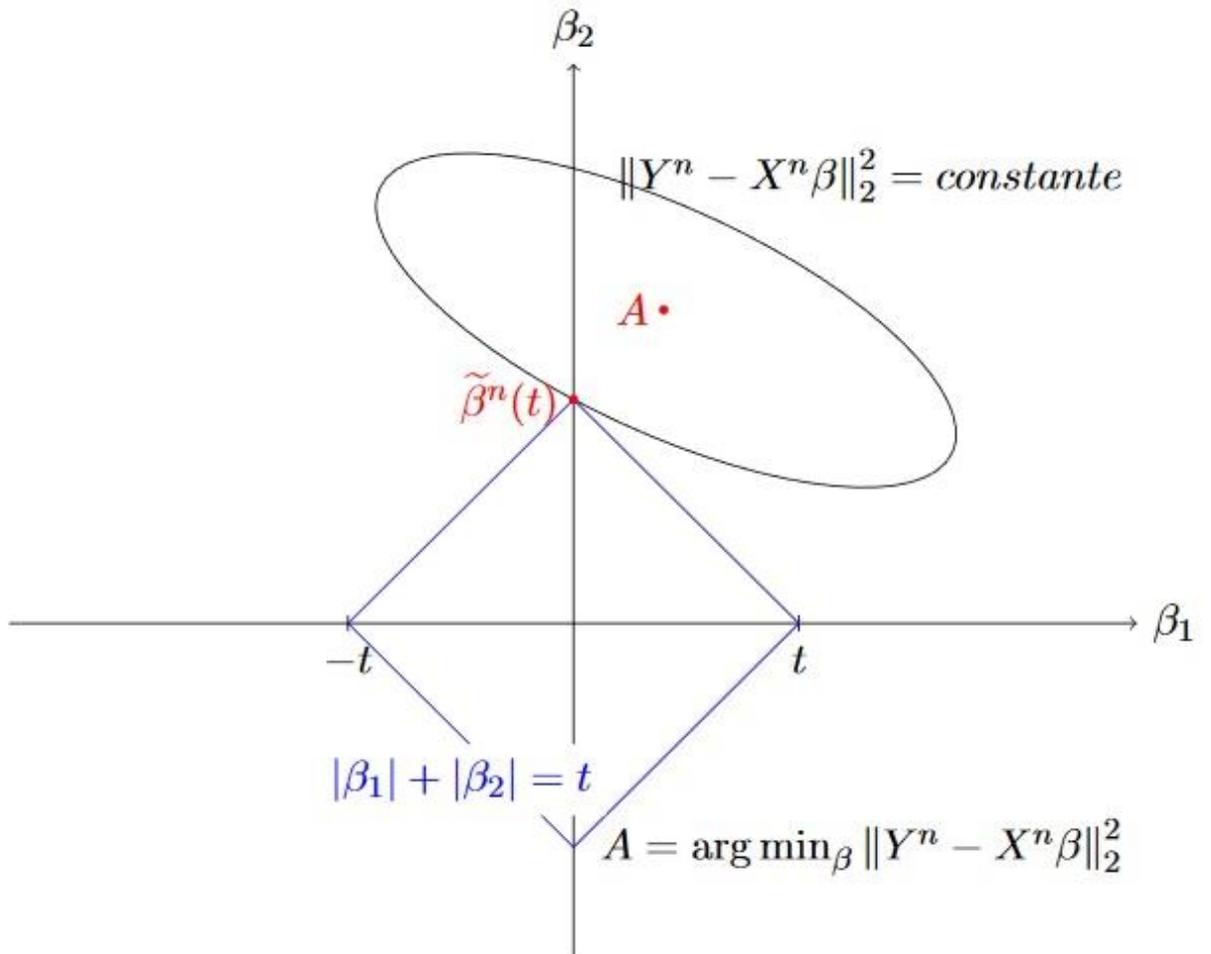
Où  $\|x\|_2^2 = \sum_{i=1}^n x_i^2$  et  $\|x\|_1 = \sum_{i=1}^p |x_i|$ .

Le paramètre  $\lambda \geq 0$  contrôle la puissance de la régularisation. Si on prend  $\lambda = 0$ , le Lasso correspond à une régression linéaire classique (si  $p \leq n$ ). Si par contre, on prend  $\lambda = \infty$ , tous les coefficients de  $\hat{\beta}^n(\infty)$  sont nuls. L'augmentation de  $\lambda$  induit la diminution de certains coefficients de  $\hat{\beta}^n(\lambda)$  vers 0 jusqu'à ce qu'ils soient exactement nuls. On peut montrer que ce modèle est équivalent au suivant :

$$\tilde{\beta}^n(t) = \arg \min_{\beta, \|\beta\|_1 \leq t} (\|Y^n - X^n \beta\|_2^2) \quad (9)$$

Dans le sens où pour tout  $\lambda \in \bar{\mathbb{R}}_+$ , il existe  $t \geq 0$  tel que  $\tilde{\beta}^n(t) = \hat{\beta}^n(\lambda)$ . En effet, il suffit de prendre  $t = \|\hat{\beta}^n(\lambda)\|_1$  alors pour tout  $\beta$  tel que  $\|\beta\|_1 \leq t$ ,  $\lambda \|\beta\|_1 \leq \lambda \|\hat{\beta}^n(\lambda)\|_1$  donc par définition de  $\hat{\beta}^n(\lambda)$ ,  $\|Y^n - X^n \beta\|_2^2 \geq \|Y^n - X^n \hat{\beta}^n(\lambda)\|_2^2$ .

Cela nous permet de comprendre intuitivement pourquoi le Lasso, dans la plupart des cas, rend exactement nuls certains coefficients de  $\hat{\beta}^n(\lambda)$ . **La figure 9** illustre en dimension 2 le cas où le Lasso annule une coordonnée. Cela dépend de la position de  $A = \arg \min_{\beta} (\|Y^n - X^n \beta\|_2^2)$  dans  $R^p$  et de  $t$ . Plus  $t$  est petit plus le Lasso annule de coordonnées. Lorsque la dimension augmente, la zone -dans laquelle A annulé au moins une coordonnée- augmente au point de devenir presque l'espace  $R^p$  tout entier.



**Figure 9.** Cas où A se trouve dans la zone qui annule le coefficient  $\beta_2$

L'algorithme du Lasso consiste donc à déterminer  $\hat{\beta}^n(\lambda)$  pour tout  $\lambda \geq 0$ , nous exposerons la méthode dans la partie suivante. Il s'agit ensuite de déterminer le bon  $\lambda$  qui permet de garder uniquement les vraies variables explicatives et d'éliminer les autres. Tout d'abord, l'existence d'un tel

$\lambda$  n'est pas immédiate et est même fautive en générale. Nous étudierons dans la deuxième partie quand est-ce qu'un tel  $\lambda$  existe et nous verrons que si tel est le cas, il est de l'ordre de  $\sqrt{n}$ .

### V.1.2. L'algorithme

Dans cette partie, on s'intéresse au problème du Lasso suivant. Déterminer pour tout  $\lambda \geq 0$ ,

$$\hat{\beta}(\lambda) = \arg \min_{\beta \in \mathbb{R}^p} \left( \frac{1}{2} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right) \quad (10)$$

Un tel problème peut a priori admettre plusieurs solutions.

❖ **Lemme 1** (Unicité). Si  $X^T X$  est définie positive, le problème (10) admet une unique solution.

Nous nous placerons à partir de maintenant dans le cas particulier où  $X^T X$  est définie positive ce qui permettra d'affirmer que la solution au problème du Lasso est unique. Une version modifiée de l'algorithme (least-angle regression) LARS [96, 97] permet alors d'en calculer la solution. Celui-ci repose sur le lemme de « la condition d'optimalité ».

❖ **Lemme 2** (Condition d'optimalité).  $\beta$  est solution de (10) si et seulement si :

$$\forall j = 1 \dots p, \begin{cases} |X_j^T(Y - X\beta)| \leq \lambda & \text{si } \beta_j = 0 \quad (L1) \\ X_j^T(Y - X\beta) = \lambda \text{ sign}(\beta_j) & \text{si } \beta_j \neq 0 \quad (L2) \end{cases} \quad (11)$$

Où  $X_j$  désigne la  $j^{\text{ème}}$  colonne de  $X$ .

Nous allons essayer de chercher une solution potentielle au problème du Lasso à partir de ce lemme. Pour  $s \in \{-1, 0, 1\}^p$ , on pose  $\Gamma_s$  l'ensemble d'indices  $j$  tel que  $s_j \neq 0$ ,  $X_{\Gamma_s}$  la matrice composée des colonnes de  $X$  correspondant à l'ensemble  $\Gamma_s$ , ainsi que  $s_{\Gamma}$  le vecteur  $s$  auquel on a retiré les coordonnées nulles. On peut alors définir pour tout  $\lambda > 0$ ,

$$\begin{cases} \beta_i^s(\lambda) = 0 & \text{si } s_j = 0 \\ \beta_{\Gamma_s}^s(\lambda) = (X_{\Gamma_s}^T X_{\Gamma_s})^{-1} (X_{\Gamma_s}^T Y - \lambda s_{\Gamma}) & \text{sinon} \end{cases} \quad (12)$$

❖ **Lemme 3**. Si  $s = \text{sign}(\beta^s(\lambda))$  alors  $\beta^s(\lambda)$  vérifie (L2) pour tout  $j$  tel que  $s_j \neq 0$ .

Nous connaissons donc une solution potentielle au problème du lasso en supposant son signe  $s$  connu. On peut remarquer que cette solution est linéaire. L'idée de l'algorithme de LARS (least-angle regression) repose dessus. On fait en sorte qu'on ait toujours  $s = \text{sign}(\beta^s(\lambda))$  et que la fonction  $\beta^s(\lambda)$  vérifie la condition (L1), on obtient ainsi une fonction linéaire par morceaux  $\beta_{LARS}$ . Notons  $(\lambda_i)_{i \geq 0}$  la famille décroissante avec  $\lambda_0 = \infty$ , telle que  $\beta_{LARS}$  est linéaire exactement (on ne peut pas les agrandir) sur les intervalles  $[\lambda_{i+1}, \lambda_i]$ . Pour tout  $i \geq 0$ ,  $\beta_{LARS} = \beta^{s_i}$ . La difficulté consiste donc à déterminer la famille  $(\lambda_i)$  et mes vecteurs de signes  $s_i$  associés.

Soit  $i \geq 1$ . Nous supposons  $\lambda_i$  et  $s_{i-1}$  connues. Donnons la méthode permettant de calculer  $s_i \in \{-1, 0, 1\}^p$ . Admettons que pour  $\lambda < \lambda_i$ , la fonction  $\beta^{s_{i-1}}(\lambda)$  ne convient plus. Il existe donc deux possibilités.

1. Soit  $\text{sign}(\beta^{s_{i-1}}(\lambda)) \neq s_{i-1}$ . Cela signifie qu'une des coordonnées disons la  $j^{\text{ème}}$  de  $\beta^{s_{i-1}}(\lambda)$  change de signe donc s'annule en  $\lambda = \lambda_i$ . On pose donc  $\varepsilon_i = 0$ .

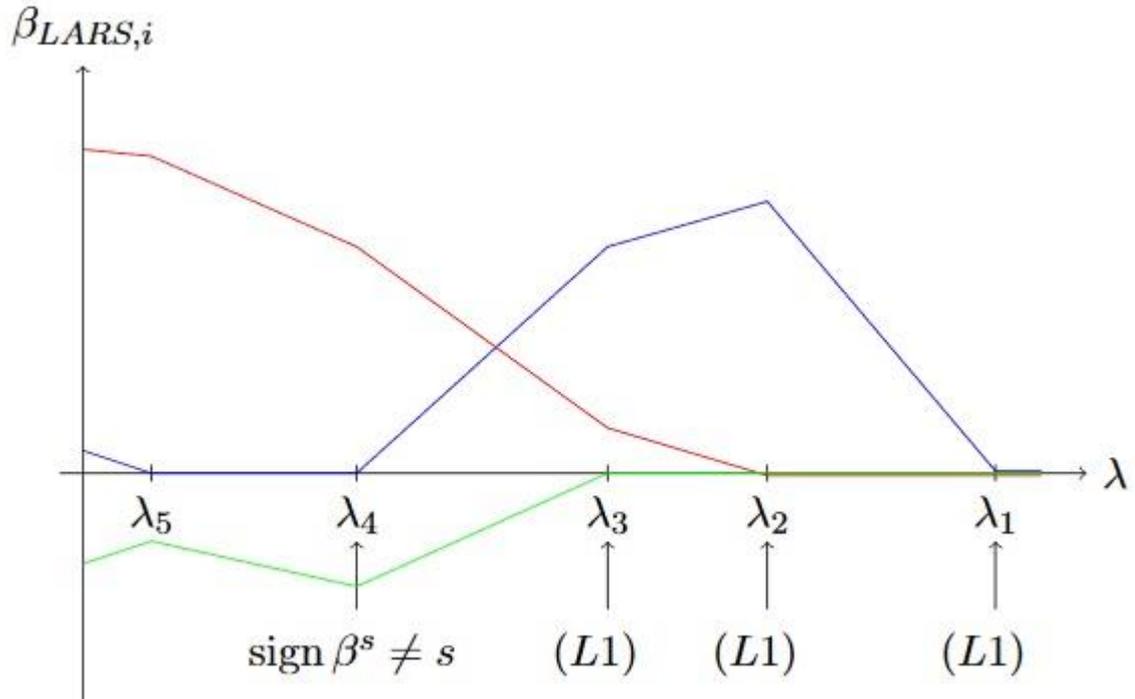
2. Soit (L1) n'est plus vérifiée pour un  $j$  tel que  $s_j = 0$ . On pose  $\varepsilon_i = \text{sign}(X_{\Gamma_{s_{i-1}}}^T (Y - X\beta^{s_{i-1}}(\lambda_i)))$ .

On définit alors  $s_i = s_{i-1}$  où l'on remplace la  $j^{\text{ème}}$  coordonnée par  $\varepsilon_i$ .

Proposons maintenant l'idée de l'algorithme de LARS pour calculer une fonction  $\beta_{LARS}$  sur  $\bar{\mathbb{R}}_+$  qui serait solution du problème du Lasso. Pour  $\lambda = \infty$ , la solution du problème du Lasso étant trivialement le vecteur nul on commence par-là puis on fait décroître  $\lambda$  au cours de l'algorithme.

- ✓ **Initialisation** : on pose  $\lambda_0 = \infty$ ,  $s_0 = 0_{R^p}$  et  $\beta_{LARS}(\lambda_0) = \beta^{s_0}(\lambda_0) = 0_{R^p}$ .
- ✓ **Étape 1** : on fait décroître  $\lambda$  en partant de  $\lambda_0$  jusqu'à ce qu'une coordonnée  $j$  de  $\beta_{LARS}(\lambda) = \beta^{s_0}(\lambda)$  ne vérifie plus (L1). On définit  $\lambda_1$  comme cet instant. La coordonnée  $j$  de  $\beta_{LARS}(\lambda)$  n'est donc plus nulle pour  $\lambda < \lambda_1$ , on calcule son signe  $\varepsilon_1 \in \{-1, 1\}$  comme vu précédemment et on pose  $s_1 = s_0$  dans lequel on a remplacé la  $j^{\text{ème}}$  coordonnée par  $\varepsilon_1$ .
- ✓ **Étape  $i \geq 2$**  : on fait décroître  $\lambda$  à partir de  $\lambda_i$ , en calculant  $\beta_{LARS}(\lambda) = \beta^{s_i}(\lambda)$  et on s'arrête soit quand une coordonnée  $j$  telle que  $(s_i)_j \neq 0$  s'annule, car alors  $\text{sign}(\beta^{s_i}) \neq s_i$  (en  $\lambda_4$  sur **la figure 10**), soit quand la condition (L1) n'est plus satisfaite pour une coordonnée telle que  $(s_i)_j = 0$  (en  $\lambda_2$  sur **la figure 10**). On note  $\lambda_{i+1}$  cet instant et on définit le nouveau vecteur de signes  $s_{i+1}$ .

On calcule ainsi la suite finie  $(\lambda_i)_{i \geq 0}$  jusqu'à ce que  $\lambda$  soit nul. De cette manière, nous avons calculé une fonction  $\beta_{LARS}$  qui devrait être solution du problème Lasso.



**Figure 10.** Calcul de la fonction  $\beta_{LARS}$

❖ **Lemme 4.** (Validité de  $\beta_{LARS}$ ). Tant que  $\lambda \in ]\lambda_{i+1}, \lambda_i[$ , alors (L1) est vérifiée pour tout  $j$  tel que  $(s_i)_j = 0$  et  $\text{sign}(\beta^{s_i}(\lambda)) = s_i$ . De plus, la suite  $(s_i)$  est bien définie.

On déduit de ce lemme et du précédent que la fonction  $\beta_{LARS}$  calculée par l'algorithme de LARS-Lasso est bien la solution du problème du Lasso. On peut cependant faire deux remarques. La première est que l'algorithme termine. C'est-à-dire qu'à partir d'un certain rang  $\lambda_i = 0$ . Cela provient du fait qu'il n'existe que  $3^p$  fonctions  $\beta^s$  différentes et chacune d'entre elles ne vérifie les conditions du lemme précédent que sur un intervalle. L'algorithme termine donc en au plus  $3^p$  itérations. Remarquons que  $3^p$  peut-être gigantesque dès que  $p$  est grand, en pratique l'algorithme termine bien plus rapidement. Ce résultat est donc purement théorique. La deuxième est que dans la définition de  $s_i$ , nous avons supposé qu'une seule des coordonnées était insatisfaisante en  $\lambda_i$ . Cependant cela arrive en général avec probabilité 1, nous restreignons donc ici notre étude à ce cas.

Nous pouvons maintenant énoncer l'algorithme LARS-Lasso, qui calcule la fonction  $\beta_{LARS}$  décrite ci-dessus.

Nous pouvons éviter les redondances de certains calculs :  $(X_\Gamma^T X_\Gamma)^{-1}, X_\Gamma^T Y$ , unifier les calculs des vecteurs de  $((X_\Gamma^T X_\Gamma)^{-1}(X_\Gamma^T Y))_{j'}$  et de  $((X_\Gamma^T X_\Gamma)^{-1}(\text{sign}(\beta_\Gamma(\lambda)))_{j'}$ . En pratique, nous implémenterons donc **l'algorithme 2 (LARS-Lasso Amélioré)**.

Le Lasso est donc une méthode pour choisir parmi un grand nombre de variables. C'est une méthode utilisée en pratique dans des domaines comme la biologie et la santé publique où le nombre de variables potentiellement explicatives est important et où nous ne pouvons pas toujours faire suffisamment d'observations. Nous avons dans cette section montrée que sous certaines hypothèses le Lasso est consistant en signes si et seulement si les conditions d'irreprésentabilité sont vérifiées que  $p$  et  $q$  soient petits ou grands. Mais en pratique, sont-elles souvent vérifiées ? Nous avons vu que la condition de forte irreprésentabilité est vérifiée dans des cadres assez spéciaux, et nous avons vu qu'il arrive qu'elle soit fautive. Nous pourrions donc nous demander à quel point cette condition est forte, afin de savoir si le Lasso est fréquemment ou rarement consistant. C'est à dire, si nous prenons  $p$  et  $q$  donnés, avec quelle probabilité un modèle choisi aléatoirement vérifie-t-il cette condition ? Nous verrions que pour  $q=1$ ,  $CI+$  est en général vérifiée mais dès que  $p$  et  $q$  ne sont plus très petits, les conditions d'irreprésentabilité ne sont vérifiées que très rarement. De plus, même lorsque par chance nous nous trouvons dans ce cas, les résultats montrés précédemment ne sont qu'asymptotiques. Il arrive que le Lasso ne puisse techniquement pas choisir le bon modèle en un temps raisonnable. Il existe de nombreuses améliorations du Lasso qui permettent de résoudre ces problèmes. Leur étude est un domaine encore très actuel.

**ENTRÉES:**  $X, Y, n, p$

$i \leftarrow 1$

$\lambda \leftarrow +\infty$

$\beta \leftarrow 0$

$\text{sign}(\beta) \leftarrow 0$

5:  $\Gamma \leftarrow \emptyset$

Resultats  $\leftarrow \emptyset$

**Tant que**  $\lambda > 0$  **faire**

« On commence par chercher la première variable insatisfaisante »

$\lambda_{\max} \leftarrow 0$

10:  $j_{\max} \leftarrow 0$

**Pour**  $j = 1, \dots, p$  **faire**

**Si**  $\text{sign}(\beta)_j = 0$  **alors**

$$\lambda_{\text{test}} \leftarrow \max \left\{ \frac{X_j^T [X_\Gamma (X_\Gamma^T X_\Gamma)^{-1} (X_\Gamma^T Y) - Y]}{[X_j^T X_\Gamma (X_\Gamma^T X_\Gamma)^{-1} (\text{sign}(\beta_\Gamma(\lambda)))] \pm 1} \right\}$$

**Sinon**

15:  $\lambda_{\text{test}} \leftarrow \left( \frac{((X_\Gamma^T X_\Gamma)^{-1} (X_\Gamma^T Y))_{j'}}{((X_\Gamma^T X_\Gamma)^{-1} (\text{sign}(\beta_\Gamma(\lambda))))_{j'}} \right)$

**Fin si**

**Si**  $\lambda_{\text{test}} < \lambda$  **et**  $\lambda_{\text{test}} > \lambda_{\max}$  **alors**

$\lambda_{\max} \leftarrow \lambda_{\text{test}}$

$j_{\max} \leftarrow j$

20: **Fin si**

**Fin pour**

« On a dans  $(\lambda_{\max}, j_{\max})$  le résultat d'insatisfaction cherché »

$\beta \leftarrow \beta_\Gamma \leftarrow (X_\Gamma^T X_\Gamma)^{-1} (X_\Gamma^T Y - \lambda_{\max} \text{sign}(\beta_\Gamma))$

$\lambda \leftarrow \lambda_{\max}$

25: **Si**  $j_{\max} > 0$  **alors**

**Si**  $\text{sign}(\beta)_{j_{\max}} = 0$  **alors**

$\text{sign}(\beta)_{j_{\max}} \leftarrow \text{sign}(X_{j_{\max}}^T (Y - X\beta))$

**Sinon**

$\text{sign}(\beta)_{j_{\max}} \leftarrow 0$

30:  $\beta_{j_{\max}} \leftarrow 0$

**Fin si**

$\Gamma \leftarrow \text{non-nuls}(\text{sign}(\beta))$

Resultats  $\leftarrow \text{Resultats} \cup \{(j_{\max}, \lambda_{\max}, \beta)\}$

**Sinon**

35: Resultats  $\leftarrow \text{Resultats} \cup \{(\lambda_{\max}, \beta)\}$

**Fin si**

$i \leftarrow i + 1$

**Fin tant que**

**Retourner** Resultats

**Figure 11.** L'algorithme 1 - LARS-Lasso

---

**ENTRÉES:**  $X, Y, n, p$

$\lambda \leftarrow +\infty$   
 $\beta \leftarrow 0$   
 $\text{sign}(\beta) \leftarrow 0$   
 $\Gamma \leftarrow \emptyset$

5: **Resultats**  $\leftarrow \emptyset$   
**Tant que**  $\lambda > 0$  **faire**  
« On commence par chercher la première variable insatisfaisante ».  
 $\lambda_{\max} \leftarrow 0$   
 $j_{\max} \leftarrow 0$

10:  $U \leftarrow (X_{\Gamma}^T X_{\Gamma})^{-1}$   
 $V \leftarrow X_{\Gamma}^T Y$   
 $W \leftarrow X_{\Gamma} U$   
 $N1 \leftarrow UV$   
 $D1 \leftarrow U \text{sign}(\beta_{\Gamma})$

15:  $F1 \leftarrow F1_{\Gamma} \leftarrow N1./D1$   
**Pour**  $j = 1, \dots, p$  **faire**  
**Si**  $\text{sign}(\beta)_j = 0$  **alors**  
 $N0 \leftarrow X_j^T (WV - Y)$   
 $D0 \leftarrow X_j^T W \text{sign}(\beta_{\Gamma})$

20:  $\lambda_{\text{test}} \leftarrow \max \left\{ \frac{N0}{D0+1}, \frac{N0}{D0-1} \right\}$   
**Sinon**  
 $\lambda_{\text{test}} \leftarrow F1_j$   
**Fin si**  
**Si**  $\lambda_{\text{test}} < \lambda$  **et**  $\lambda_{\text{test}} > \lambda_{\max}$  **alors**

25:  $\lambda_{\max} \leftarrow \lambda_{\text{test}}$   
 $j_{\max} \leftarrow j$   
**Fin si**  
**Fin pour**  
« On a dans  $(\lambda_{\max}, j_{\max})$  le résultat d'insatisfaction cherché ».

30:  $\beta \leftarrow \beta_{\Gamma} \leftarrow U(V - \lambda_{\max} \text{sign}(\beta_{\Gamma}))$   
 $\lambda \leftarrow \lambda_{\max}$   
**Si**  $j_{\max} > 0$  **alors**  
**Si**  $\text{sign}(\beta)_{j_{\max}} = 0$  **alors**  
 $\text{sign}(\beta)_{j_{\max}} \leftarrow \text{sign}(X_{j_{\max}}^T (Y - X\beta))$

35: **Sinon**  
 $\text{sign}(\beta)_{j_{\max}} \leftarrow 0$   
 $\beta_{j_{\max}} \leftarrow 0$   
**Fin si**  
 $\Gamma \leftarrow \text{non-nuls}(\text{sign}(\beta))$

40: **Resultats**  $\leftarrow \text{Resultats} \cup \{(j_{\max}, \lambda_{\max}, \beta)\}$   
**Sinon**  
**Resultats**  $\leftarrow \text{Resultats} \cup \{(\lambda_{\max}, \beta)\}$   
**Fin si**  
**Fin tant que**

45: **Retourner** **Resultats**

**Figure 12.** L'algorithme 2 (LARS-Lasso Amélioré)

## **MATÉRIEL ET MÉTHODES**

Dans un premier temps, la cohorte de PALOMB et la population d'étude qui la compose seront présentées. Le suivi des patients dans la cohorte PARIS sera ensuite décrit, en s'attardant sur la modélisation des trajectoires des patients BPCO. Enfin, l'exploitation statistique des données recueillies sera détaillée.

## **I. Présentation générale du projet PALOMB**

Les porteurs du projet PALOMB – BPCO se sont donné pour missions :

1. Créer un observatoire de la BPCO en Aquitaine et en Charentes afin de mieux connaître les patients et leur prise en charge (mise en place d'une cohorte) ;
2. Mettre en place des études épidémiologiques dans un futur proche ;
3. Développer un portail Internet à destination des professionnels de santé et des patients (véritable centre de ressources sur la maladie et outil de formation médicale continue).

### **I.1. La cohorte PALOMB**

En France, actuellement il existe peu de données épidémiologiques portant sur la prévalence et la prise en charge de la BPCO, tant en médecine générale, qu'en médecine spécialisée. Dans ce contexte, la cohorte PALOMB a pour objectif de mettre en place en Aquitaine et en Charente-Maritime un observatoire régional de la BPCO, qui permette de décrire la population des patients BPCO pris en charge par les pneumologues quel que soit leur mode d'exercice (hospitalier ou libéral). La cohorte PALOMB est une initiative commune de l'Association des Pneumologues Privés du Sud-Ouest (APPSO) et des Pneumologues Hospitalo-universitaires du Service des Maladies Respiratoires du CHU de Bordeaux avec le soutien de la Fondation de Bordeaux. Le comité scientifique de ce projet est composé par les membres suivants : Dr F. Pellet, Pr C. Raheison, Dr A. Bernady, Dr J. Casteigt, Dr C. Nocent, Dr L. Falque, Dr F. Le Guillou, Dr J. Moinard, Pr M. Molimard, Dr L. Nguyen, Dr A. Ozier et Dr C. Roy.

### **I.2. Objectifs pour les professionnels de santé**

Ce projet doit permettre à un professionnel de santé, pneumologue quel que soit son mode d'exercice, ou médecin de premier recours, suivant son niveau d'intervention dans la BPCO :

1. De recenser les patients BPCO qu'il suit
2. D'en apprécier les stades de sévérité, l'âge, les comorbidités, la prise en charge
3. D'améliorer le suivi des patients BPCO
4. De connaître les recommandations médicales existantes dans la BPCO
5. De rationaliser les indications thérapeutiques-de connaître et d'avoir un accès facile au traitement plus global de la maladie chronique BPCO, à savoir la réhabilitation respiratoire, l'éducation thérapeutique, le sevrage tabagique.

### **I.3. Objectif d'évaluation professionnelle**

Cet outil permet également aux professionnels de santé de réaliser une évaluation personnelle de leur activité sur le thème de la BPCO et de s'engager ainsi dans des programmes de DPC (développement professionnel continu).

## **I.4. Objectif de santé publique**

Ce programme permet de décrire la population de BPCO suivis par les pneumologues et leur prise en charge. Cette étude a fait l'objet d'une déclaration auprès de la Commission Nationale de l'Informatique et des Libertés (CNIL) en effet, compte tenu des nécessités de la recherche et de son analyse ultérieure, les données recueillies concernant les patients ont été informatisées dans des conditions de façon à garantir leur confidentialité. Tout médecin souhaitant participer à la cohorte devait s'inscrire pour ensuite enregistrer ses patients présentant une BPCO, les données simples concernant le diagnostic, le bilan, la prise en charge thérapeutique et les objectifs fixés avec le patient. Cette Plateforme permet au professionnel de santé de visualiser son activité, éventuellement de s'évaluer dans le domaine de la BPCO et également lui apporter une aide pratique par la diffusion des recommandations.

## **II.1. Suivi longitudinal**

Le suivi des patients de la cohorte PALOMB repose sur le recueil régulier de données cliniques, fonctionnelles, les comorbidités, les traitements inhalés, les vaccinations, le sevrage tabagique et la réhabilitation respiratoire. Ces données sont notamment relatives à la symptomatologie respiratoire des patients et à l'évolution de leur mode et cadre de vie.

Après la première visite clinique d'inclusion, le suivi des patients inclus dans la cohorte PALOMB est basé sur des questionnaires adressés régulièrement par les pneumologues en utilisant la plateforme du projet PALOMB ( [www.palomb.fr](http://www.palomb.fr) ) aux dates de point choisis par les pneumologues (15 jours ; 1 mois ; 3 mois ; 6 mois ; 1 an ou autre).

## **II.2. Aspects éthiques**

L'observatoire doit permettre initialement aux pneumologues puis vraisemblablement secondairement aux médecins traitants, puis dans le futur à l'ensemble des professionnels de santé concernés par la BPCO, de recenser des patients de façon anonyme, suivant des critères cliniques, paracliniques et thérapeutiques. Ces critères ont fait l'objet d'une déclaration auprès de la Commission Nationale de l'Informatique et des Libertés (CNIL) en effet, compte tenu des nécessités de la recherche et de son analyse ultérieure, les données recueillies concernant le patient touché par cette maladie seront informatisées dans des conditions de façon à garantir leur confidentialité.

La CNIL et le CCTIRS (Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé) ont approuvé l'étude, et un consentement éclairé a été obtenu avant l'inclusion. Les responsables du projet PALOMB avaient demandé au comité d'éthique local un retour sur la nécessité d'une autorisation éthique pour une telle analyse rétrospective et ont été informés que cela n'était pas justifié.

Le médecin voulant participer à ce projet doit tout d'abord s'inscrire et il pourra ensuite enregistrer ses patients présentant une BPCO, des données simples concernant le diagnostic, le bilan, la prise en charge thérapeutique et les objectifs fixés avec le patient. Cette Plateforme permet au professionnel de santé de visualiser son activité, éventuellement de s'évaluer dans le domaine de la BPCO et également lui apporter une aide pratique par la diffusion des recommandations. Elle doit permettre aussi le lien avec les prises en charge autres que médicamenteuses, par le recensement des possibilités d'éducation thérapeutique, de sevrage tabagique, de réhabilitation respiratoire dans son environnement pour orienter au mieux son patient.

## **II.3. Evaluation de l'état de santé des patients BPCO de la cohorte PALOMB**

L'état de santé jusqu'à 6 ans de suivi des patients BPCO de la cohorte PALOMB a été évalué grâce à l'ensemble des données recueillies depuis l'inclusion par les questionnaires régulièrement renseignés par les pneumologues et grâce aux données obtenues lors des différentes visites de suivi.

### **II.3.1. Les données cliniques**

Les informations cliniques relatives à la première visite d'inclusion ont été collectées par le biais des questionnaires de la plateforme PALOMB. Cette section est subdivisée en trois sous-sections.

Premièrement, les pneumologues commencent par collecter la taille et le poids afin de pouvoir mesurer l'indice de masse corporelle, ensuite ils mesurent la tension artérielle du patients (soit maximale ou minimale) en millimètres de mercure (mmHg), la fréquence cardiaque c'est-à-dire le nombre de battements cardiaques (ou pulsations) par unité de temps (la minute), le périmètre cervical et abdominal (en cm).

Dans une deuxième partie, les pneumologues collectent des données sur les facteurs de risque liée à la BPCO à savoir : Les données relatives au statut tabagique (fumeur actif ou passif, ancien fumeur, Année d'arrêt, Nombre de paquet-année et conseil minimal) ; la consommation de cannabis (Actif ou passif et la fréquence de consommation par semaine) ; la consommation de cigarette électronique (le mode de la consommation, la fréquence et le dosage) ; les données relatives à l'environnement (Amiante, tabagisme passif et l'exposition à des toxiques) et les Antécédents génétiques (emphysème et déficit A1AT).

Enfin, les pneumologues se concentrent sur le recueil des données cliniques en commençant par le motif de consultation, l'année de début des symptômes, la présence de la toux et l'expectoration, la dyspnée avec l'échelle mMRC ou l'échelle visuelle analogique (EVA), la présence de la douleur, le nombre d'exacerbations durant les 12 derniers mois, le score CAT (COPD Assessment

Test) pour mesurer la qualité de vie, le temps de marche quotidien (la minute), l'auscultation (normale, sibilants, crépitants, sous crépitants, ronchi et l'hypoventilation), le caractère (unilatéral et/ou bilatéral) et l'évaluation amyotrophique (Périmètre quadricipital 10 cm au-dessus de l'extrémité de la rotule).

### II.3.2. Les explorations fonctionnelles respiratoires (EFR)

L'Exploration Fonctionnelle Respiratoire (EFR) est un examen qui mesure les capacités respiratoires des poumons. Cet examen aide à faire le diagnostic de certaines maladies respiratoires, à en évaluer la sévérité, à en surveiller l'évolution et l'efficacité des traitements. La coopération du patient au cours de cet examen, simple et indolore, est capitale. Un médecin ou une infirmière guidera le patient pendant l'examen. En effet, le patient portera un pince-nez pour ne respirer que par la bouche. Soit il effectuera l'examen simplement assis (spirométrie), soit il sera dans une cabine transparente (pléthysmographie). On lui demandera de remplir ces poumons, retenir la respiration, puis souffler le plus vite et le plus fort possible à travers un embout buccal relié à un capteur. Cette manœuvre sera répétée plusieurs fois. Ce test mesure la quantité d'air inspiré par les poumons du patient et la vitesse à laquelle il va expulser cet air. On pourra lui demander de recommencer le test après avoir inhalé un médicament, afin d'apprécier l'éventuelle amélioration de la fonction respiratoire avec ce traitement. D'autres tests pourront être effectués : mesure des gaz du sang, test de transfert du monoxyde de carbone, test de broncho provocation, test d'effort...

L'exploration fonctionnelle respiratoire (EFR) réalisée au repos comprend l'évaluation de la capacité ventilatoire : spirométrie (débits expiratoires forcés et volumes mobilisables) et mesure des volumes pulmonaires statiques, notamment par pléthysmographie. La spirométrie permet de détecter l'existence d'un trouble ventilatoire obstructif tandis que la mesure des volumes statiques permet de détecter un trouble ventilatoire restrictif (diminution de la capacité pulmonaire totale) ou d'une distension thoracique (augmentation des volumes statiques). Il est important de savoir que cette évaluation est partielle et qu'il faut souvent y associer une évaluation de la demande ventilatoire, notamment lors du bilan d'une dyspnée : mesure de l'hématose de repos (permettant de définir l'éventuelle insuffisance respiratoire) et mesure du coefficient de transfert du CO, permettant avec la mesure du volume alvéolaire de calculer la DLCO (exploration de la barrière alvéolo-capillaire et du volume capillaire pulmonaire). L'ensemble de ces explorations fonctionnelles a fait l'objet d'une conférence de consensus européenne et américaine en 2005, dont la traduction française est parue en 2007. L'interprétation des EFR doit donc logiquement s'effectuer selon ces recommandations internationales et nécessite de se référer aux limites inférieures (5<sup>e</sup> percentile) et parfois supérieures (95<sup>e</sup> percentile) de valeurs théoriques ou prédites. La recherche d'une hyperréactivité bronchique sera seulement réalisée en cas de doute diagnostique d'un asthme (probabilité clinique pré-test intermédiaire, situation peu fréquente). L'exploration fonctionnelle à l'exercice la plus simple est le test de marche de 6 minutes qui donne une idée de la performance à la marche, permet de rechercher une désaturation artérielle et de quantifier l'éventuelle plainte dyspnéique.

La spirométrie peut faire suspecter un trouble ventilatoire restrictif, mais sa confirmation nécessite la mesure des autres volumes pulmonaires. Le volume expiratoire en 1 seconde (VEMS) est le principal index de la sévérité du TVO, il a une valeur pronostique importante. Il s'agit du paramètre de suivi le plus reproductible et sert d'un fil conducteur pour le choix du traitement. Nous évoquons une spirométrie pathologique si les valeurs des mesures spirométriques sont trop basses, relativement

à des valeurs prédites. Différentes manières existent pour effectuer cette comparaison. Si les valeurs de spirométrie étaient inférieures à 80% des valeurs prédites, elles sont considérées comme trop basses. Cependant, pour l'indice de Tiffeneau (VEMS/CVF), nous ne comparons pas le résultat aux valeurs prédites, mais plutôt à une valeur fixe de 70%. Un Tiffeneau inférieur à 70% était considéré comme une maladie pulmonaire obstructive.

### II.3.3. Les vingt-cinq comorbidités

Nous avons recueilli p= 25 comorbidités dans le cadre de notre étude. Elles peuvent être classées en 6 catégories :

**Tableau 5.** Catégorisation des comorbidités.

<b>Pathologies pulmonaires</b>	SAS, Asthme, DDB, Cancer de poumon.
<b>Cardiovasculaire</b>	Hypertension, Cardiopathie Ischémique, Trouble du rythme, Insuffisance cardiaque gauche, Athérome, AVC, HTAP, Artériopathie, Maladie thromboembolique.
<b>Métabolique</b>	Diabète, Dyslipidémie, Syndrome métabolique, Dénutrition, Problèmes digestifs.
<b>Ostéo-articulaire</b>	Ostéoporose, Ostéodensitométrie.
<b>Neuropsychiatrique</b>	Anxiété, Dépression.
<b>Autres comorbidités</b>	Manifestations sinusiennes, autres cancers.

Plusieurs études ont montré que les comorbidités sont sous-diagnostiquées, en réalité. Cela pourrait également être le cas dans notre étude pour la plupart des maladies à l'exception du SAS, car le SAS a été diagnostiquée par une polysomnographie réalisée par le même pneumologue qui a diagnostiqué la BPCO. Cependant, nous pensons que ce biais est limité dans notre étude, car nous avons trouvé une corrélation significative entre les comorbidités déclarées et le traitement qui leur est administré.

### II.3.4. Les médicaments

Les données sur les médicaments utilisés chez les patients BPCO relatives à la première visite d'inclusion ont été collectées par le biais des questionnaires de la plateforme PALOMB. Cette section est subdivisée en cinq sous-sections.

Premièrement, les pneumologues ont commencé par collecter les informations relatives aux Médicaments hors respiratoire à savoir : bêta-bloquant ; Statines ; Anti-coagulants ; Anti-agrégants plaquettaires ; Substitution calcium-vitamine D ; Anti-dépresseurs ; Anxiolytiques ; Anti-diabétiques ; Diurétiques ; Traitement substitutif de la ménopause ; inhibiteurs de l'enzyme de conversion et d'autres types de médicaments hors respiratoire.

Deuxièmement, les médicaments respiratoires de la BPCO : Les bronchodilatateur courte durée d'action (Salbutamol, Terbutaline, Ipratropium et les deux Associations : Salbutamol + ipratropium et Fénotérol + ipratropium) ; Les bronchodilatateur longue durée d'action (Formotérol, Salmétérol, Indacatérol, Olodaterol, Bambutérol, Tiotropium, Umeclidinium et Glycopyrronium) ; les Corticoïdes Inhalés (Budésonide, Béclométasone, Fluticasone et Furoate de mométasome) ; Corticoïdes par voie générale ; les Associations fixes (LABA + LAMA, LABA + CSI et LABA + LAMA + CSI) ; la Théophylline et les autres médicaments respiratoires (Singulair et zithromax).

Troisièmement, les différentes activités autour de l'assistance respiratoire : L'oxygénothérapie (Long terme, Nocturne, Déambulation et Courte durée) ; La ventilation non invasive ; La ventilation en pression positive continue et la ventilation sur trachéotomie.

Quatrièmement, les informations relatives aux interventions non médicamenteuses : La tentative antérieure de sevrage tabagique ; l'aide au sevrage tabagique ; la Kinésithérapie respiratoire (Ponctuelle, Au long cours et déambulation) ; l'éducation thérapeutique ; La réhabilitation respiratoire dans l'année écoulée ; Le programme de réentraînement à l'effort et les compléments alimentaires.

Cinquièmement, les données relatives à la prévention : La vaccination antigrippale annuelle ; La vaccination antipneumococcique ; Le suivi psychologique ; Le suivi psychiatrique ; Le plan de gestion des exacerbations ; L'activité physique et la prise en charge dans un parcours de soins établi (Ergothérapeute, Diététicien et Nutritionniste).

## **III. Exploitation des données**

### **III.1. Gestion des données**

L'ensemble des données (démographiques, cliniques, l'exploration fonctionnelle respiratoire et les comorbidités) récupérées à l'inclusion et tout au long du suivi grâce aux questionnaires ont été saisies par les pneumologues sur la plateforme ( [www.PALOMB.fr](http://www.PALOMB.fr) ).

Les différents fichiers de données ont ensuite été exportés au format XLSX afin de constituer une unique base de données, dans laquelle j'ai intégré les données concernant et les traitements. Chaque patient est désigné dans cette base par une ligne d'informations et ses deux numéros d'anonymat, l'un attribué au pneumologue ayant inclus le patient à la première visite d'inclusion, l'autre relatif au patient lui-même.

Avant de procéder à une analyse statistique, nous avons examiné le cas des valeurs manquantes et des individus aberrants. Nous avons effectué une extraction de ces individus afin de procéder à une analyse univariée, ainsi que bivariée vis-à-vis des classifications GOLD.

## **III.2. Analyse statistique**

L'analyse statistique des données de la cohorte PALOMB a été réalisée sur le logiciel SAS (versions 9.4 ; Statistical Analysis System) et le logiciel R (R version 4.0.0 ; Foundation for Statistical Computing, Vienne, Autriche).

### **III.2.1 Descriptif des patients BPCO de la cohorte PALOMB**

Dans un premier temps, la participation des patients dans la cohorte PALOMB au fil des ans a été décrite. Les principales caractéristiques des patients toujours suivis ont été comparées à celles des patients sortis du suivi par des tests du Chi-2. Dans un second temps, le descriptif de la santé respiratoire des patients a été réalisé. La prévalence des symptômes a été calculée à 4, 5 et 6 ans de suivi. La prévalence des comorbidités, l'âge auquel les diagnostics ont été posés ainsi que la proportion des patients BPCO traités ont été décrits, tout comme les données relatives au décès à 5 ans. Les prévalences de la sévérité de la BPCO en fonction des différents symptômes et comorbidités ont par ailleurs été comparés par des tests de Chi-2.

Les distributions des volumes et débits respiratoires mesurés lors de l'EFR (VEMS, CPT et CVF exprimés en valeur brute ou en pourcentage de leur valeur théorique) et de la VR, FRC et CI ont également été établie. L'association des paramètres de la fonction respiratoire avec les différents symptômes et comorbidités a été étudiée par le biais de régressions linéaires ajustées sur l'âge, la taille, le sexe du patient et les classes thérapeutiques de ses patients. L'association entre la sévérité du VEMS et ces symptômes/comorbidités a quant à elle été étudiée par des tests de Kruskal-Wallis.

### **III.2.2 Identification de Profils de comorbidités - Analyse en cluster**

Une méthode statistique de classification non-supervisée a été utilisée afin d'identifier, de manière objective, des profils de comorbidités basés sur les trajectoires de plusieurs patients BPCO dans la cohorte PALOMB.

L'analyse de la variance a été utilisée pour les variables continues et le test de Chi-2 a été utilisé pour les variables catégorielles. Nous avons effectué une analyse univariée entre les comorbidités et la sévérité de la BPCO (GOLD 2011). Afin d'effectuer une analyse plus approfondie, nous avons décidé de retenir que les comorbidités significativement plus fréquentes aux stades plus sévères de la BPCO (B, C, D) qu'au stade léger (A). Ensuite, nous avons effectué une analyse en composantes principales et une analyse de cluster non hiérarchique (K-means) pour identifier les clusters stables de comorbidités. Pour mieux définir le nombre optimal de cluster, nous avons utilisé trois méthodes statistiques : la méthode des éboulis, le pourcentage de variance expliqué et la méthode de Kaiser-Guttman.

Enfin, nous avons effectué une analyse en cluster (méthode de Ward) pour assurer la stabilité des différents clusters. Toutes les analyses ont été effectuées avec la version 9.4 du logiciel SAS. Tous les tests statistiques étaient bilatéraux, les résultats avec  $p < 0,05$  sont considérés comme étant statistiquement significatifs.

Cette classification a été basée sur la présence à l'inclusion, des comorbidités cardiovasculaire, métabolique, ostéo-articulaire, neuropsychiatrique et/ou pathologies pulmonaires suivants : Le syndrome d'apnée du sommeil, l'asthme, la dilatation des bronches, le cancer de poumon, l'hypertension, la cardiopathie ischémique, le trouble du rythme, l'insuffisance cardiaque gauche, l'athérome, l'accident vasculaire cérébral, l'hypertension artérielle pulmonaire, l'artériopathie, la maladie thromboembolique, le diabète, la dyslipidémie, le syndrome métabolique, la dénutrition, les problèmes digestifs, l'ostéoporose, l'ostéodensitométrie, l'anxiété, la dépression, les manifestations sinusiennes et les autres types des cancers.

L'analyse a été réalisée chez les 1584 patients de la cohorte PALOMB qui avaient des données relatives à la symptomatologie respiratoire, l'EFR et aux 25 comorbidités disponibles. Les comorbidités ( $n = 19$ ) ont été enregistrées systématiquement de manière standardisée par le pneumologue. Le diagnostic de comorbidité a d'abord été évalué par le rapport du patient, puis confirmé soit en examinant la liste des médicaments du patient, soit lorsque des tests complémentaires étaient disponibles dans les dossiers médicaux. La bronchectasie a été enregistrée selon des critères cliniques et / ou radiologiques, comme d'habitude en pratique clinique. Les données manquantes ont été imputées par interpolation linéaire. La classification a été répétée en faisant varier le nombre désiré de profils obtenus de deux à dix. Le choix du nombre optimal de profils et donc de la « meilleure » classification a été déterminé au regard du critère d'information bayésien BIC [98] et de la pertinence clinique des profils identifiés.

De plus, une analyse de sensibilité a été réalisée en appliquant la même méthode de classification aux patients ayant des données complètes.

### **III.2.3 Identification des déterminants de la dyspnée par une méthode de régression pénalisée**

La dyspnée est un des symptômes cardinaux de la BPCO, et fait partie des éléments d'évaluation pour apprécier la sévérité de la BPCO et décider de la prise en charge thérapeutique.

L'évaluation de la dyspnée en pratique clinique n'est pas chose aisée, et il est parfois difficile d'apprécier la part attribuable respiratoire de la dyspnée chez des patients ayant des comorbidités.

Une analyse descriptive a été réalisée puis une analyse multivariée afin de déterminer les facteurs associés à la dyspnée définie par mMRC $\geq$ 2. Les variables catégorielles sont présentées sous forme de proportions et les variables continues sont présentées sous forme de moyenne et écart-type. Nous avons effectué une analyse de régression logistique univariée entre l'échelle de dyspnée mMRC et le VEMS, le nombre d'exacerbations, le sexe, la distension thoracique et les comorbidités. Les différences statistiquement significatives entre les groupes ont été testées par le test de Wilcoxon, le test du Chi-2 ou les tests exacts de Fisher, selon le cas. Nous avons effectué une analyse croisée approfondie de différents facteurs, pour décrire les similitudes / différences existant entre les groupes de dyspnée. Les relations entre l'échelle de la dyspnée mMRC et les variables cliniques ont été examinées par les coefficients de corrélation de Pearson et de Spearman. Un modèle multivarié de régression logistique avec une procédure de type (Sélection BACKWARD basée sur le Test de Wald) a été utilisé pour identifier les facteurs associés au groupe de dyspnée plus élevé. Ce modèle comprenait les comorbidités étudiées, l'âge, le sexe, la distension thoracique, la toux, les expectorations, le nombre d'exacerbations et d'autres paramètres. Les variables indépendantes incluses dans le modèle étaient également celles montrant un P <0,05 en analyse bivariée.

Nous avons également effectué une analyse de sensibilité pour une dyspnée plus élevée avec différentes valeurs seuils : échelle mMRC  $\geq$  1 ; échelle mMRC  $\geq$  2 ; Échelle mMRC  $\geq$  3 et échelle mMRC  $\geq$  4.

L'estimateur des moindres carrés pénalisé par la norme L<sub>1</sub> du coefficient, connu sous le nom de LASSO [99] (The least absolute shrinkage and selection operator : une méthode d'analyse de régression par apprentissage automatique qui effectue à la fois la sélection et la régularisation des variables afin d'améliorer la précision de prédiction et l'interprétabilité du modèle statistique) a été utilisé pour sélectionner les variables associées à l'échelle de dyspnée mMRC dans un sous-groupe de patients qui n'avaient pas de valeurs RV, IC et FRC manquantes.

### **III.2.4 Identification de profils d'exacerbations - classification non-supervisée**

Une méthode statistique de classification non-supervisée a été utilisée afin d'identifier, de manière objective, des profils basés sur les trajectoires des patients en terme de la fréquence d'exacerbations entre l'inclusion et la troisième année de suivi dans la cohorte PALOMB. Cette méthode, Kml, est un algorithme de partitionnement, adapté de celui des k-means, qui permet de synthétiser l'étude conjointe de trajectoires d'un ou plusieurs variables sous la forme d'une seule variable nominale, correspondant aux différents groupes (ou profils) ainsi observés dans l'échantillon. Les patients sont répartis dans ces différents groupes afin de minimiser leurs distances euclidiennes aux trajectoires moyennes de leurs groupes. Cette classification a été basée sur la survenue, à 1, 2 et 3 ans, des exacerbations.

L'analyse a été réalisée chez les 932 patients de la cohorte PALOMB qui avaient des données relatives à la fréquence des exacerbations disponibles pour plus de la totalité des trois dates de points. La classification a été répétée en faisant varier le nombre désiré de profils obtenus de deux à six. Le

choix du nombre optimal de cluster et donc de la « meilleure » classification a été déterminé au regard du critère d'information bayésien BIC et de la pertinence clinique des profils identifiés.

L'étude a commencé le 1er janvier 2014 (date du début d'inclusion de la cohorte PALOMB) et s'est terminée le 1er janvier 2018 (décès, abandon de la pratique ou fin de suivi). La première analyse, réalisée à l'inclusion, nous a permis de décrire et de caractériser les patients ayant des exacerbations fréquentes ( $\geq 2$ ). La seconde, appelée analyse approfondie, était une analyse de confirmation utilisant des informations complètes sur les exacerbations annuelles, au cours de chacune des 3 années de suivi. Nous avons utilisé l'analyse KML, qui est une implémentation d'un algorithme non paramétrique conçu pour travailler spécifiquement sur des données longitudinales [100]. L'algorithme a identifié des groupes de patients avec une évolution similaire de la fréquence des exacerbations sur les 3 ans. Cinq différents indices ont été utilisés pour choisir le nombre optimal de clusters : l'indice de Calinski et Harabatz (version 1, 2 et 3), l'indice de Ray et turi et l'indice de Davies et Bouldin [100]. L'analyse ROC (Receiver Operating Characteristic) a été réalisé pour identifier le meilleur seuil pour séparer les clusters.

Les différences statistiquement significatives entre les patients et les différences entre les groupes ont été examinées en utilisant le test t de Student, le test du chi-2 ( $\chi^2$ ) ou le test exact de Fisher, selon le cas. Les variables indépendantes incluses dans le modèle étaient celles montrant un  $P < 0,05$  en analyse bivariée. Un modèle multivarié de régression logistique avec une procédure de type (Sélection BACKWARD basée sur le Test de Wald) a été utilisé pour identifier les facteurs associés aux groupes d'exacerbations fréquentes. Le modèle incluait les comorbidités étudiées et a été ajusté pour les traitements et d'autres paramètres. Toutes les analyses ont été effectuées avec le logiciel SAS version 9.4 et R version 3.3.2. Les données ont été analysées à l'aide des packages KML, version 2.4.1, pROC, version 1.8 et dplyr version 0.5.0 [100, 101].

### **III.2.5 L'analyse en cluster la plus optimale (supervisé vs non-supervisé)**

L'identification de sous-groupes de patients atteints de BPCO et partageant des caractéristiques cliniques et/ou paracliniques communes est à l'origine du concept de phénotypes dans cette pathologie. De nombreuses études de cohortes observationnelles sont en cours, pour tenter d'identifier de nouveaux phénotypes de patients BPCO. L'objectif de cette étude est de déterminer l'analyse en cluster la plus optimale (supervisé vs non-supervisé) pour identifier de façon robuste les phénotypes cliniques dans le cadre de la cohorte PALOMB.

Les patients BPCO (VEMS / CVF  $< 70\%$  post BD) ont été inclus par leur pneumologue depuis janvier 2014 jusqu'au février 2020. Les pneumologues ont rempli un questionnaire en ligne comportant les données cliniques, fonctionnelles, les comorbidités et les traitements inhalés. Après cinq ans de suivi, les données sur leur statut vital ont été obtenues en utilisant le RNIPP (Répertoire National d'identification des Personnes Physiques) [102]. En premier lieu, une analyse en composantes principales, suivie d'une classification hiérarchique sur les composantes principales a été réalisée, suivie de deux algorithmes de classification non-supervisée (k-means et PAM). Nous avons procédé à plusieurs classifications, retenu celle qui semble la "robuste" en fonction des paramètres suivants :  $R^2$ , homogénéité des clusters, l'indice de connectivité, l'indice de Dunn et l'indice de silhouette.

Afin d'évaluer l'impact des comorbidités sur la sévérité des patients BPCO, nous avons construit un score non pénalisé, qui affecte, pour chaque patient, le nombre des comorbidités dont il est atteint simultanément. Il est ainsi construit en sommant les colonnes des 25 variables de comorbidités de notre base de données.

Le critère d'évaluation principal était la mortalité toutes causes confondues et des analyses de régression de Cox à plusieurs variables ont été utilisées pour évaluer les associations entre les comorbidités de la BPCO et les paramètres de l'étude. Des arbres de classification et de régression (CART : classification and regression trees) ont été utilisés pour répartir les patients entre les clusters et la pertinence clinique a été déterminée en comparant la mortalité à 5 ans.

Les deux critères de jugement principaux étaient le taux de mortalité et le délai avant le décès. Le délai de mort allait jusqu'à 5 ans et 4 mois (du 1er janvier 2014 au 30 avril 2019) avec ces données. La période de suivi pour le calcul de la survie a été prise de la date de l'inclusion, à la date du décès, ou le 30 avril 2019, selon la première éventualité. Les taux de mortalité ont été suivis pendant 64 mois (5 ans et 4 mois : de janvier 2014 à avril 2019). Les patients non répertoriés dans le RNIPP étaient supposés vivants à la fin de la période d'étude.

### **III.2.6 L'analyse de survie – Principaux déterminants associés à la mortalité**

Les causes de mortalité chez les patients BPCO sont de reconnaissance récente et incomplète. Le risque de mortalité chez les patients BPCO est influencé par l'âge, la sévérité de l'obstruction bronchique et les comorbidités. Cependant, l'impact des comorbidités semble varier entre les patients BPCO. Nous avons cherché à étudier les associations entre les comorbidités associés aux patients BPCO et les pronostics évalués par la mortalité toutes causes dans une cohorte réelle de patients BPCO bien caractérisés.

Les variables catégorielles sont présentées sous forme de proportions. Les variables continues sont présentées sous forme de moyenne et écart-type si elles sont normalement distribuées, ou médiane et interquartile si elles ne sont pas distribuées normalement. La comparaison des moyennes entre deux groupes d'étude (patients décédés vs patients vivants) était basée sur l'analyse de la variance (et les tests de Wilcoxon ou Kruskal – Wallis, selon le cas), avec le test du chi-2 (ou les tests exacts de Fisher, selon le cas) et Bonferroni correction pour la comparaison des proportions. La mortalité toutes causes confondues a été évaluée et la survie à cinq ans entre les groupes a été comparée à l'aide de l'analyse de Kaplan-Meier.

Nous avons d'abord effectué des analyses univariées basées sur le modèle de risques proportionnels de Cox en utilisant chacun des prédicteurs potentiels de la mortalité comme variables

indépendantes et le statut de survie comme variable dépendante. Les courbes de survie pour les groupes (patients décédés vs vivants) ont été estimées par la méthode de Kaplan-Meier (l'estimateur produit-limite) et comparées avec le test du log-rank.

Un modèle multivarié de régression logistique avec une procédure de type (Sélection BACKWARD basée sur le Test de Wald) a été utilisé pour identifier les facteurs associés à la mortalité chez les patients atteints de BPCO. Les prédicteurs significatifs de l'analyse univariée ont été introduits par étapes dans un modèle de risques proportionnels de Cox pour tester l'effet indépendant de chaque facteur de risque candidat. Les covariables ont été incluses dans le modèle final si elles contribuaient de manière significative à l'ajustement du modèle et étaient cliniquement significatives. Les résultats de la modélisation des risques proportionnels de Cox sont présentés sous forme de rapports de risque (HR) et d'intervalles de confiance (IC) à 95%. Le risque de décès a été calculé sur le temps-patient à risque. Le temps écoulé jusqu'au décès a été censuré à la fin du suivi (30 avril 2019). Les rapports de risque ont été ajustés en fonction de l'âge et de l'indice BOD [103], qui étaient liés au décès. L'ajustement sur la période d'entrée dans l'étude n'a pas changé les résultats.

Les analyses des phénotypes étaient basées sur l'arbre de classification et de régression (CART) développé à partir des cohortes françaises / belges de BPCO regroupées [104]. En bref, un algorithme basé sur des variables cliniques (y compris les comorbidités cardiovasculaires, le diabète et les caractéristiques respiratoires) permet de répartir les patients en cinq clusters correspondant à des phénotypes identifiés par une analyse en clusters avec différents taux de mortalité toutes causes confondues et de décès. Toutes les analyses ont été effectuées avec le logiciel SAS version 9.4 et R version 3.3.2.

## RÉSULTATS

Ce chapitre présente les résultats de ce travail de thèse. Dans un premier temps, un descriptif détaillé des patients de la cohorte PALOMB sera réalisé (partie I), puis les profils de comorbidités identifiés par l'analyse en clusters seront décrits (partie II). Ensuite, les phénotypes et les déterminants d'exacerbations identifiés par une classification non-supervisée seront présentés (partie III). Les déterminants de la dyspnée par la régression pénalisée et l'analyse de sensibilité seront décrits (partie IV). Enfin, l'analyse de mortalité et les phénotypes cliniques des patients BPCO identifiés par l'analyse en cluster la plus optimale (supervisé vs non-supervisé), sera présentée dans les parties IV et V.

L'inclusion des patients BPCO de la cohorte PALOMB a commencé en janvier 2014. On a différentes tailles d'échantillons sur lesquels ont porté ces différentes analyses. Les effectifs dépendaient des données disponibles au moment de l'exploitation statistique de celles-ci.

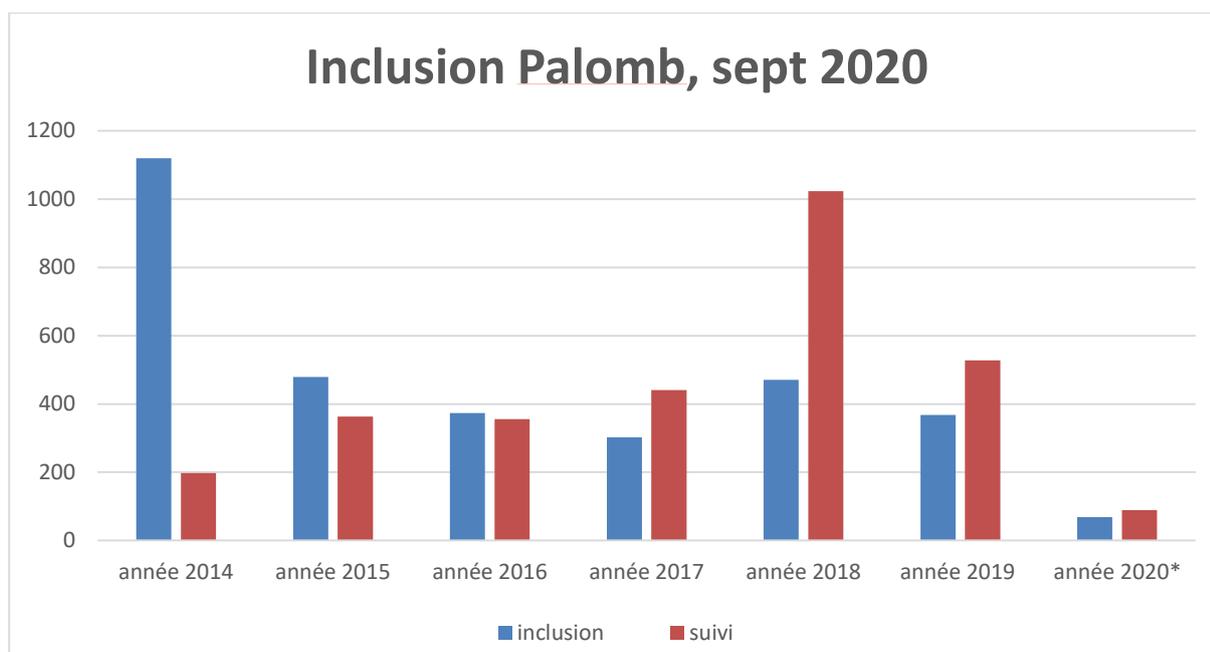
## I. Descriptif des patients de la cohorte PALOMB (n=1391)

Le but de ce chapitre est d'offrir un descriptif complet de la population d'étude de la cohorte PALOMB. Le profil des patients, les principales caractéristiques cliniquement pertinentes, l'état de leur santé respiratoire, les prévalences des différentes comorbidités et les caractéristiques de leur prise en charge thérapeutique seront successivement présentés en détail.

### I.1. Population d'étude

La Figure 13 représente l'évolution de la participation dans la cohorte PALOMB depuis l'inclusion des patients en janvier 2014 jusqu'au septembre 2020. Durant l'année 2014, 1120 inclusions ont pu être réalisés, contre 479 inclusions en 2015, 373 inclusions en 2016, 302 inclusions en 2017, 471 inclusions en 2018, 368 inclusions en 2019 et 69 inclusions jusqu'au septembre 2020. C'est-à-dire que les pneumologues de la cohorte PALOMB ont pu inclure 3182 patients BPCO entre janvier 2014 et septembre 2020. Le nombre de visite de suivi est de 2998 jusqu'au septembre 2020. Ce nombre varie selon les patients entre 0 et 10 visites de suivi par patient.

La cohorte a débuté en janvier 2014 et un premier échantillon de 1391 patients a été analysé. Les caractéristiques démographiques et cliniques des patients sont présentées dans le **tableau 6**. L'échantillon est composé majoritairement d'hommes (910 hommes soit 65.4% de la population totale de l'étude). Pour 81% des patients (1125 patients), les premiers symptômes ont débuté entre 50 et 79 ans. La plupart des patients de notre étude sont des fumeurs (1323 patients, 95.1%) dont 64 % sont ex-fumeurs et 36% fumeurs actifs.



**Figure 13.** Diagramme d'inclusion des patients BPCO de la cohorte PALOMB participant aux différentes parties de l'exploitation statistique réalisée dans le cadre de ce travail de thèse.

**Tableau 6.** Caractéristiques cliniques et démographiques de la population totale.

Variables		n=1391 (%)	
<b>Sexe</b>	Homme	910	(65,4)
	femme	481	(34,6)
<b>Tabagisme</b>	Non –fumeur	68	(4,9)
	Fumeur actif	487	(35,8)
	Ex fumeur	873	(64,2)
<b>VEMS en % de la valeur prédite</b>	≥ 80 %	179	(12,9)
	50 - 80 %	750	(53,9)
	30 - 50 %	369	(26,5)
	≤ 30 %	93	(6,7)
<b>Stade de la dyspnée à l'état basal selon MMRC</b>	0	137	(10)
	1	381	(27,9)
	2	428	(31,3)
	3	342	(25)
	4	80	(5,9)
<b>Toux</b>		794	(57,1)
	<b>Expectoration</b>	569	(40,9)
<b>Nombre d'exacerbation par an</b>	0	515	(37)
	1	456	(32,8)
	2	216	(15,5)
	3	138	(9,9)
	≥4	66	(4,7)
<b>IMC</b>	≤ 20	223	(16)
	20 –25	480	(34,5)
	25 –30	280	(20,1)
	≥ 30	408	(29,3)

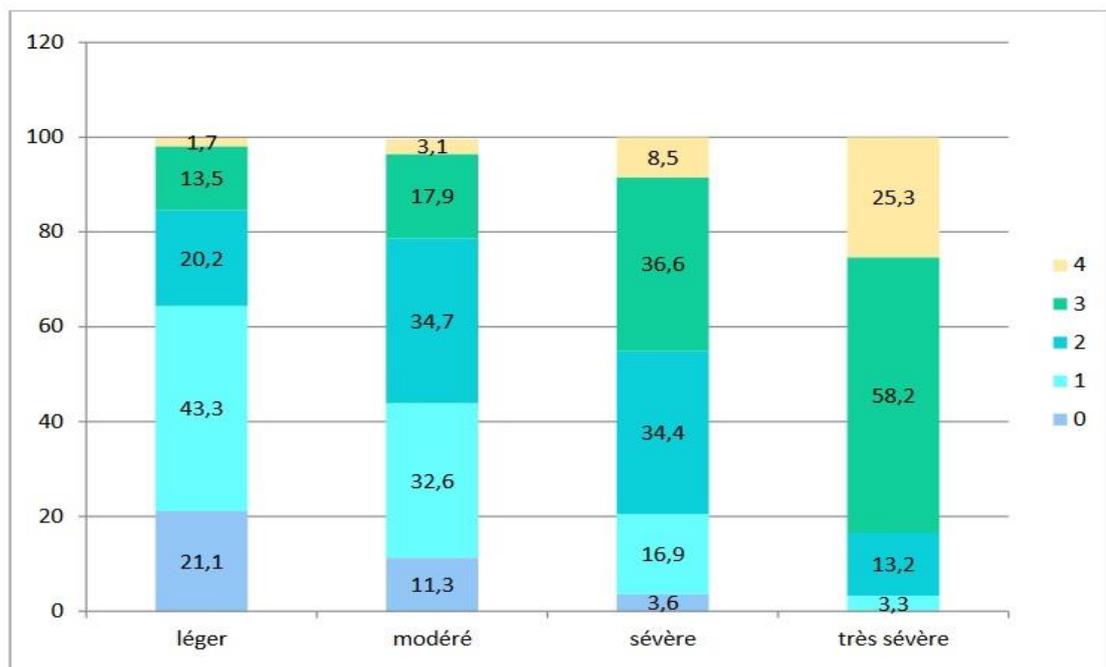
De façon globale, plus de la moitié des patients fumeurs ont une consommation tabagique > 30 PA. La consommation tabagique est d'autant plus importante que les patients sont symptomatiques (Tableau 7).

**Tableau 7.** Caractéristiques du tabagisme en fonction du stade de sévérité.

Variabes	Léger n=179(%)	Modéré n=750(%)	Sévère n=369(%)	Très sévère n=93(%)	Total n=1391(%)	p
Tabagisme	171 (95,5)	717 (95,6)	347 (94)	88 (94,6)	1323 (95)	0,7021
Tabac actif	81 (45,3)	266 (35,5)	113 (30,6)	27 (29)	487 (35,8)	0,0055
<b>Nombre PA</b>						
<20 PA	16 (8,9)	50 (6,7)	11 (3)	1 (1,1)	78 (5,6)	0,0018
20-29 PA	29 (16,2)	81 (10,8)	31 (8,4)	6 (6,5)	147 (10,6)	
30-39 PA	58 (32,4)	207 (27,6)	112 (30,4)	22 (23,7)	399 (28,7)	
>40 PA	66 (36,9)	356 (47,5)	180 (48,3)	57 (61,3)	659 (47,4)	
Cannabis	6 (3,4)	17 (2,3)	4 (1,1)	3 (3,2)	30 (2,2)	0,2873

## I.2. Les principales caractéristiques cliniquement pertinentes

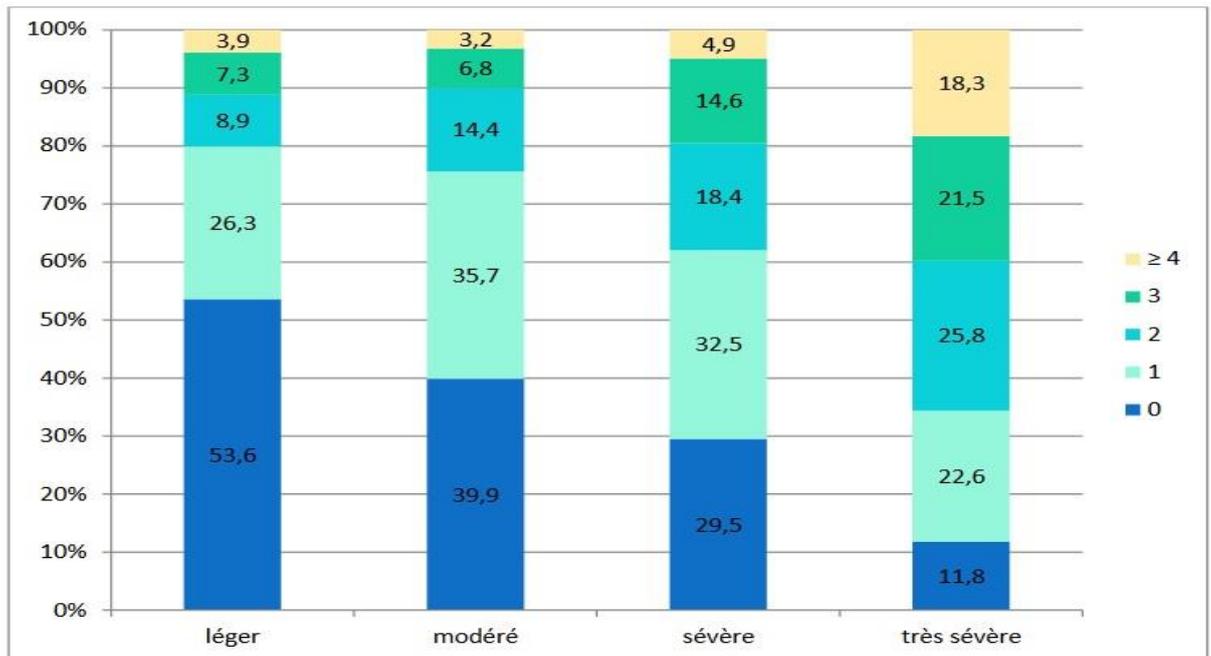
Sur le plan clinique, plus de la moitié des patients ont une BPCO modérée (n=750, soit 54%). 13% (n=179) une BPCO légère, 26% (n=369) une BPCO sévère et 7% (n=93) une BPCO très sévère (Figure 10). Les patients BPCO suivis en consultation par les pneumologues sont symptomatiques quel que soit le degré de sévérité de la maladie avec toutefois des symptômes plus fréquents et plus sévères pour les stades avancés. La dyspnée est quasiment constante dans notre population puisque 90% des patients ont une dyspnée de stade  $\geq 1$  selon l'échelle mMRC. Celle-ci est significativement plus sévère pour les stades sévères et très sévères ( $p < 0.00.1$ ). 31,3% des patients présentent une dyspnée de stade 2 selon l'échelle mMRC (Tableau 6 et figure 14)



**Figure 14.** Dyspnée en fonction du stade de sévérité de la BPCO.

57 % des patients présentent une toux et 41% des expectorations chroniques (**Tableau 6**). Ces symptômes sont présents dès le stade léger, mais leur fréquence est significativement plus élevée pour les stades sévères et très sévères ( $p < 0.001$  pour les expectorations chroniques et  $p = 0.001$  pour la toux).

Plus d'un tiers des patients ne présentent pas d'exacerbation dans l'année ( $n = 515$ , soit 37%). 20% des patients de stade léger présentent au moins 2 exacerbations par an. Ce pourcentage augmente avec le stade de sévérité : 24% dans le stade modéré, 38% dans le stade sévère et 66% dans le stade très sévère ( $p < 0.001$ ) (**figure 15**).



**Figure 15.** Nombre d'exacerbations par an selon le stade de sévérité.

### I.3. Les prévalences des différentes comorbidités

38% des patients présentent une HTA, 25% une dyslipidémie, 21% une cardiopathie ischémique et 13% un diabète. 19% des patients présentent un syndrome d'apnées du sommeil (dont 97 % de type obstructif). Enfin, l'anxiété est présente chez 19% des patients et la dépression 14% des patients. L'HTA, le SAS, la dépression et l'anxiété sont significativement plus présents dans les stades modérés et très sévères. L'asthme est associé à la BPCO dans 12% des cas et sa répartition selon la sévérité de la maladie ne diffère pas selon le stade de sévérité de la maladie ( $p = 0.845$ ). Le cancer est associé à la BPCO dès le stade léger et 17% des patients de la population totale ont un cancer (**Tableau 8**).

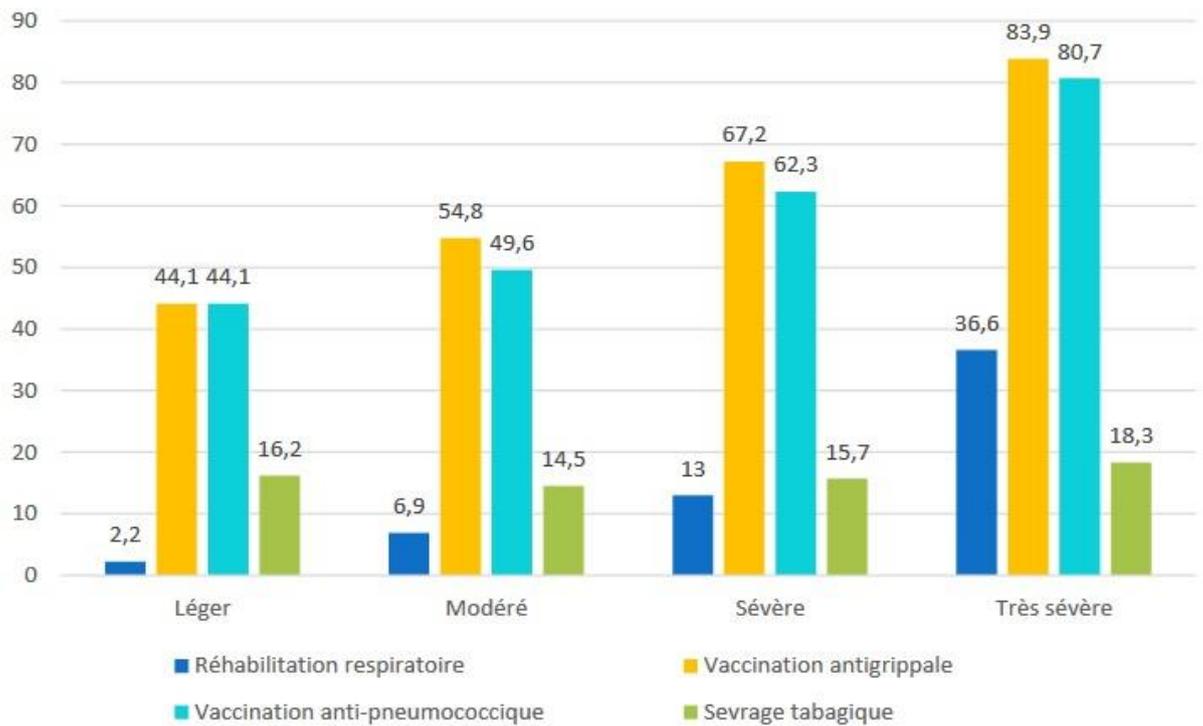
**Tableau 8.** Répartition des comorbidités selon le stade de sévérité.

Variétés	Léger n=179(%)	Modéré n=750(%)	Sévère n=369(%)	Très sévère n=93(%)	Total n=1391(%)	p
HTA	46 (25,7)	307 (40,9)	140 (37,9)	39 (41,9)	532 (38,3)	0.002
Dyslipidémie	44 (24,6)	203 (27,1)	84 (22,8)	19 (20,4)	350 (25,2)	0.295
Cardiopathie ischémique	26 (14,5)	165 (22)	76 (20,6)	18 (19,4)	285 (20,5)	0.169
Anxiété	28 (15,6)	130 (17,3)	71 (19,2)	40 (43)	269 (19,3)	<0.001
SAS	48 (26,8)	163 (21,7)	41 (11,1)	5 (5,4)	257 (18,5)	<0.001
Dépression	23 (12,9)	100 (13,3)	50 (13,6)	26 (28)	199 (14,3)	0.002
Cancer	31 (17,3)	125 (16,7)	57 (15,5)	13 (14)	226 (16,3)	0.856
Trouble du rythme	13 (7,3)	122 (16,3)	46 (12,5)	13 (14)	194 (14)	0.014
Diabète	16 (8,9)	107 (14,3)	47 (12,7)	7 (7,5)	177 (12,7)	0.103
Asthme	21 (11,7)	94 (12,5)	42 (11,4)	9 (9,7)	166 (11,9)	0.845
Athérome	9 (5)	55 (7,3)	29 (7,9)	6 (10,8)	99 (5,8)	0.658
Insuffisance cardiaque	3 (1,7)	31 (5,1)	30 (8,1)	10 (0,7)	81 (5,8)	0.003
DDB	7(3,9)	31 (4,1)	34 (9,2)	6 (6,5)	99 (5,6)	0.004
Syndrome métabolique	5 (2,8)	28 (3,7)	17 (4,6)	5 (5,4)	55 (4)	0.650
AVC	9 (5)	28 (3,7)	10 (2,7)	4 (4,3)	51 (3,7)	0.570
Ostéoporose	7 (3,9)	21 (2,8)	20 (5,4)	10 (10,8)	58 (4,2)	0.012
HTAP	1 (0,6)	11 (1,5)	11 (3)	12 (12,9)	35 (2,5)	<0.001
Dénutrition	1 (0,6)	7 (0,9)	15 (4,1)	9 (9,7)	32(2,3)	<0.001

## I.4. Prise en charge des patients BPCO

Le conseil minimal est prodigué chez 36% des patients tous stades de sévérité confondus, sans différence significative entre les différents stades ( $p=0,852$ ). Le sevrage tabagique est obtenu chez 15% des patients sans différence significative entre les stades ( $p=0,768$ ) (**Figure 16**). Une aide au sevrage tabagique est prescrite chez 9% des patients de la population totale. Parmi les patients ayant reçu le conseil minimal, 22% ont eu une aide au sevrage tabagique, alors que pour les patients n'ayant pas reçu le conseil minimal, une aide au sevrage tabagique est apportée chez 3% des patients (**Tableau 9**). Les substituts nicotiques sont prescrits chez 6,3% des patients, les autres moyens d'aide au sevrage tabagique sont anecdotiques (champix : 0,46% de la population soit 6 patients, zyban : 0,15% soit 2 patients, thérapie cognitivo-comportementale : 0,53% soit 7 patients, acupuncture 0,15% soit 2 patients, hypnose 0,53% soit 7 patients, magnétiseur : 0,08% soit 1 patient).

59% des patients de la population totale sont vaccinés contre la grippe et 54% contre le pneumocoque. La couverture vaccinale augmente avec la sévérité de la maladie ( $p<0.001$ ) (**Figure 16**). 10% des patients de la population totale sont adressés en réhabilitation respiratoire. Celle-ci est principalement réalisée en centre (dans 66% des cas). Le réentraînement à l'effort est réalisé chez 9% des patients. La réhabilitation respiratoire est davantage prescrite pour les stades très sévères avec une différence significative ( $p<0.001$ ) (**Figure 16**).



**Figure 16.** Traitement non médicamenteux en fonction du stade de sévérité.

**Tableau 9.** Aide au sevrage tabagique selon le conseil minimal.

Variables	<u>Conseil Minimal</u>		<u>Absence de Conseil Minimal</u>		p
	Aide au Sevrage tabagique	Absence Aide au sevrage tabagique	Aide au sevrage tabagique	Absence Aide au sevrage tabagique	
<b>Léger (n=179)</b>	14/66 (21,2)	52/66 (78,8)	8/105 (7,6)	97/105 (92,4)	0,01
<b>Modéré (n=750)</b>	52/252 (20,6)	200/252 (79,4)	9/462 (2)	453/462 (98,1)	<0,001
<b>Sévère (n=369)</b>	29/127 (22,8)	98/127 (77,2)	4/217 (1,8)	213/217 (98,2)	<0,001
<b>Très sévère (n=93)</b>	8/31 (25,8)	23/31 (74,2)	2/57 (3,5)	55/57 (96,5)	0,002
<b>Total (n=1391)</b>	103/476 (21,3)	373/476 (78,4)	23/241 (2,7)	818/841 (97,3)	<0,001

75% des patients de la population totale ont une prescription de traitement au moment de la consultation avec le pneumologue. Au décours de la consultation avec le pneumologue, cette proportion de patients traités est évaluée à 79.3%. Les patients sont d'autant plus traités quand ils sont à un stade avancé de leur BPCO ( $p < 0.001$ ).

Parmi les patients qui recevaient un traitement avant la consultation avec le pneumologue, 7% d'entre eux ont eu un arrêt de traitement. Inversement, 39% des patients ont eu une instauration de traitement alors qu'ils n'en recevaient pas avant la consultation. Dans les stades légers, les traitements sont arrêtés chez 11% des patients et un traitement est introduit chez 35% des patients. Dans les stades très sévères, les traitements sont arrêtés chez 1% des patients et 40% des patients ont eu une instauration de traitement médicamenteux.

Les bronchodilatateurs à courte durée d'action sont prescrits chez 57% des patients. Tous stades de sévérité confondus, le salbutamol est le plus souvent prescrit (38% des cas), la terbutaline et le combivent sont prescrits dans moins de 5% des cas. L'ipratropium est quant à lui prescrit dans 13% des cas. (**Tableau 10**). Les bronchodilatateurs à longue durée d'action sont prescrits chez 44% des patients, avec une prescription plus fréquente pour l'indacatérol (28%). Le tiotropium représente 9% des bronchodilatateurs à longue durée d'action et le salmétérol 2%.

**Tableau 10.** Traitement BD (CDA et LDA) selon le stade de sévérité.

Variables	Léger n=179	Modéré n=75	Sévère n=369	Très Sévère n=93	Total n=1391	p
<b>BD CDA</b>	94 (53)	374 (50,1)	233 (63,3)	86 (92,4)	787 (56,6)	
<b>Salbutamol</b>	69 (39)	270 (36)	143 (38,9)	48 (51,6)	530 (38,1)	0,004
<b>lpratropium</b>	21 (11,7)	76 (10,3)	59 (16)	19 (20,4)	175 (12,6)	0,004
<b>terbutaline</b>	3 (1,7)	23 (3,1)	22 (6)	15 (16,1)	63 (4,5)	<0,001
<b>combivent</b>	1 (0,6)	5 (0,7)	9 (2,4)	4 (4,3)	19 (1,4)	0,005
<b>BD LDA</b>	58 (32,4)	343 (45,7)	172 (46,6)	40 (43)	613 (44,1)	
<b>Indacatérol</b>	35 (19,6)	222 (29,6)	106 (28,7)	19 (20,4)	382 (27,5)	0,019
<b>Tiotropium</b>	13 (7,3)	65 (8,7)	37 (10)	8 (8,6)	123 (8,8)	0,745
<b>Formoterol</b>	7 (3,9)	35 (4,7)	21 (5,7)	8 (8,6)	71 (5,1)	0,329
<b>Salmeterol</b>	2 (1,1)	13 (1,7)	5 (1,4)	2 (2,2)	22 (1,6)	0,879
<b>Glycopyrronium</b>	1 (0,6)	8 (1,1)	3 (0,8)	3 (3,2)	15 (1,1)	0,193

Les CSI seuls sont prescrits chez 6% de la population et les associations fixes chez 29% des patients (l'association fluticasone-salmétérol représentant plus de la moitié des prescriptions de traitement combiné). 1,4% des patients reçoivent des corticoïdes par voie générale (**Tableau 11**).

**Tableau 11.** Traitement corticoïdes selon le stade de sévérité.

<b>Variables</b>	<b>Léger n=179 (%)</b>	<b>Modéré n=750 (%)</b>	<b>Sévère n=369 (%)</b>	<b>Très Sévère n=93 (%)</b>	<b>Total n=1391 (%)</b>	<b>P</b>
<b>Corticoïdes inhalés</b>	2 (1,1)	36 (4,8)	36 (9,8)	10 (10,8)	84 (6,1)	
<b>Budésonide</b>	2 (1,1)	22 (2,9)	16 (4,3)	4 (4,3)	44 (3,2)	0,201
<b>Béclométasone</b>	0	4 (0,5)	5 (1,4)	3 (3,2)	12 (0,9)	0,022
<b>Fluticasone</b>	0	10 (1,3)	15 (4,1)	3 (3,2)	28 (2)	0,003
<b>Associations Fixes</b>	30 (16,8)	183 (24,4)	139 (37,7)	46 (49,5)	398 (28,6)	
<b>Fluticasone-Salmétérol</b>	17 (9,5)	109 (14,5)	74 (20,1)	25 (26,9)	225 (16,2)	0,002
<b>Budésonide-Formotérol</b>	11 (6,2)	58 (7,7)	49 (13,3)	17 (18,3)	135 (9,7)	0,003
<b>Béclométasone-Formotérol</b>	2 (1,1)	16 (2,1)	16 (4,3)	4 (4,3)	38 (2,7)	0,008
<b>Corticoïdes par voie générale</b>	1 (0,6)	7 (0,9)	6 (1,6)	5 (5,4)	19 (1,4)	0,004

## II. Les profils de comorbidités identifiés par l'analyse en clusters (n=1584)

Ce travail a fait l'objet d'une publication sous forme d'article original dans la revue scientifique BMC Pulmonary Medicine et d'une communication orale au congrès international du CPLF (21<sup>ème</sup> édition du congrès de pneumologie de langue française) en janvier 2017 à Marseille. L'article est résumé dans l'encadré suivant, et est présenté dans son intégralité dans les pages suivantes. Le supplément de cet article est disponible en Annexe.

La BPCO est responsable d'une morbidité et d'une mortalité lourde. Dans ce contexte, l'évaluation des comorbidités est un élément majeur de la prise en charge. Il existe peu de données sur les profils de comorbidités associés à la sévérité de la BPCO en vraie vie. L'objectif de cette étude est d'analyser les clusters de comorbidités chez des patients BPCO en fonction de leur stade de sévérité.

Les données utilisées ont été recueillies prospectivement dans le cadre de la cohorte Palomb. Les pneumologues ont rempli un questionnaire en ligne comportant les données cliniques, fonctionnelles, les comorbidités et leur traitement. La classification de la BPCO a été faite selon GOLD 2011 (dyspnée, exacerbation, VEMS). Après l'analyse univariée, une analyse en cluster a été faite afin de déterminer les profils de comorbidités. Nous avons effectué une analyse en composantes principales, puis implémenté un algorithme non supervisé de clustering non hiérarchique (K-moyennes) pour identifier les profils de comorbidités. Pour mieux définir le nombre de clusters appropriées, nous avons utilisé trois méthodes statistiques : la méthode des éboulis, le pourcentage de variance expliqué et la méthode de Kaiser-Guttman. Enfin, nous avons réalisé une analyse en clusters (méthode de Ward) pour assurer la stabilité des différents clusters.

Au total, 1584 patients ont été inclus dans cette analyse durant les deux premières années d'inclusion. La distribution de la sévérité était selon GOLD 2011 est la suivante : 27,4 % stade A, 24,7 % stade B, 11,2 % stade C, et 36,6 % stade D sans différence notable entre les hommes et les femmes. La moyenne d'âge est de 66,56 ans (SD : 11), et la population compte 30 % de femmes. Au total, 28,4 % des patients n'avait aucune comorbidité. En moyenne, les patients avaient deux comorbidités associées [min : 0, max : 6]. La proportion de patients ayant plus de deux comorbidités était significativement plus élevée chez les patients de stade B (50,4 %) et D (53,1 %) comparativement aux patients de stade A (35,4 %) et C (34,3 %). L'HTA, le SAOS, la dyslipidémie, la cardiopathie ischémique et les troubles du rythme étaient significativement plus fréquents chez les patients GOLD B et D, comparativement aux patients GOLD A et C. L'analyse en cluster a permis de différencier 5 clusters de comorbidités :

- cardiovasculaire (22,7 %) ;
- syndrome métabolique et SAOS (14,7 %) ;
- dénutrition (20,6 %) ;
- patient avec bronchectasie (14,7 %) ;
- moins de comorbidités (27 %).

L'ensemble des clusters était retrouvé dans tous les stades de sévérité de la BPCO, cependant la fréquence de chaque cluster variait significativement en fonction du stade sévérité de la BPCO. L'analyse des comorbidités dans la BPCO reste complexe. Notre étude montre qu'au sein de la cohorte Palomb, la multimorbidité est fréquente, hétérogène et varie en fonction du stade de sévérité de la BPCO.

- **Article :**

Chantal Raheison, El-Hassane Ouaalaya, Alain Bernady, Julien Casteigt, Cecilia Nocent-Eijnani, Laurent Falque, Frédéric Le Guillou, Laurent Nguyen, Annaig Ozier and Mathieu Molimard. Comorbidities and COPD severity in a clinic-based cohort. BMC Pulm Med. 2018 Jul 16; 18(1) :117. DOI: <https://doi.org/10.1186/s12890-018-0684-7>

- **Communication orale :**

C. Raheison-Semjen, H. Ouaalaya, A. Bernady, J. Casteigt, C. Nocent-Eijnani, L. Falque, F. Le-Guillou, L. Nguyen, J. Moinard, A. Ozier, C. Roy, E. Berteaud, M. Molimard, F. Pellet. Profils de comorbidités et sévérité de la BPCO : données de la cohorte PALOMB. Revue des Maladies Respiratoires ; Volume 34, Supplement, January 2017, Pages A4-A5. DOI: <https://doi.org/10.1016/j.rmr.2016.10.014> [CPLF - Marseille 2017 (21e édition du Congrès de Pneumologie de Langue Française) : sélection via résumé]

RESEARCH ARTICLE

Open Access



# Comorbidities and COPD severity in a clinic-based cohort

Chantal Raherison<sup>1,2,10\*</sup> , El-Hassane Ouaalaya<sup>1</sup>, Alain Bernady<sup>3</sup>, Julien Casteigt<sup>4</sup>, Cecilia Nocent-Eijnani<sup>5</sup>, Laurent Falque<sup>6</sup>, Frédéric Le Guillou<sup>7</sup>, Laurent Nguyen<sup>8</sup>, Annaig Ozier<sup>8</sup> and Mathieu Molimard<sup>9</sup>

## Abstract

**Background:** Chronic obstructive pulmonary disease (COPD) is an important cause of morbidity and mortality around the world. The aim of our study was to determine the association between specific comorbidities and COPD severity.

**Methods:** Pulmonologists included patients with COPD using a web-site questionnaire. Diagnosis of COPD was made using spirometry post-bronchodilator FEV1/FVC < 70%. The questionnaire included the following domains: demographic criteria, clinical symptoms, functional tests, comorbidities and therapeutic management. COPD severity was classified according to GOLD 2011. First we performed a principal component analysis and a non-hierarchical cluster analysis to describe the cluster of comorbidities.

**Results:** One thousand, five hundred and eighty-four patients were included in the cohort during the first 2 years. The distribution of COPD severity was: 27.4% in group A, 24.7% in group B, 11.2% in group C, and 36.6% in group D. The mean age was 66.5 (sd: 11), with 35% of women. Management of COPD differed according to the comorbidities, with the same level of severity. Only 28.4% of patients had no comorbidities associated with COPD. The proportion of patients with two comorbidities was significantly higher ( $p < 0.001$ ) in GOLD B (50.4%) and D patients (53.1%) than in GOLD A (35.4%) and GOLD C ones (34.3%). The cluster analysis showed five phenotypes of comorbidities: cluster 1 included cardiac profile; cluster 2 included less comorbidities; cluster 3 included metabolic syndrome, apnea and anxiety-depression; cluster 4 included denutrition and osteoporosis and cluster 5 included bronchiectasis. The clusters were mostly significantly associated with symptomatic patients i.e. GOLD B and GOLD D.

**Conclusions:** This study in a large real-life cohort shows that multimorbidity is common in patients with COPD.

**Keywords:** COPD, Comorbidities, Cluster analysis, Management

## Background

COPD has emerged as the most important respiratory disease worldwide. The epidemiology of COPD had changed in recent years, with more women affected [1], fewer old subjects and more medications available for health providers.

To improve the management of COPD and take into account the heterogeneity of the disease, the Global Obstructive Lung Disease Initiatives [2] proposed a new

classification in 2011 that takes into account respiratory symptoms, the burden of exacerbations and lung function.

Comorbidities in COPD have received considerable attention as COPD patients frequently suffer from comorbidities such as cardiovascular and cerebrovascular disease, lung cancer and diabetes, with a significant impact on mortality that was termed by Divo et al. known as the “comorbidome” [3]. They constructed a comorbidity index (COTE index) based on 12 comorbidities that seem to negatively influence survival. However, the use of indexes like the COTE and the BODE [4] in clinical practice needs to be clarified. The validity of the COTE has been questioned since patients with GOLD B

\* Correspondence: Chantal.rahherison@chu-bordeaux.fr

<sup>1</sup>Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, team EPICENE, UMR 1219, F-33000 Bordeaux, France

<sup>2</sup>Pole cardiopulmonaire, Respiratory Diseases Department, CHU de Bordeaux, F-33000 Bordeaux, France

Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

seem to have worse survival than patients with GOLD C, because of the particular heart disease found in a very large population study in Copenhagen area [5]. Some authors suggest that the presence of comorbidities should influence the relationship between the GOLD score and lung function measurements, the former perhaps being more representative of morbidity than of COPD severity [6].

The distribution and the type of comorbidities seem to vary between studies, except for cardiovascular disease which seems to be stable across them [7]. The complexity of COPD was reported in the Eclipse cohort, suggesting that COPD includes several different phenotypes when taking into account clinical parameters, survival, hospitalization, comorbidities and systemic inflammation [8]. Recently, in a complex analysis using network analysis, Divo et al. showed that comorbidities in COPD do not occur by chance [9].

In 213 patients included in a rehabilitation center, Vanfleteren et al. identified 13 comorbidities and five comorbidity clusters: less comorbidity, cardiovascular, cachectic, metabolic and psychological [10]. However, little is known about the reproducibility of these comorbidity phenotypes in COPD patients in real life and their association with COPD severity. Recently, the ATS/ERS consensus statement recommended that studies be performed either to confirm or rule out an association between specific comorbidities and COPD [11]. In an ongoing prospective observational cohort of outpatients with COPD followed up by pulmonologists, the aim of our study was to determine the association between specific comorbidities and COPD severity.

## Methods

### Study design and population

The Palomb cohort is an ongoing, prospective, multicenter, observational study of subjects with COPD recruited in pulmonary clinics in South-West of France in a real-life setting since the first January 2014 and followed up for 3 years (Additional file 1).

The CNIL (National Data Protection and Privacy Commission) and the CCTIRS (Advisory Committee for Data Processing in Health Research) approved the study, and informed consent was obtained before enrollment. The authors had asked the local ethics committee for feedback regarding the need for ethical clearance for such a retrospective analysis, and were advised that this was not warranted.

Consent for publication statement is not applicable as no personal information is provided in this manuscript.

Between January 2014 and February 2016,  $n = 1584$  patients were enrolled in the study by pulmonologist and followed up in everyday practice, the data was obtained

using a web-site questionnaire fulfilled by the pulmonologist on a secure platform and with specific agreement obtained for the storage of health data.

The inclusion criterion was a diagnosis of COPD on the basis of a lung function test according to the ATS/ERS standards [12] and made using spirometry with post-bronchodilator FEV1/FVC < 70%. Patients were excluded if they didn't have the lung function criteria of COPD.

### Measurements

The website questionnaire included the following domains: demographic criteria (age, gender), smoking habits, clinical symptoms (mMRC dyspnea, chronic cough, exacerbations during past 12 months), body mass index (BMI) [4], lung function, comorbidities and therapeutic management (vaccinations, pulmonary rehabilitation, smoking cessation and prescribed inhaler medication).

### Comorbidities

Comorbidities ( $n = 19$ ) were recorded systematically in a standardized manner by the pulmonologist. The diagnosis of comorbidity was assessed first by patient report then confirmed by either reviewing the patient's medication list or when complementary tests were available in medical records. Bronchiectasis was recorded by clinical and/or radiologic criteria, as usual in clinical practice.

### COPD severity

Severity of COPD was classified as A,B,C or D according to GOLD 2011 [2]: 1) high/low symptoms using the mMRC dyspnea score < or  $\geq 2$ ; 2) the severity of airflow limitation [13]; and 3) the number of exacerbations per year. Despite the recent publication of GOLD 2017, GOLD 2011 was chosen because this classification was used by the pulmonologist in clinical practice during the study.

### Statistical analysis

Analysis of variance was used for continuous variables and Chi-squared tests were used for categorical variables. We performed a univariate analysis between comorbidities and COPD severity (GOLD 2011). We decided to retain for further analysis only the comorbidities that were significantly more frequent in more severe COPD stages (B,C,D) than in the mild stage (A). Then we performed a principal component analysis and a non-hierarchical (K-means) cluster analysis to describe the cluster of comorbidities. To better define the number of appropriate clusters, we used three statistical methods: the scree plot method, the percentage of variance explained, and the Kaiser-Guttman method.

Lastly, we performed a cluster analysis (ward method) to ensure stability of the different clusters. All analyses were performed with SAS software version 9.4. All statistical tests were two sided, with  $P < 0.05$  considered to indicate statistical significance.

**Results**

**Patients' characteristics**

A total of 1584 patients were included in the cohort during the first 2 years.

The GOLD 2011 distribution was as follows: 27.4% in group A, 24.7% in group B, 11.2% in group C, and 36.6% in group D, with no significant gender difference. The mean age was 66.5 (sd: 11), with 30% of women. Clinical symptoms, exacerbations, lung function and management according to COPD severity are presented in Table 1. 28.7% of the patients had had 2 or more exacerbations during the past year.

The onset of symptoms (i.e. cough, exacerbations or dyspnea) occurred between 25 and 49 years in 6% of the patients, 49–59 years in 20%, 59–69 years in 35%, 69–79 years in 25.3% and after 79 years in 13.26%.

16% had a BMI below 21 kg/m<sup>2</sup>, 34.8% between 21 and 26 kg/m<sup>2</sup>, 20% were overweight and 28.7% were obese (BMI > 29 kg/m<sup>2</sup>). BMI was significantly associated with severity of dyspnea and age (data not shown).

Prescription of pulmonary rehabilitation was rather infrequent in the whole population and was significantly more frequent in GOLD D patients (Table 1). By contrast, influenza vaccination was more frequent in GOLD B and D patients than pneumococcal vaccination, which increased with COPD severity (trend).

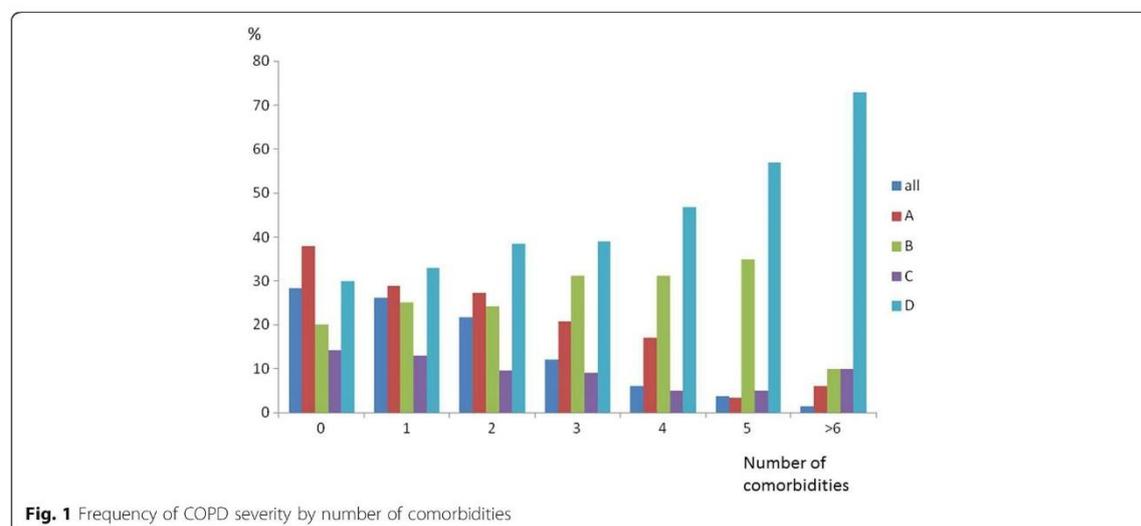
**Frequency of comorbidities**

Only 28.4% of patients had no comorbidities associated with COPD, whatever the severity of their COPD. The

**Table 1** Description of 1584 subjects with COPD according to GOLD 2011 Classification (frequency of each variable by COPD severity)

	A N = 435	B N = 391	C N = 178	D N = 580	<i>p</i>
Males, %	68.5	62.6	62.9	64.6	0.3006
Age, yr (SD)	63.2(10.4)	69(10.8)	63.6(10.6)	68.2(11)	0.0006
BMI, (kg/m <sup>2</sup> ), % < 21					0.0001
(underweight) [21–26](normal)	12.4	13.3	21.3	19.6	
[26–29](overweight)	37.7	30.2	33.2	36.4	
> 29 (obese)	21.8	21.2	22.5	17.6	
	28.3	35.3	23	26.4	
FEV1, % pred					0.0001
> 80%	109(25)	55(14)	26(14.6)	13(2.2)	
50–80%	326(75)	336(86)	62(34.9)	141(24.3)	
30–50%	–	–	85(47.8)	328(56.5)	
< 30%	–	–	5(2.8)	98(16.9)	
mMRC 0–1, n(%)	435(100)	0	178(100)	0	0.0001
mMRC > =2, n(%)	0	391(100)	0	580(100)	0.0001
Chronic cough, n(%)	190(43.6)	222(56.7)	96(53.9)	402(69.3)	< 0.0001
Current smokers, n(%)	170(39.8)	131(34.5)	68(39.5)	195(34.5)	0.2394
0–1 exacerbation previous year, n(%)	435(100)	391(100)	71(40)	231(39.8)	0.0001
> = 2 exacerbations, previous year, n(%)	0	0	107(60)	349(60.2)	0.0001
Pulmonary rehabilitation, n(%)	12(2.7)	23(5.8)	10(5.6)	97(16.7)	< 0.0001
Smoking cessation, n(%)	35(8.0)	23(5.8)	26(14.6)	58(10)	0.0053
SABA, %	28.7	44.2	37	47	< 0.0001
LABA, or LAMA, %	53.1	72.1	66.8	71.2	< 0.0001
ICS and LABA, %	18.1	24.8	24.2	39	< 0.0001
Annual influenza vaccination, n(%)	166(38.1)	230(58.8)	93(52.2)	410(70.6)	< 0.0001
Pneumococcal vaccination, n(%)	141(32.4)	204(52.1)	100(56.1)	379(65.3)	< 0.0001

Abbreviations: SABA short-acting bronchodilators LABA long-acting bronchodilators. LAMA long-acting muscarinic antagonist



number of comorbidities by COPD severity is shown in Fig. 1, the number of comorbidities increased with severity of COPD. Cardiac comorbidities were more frequent in men whereas anxiety-depression and osteoporosis were more frequent in women.

Hypertension, ischemic cardiopathy, heart rhythm disorder and left cardiac insufficiency were significantly higher in overweight and obese subjects ( $p$ : 0.0001).

Prevalence of obstructive apnea syndrome (OAS) was higher in group A. Hypertension, OAS, dyslipidemia,

**Table 2** Frequency of comorbidities in 1584 subjects with COPD according to GOLD 2011 Classification (frequency of each variable by COPD severity)

	A N = 435	B N = 391	C N = 178	D N = 580	$p$
Hypertension	139(31.9)	173(44.2)	52(29.2)	237(40.8)	0.0001
Obstructive apnea syndrome	118(27.1)	75(19.2)	23(12.9)	63(10.9)	0.0001
Dyslipidemia	97(22.3)	114(29.2)	37(20.8)	143(24.7)	0.072
Cancer, all causes	79(18.2)	70(17.9)	20(11.2)	90(11.5)	0.14
Ischemic cardiopathy	59(12.5)	99(25.3)	30(16.8)	127(21.9)	0.0001
Past asthma	47(10.8)	48(12.3)	19(10.7)	71(12.3)	0.84
Depression	43(9.9)	50(12.8)	15(8.4)	113(19.5)	0.0001
Anxiety	30(6.9)	70(17.9)	23(12.9)	154(26.5)	0.0001
Heart rhythm disorder	41(9.4)	65(16.6)	24(13.4)	86(14.8)	0.016
Diabetes	41(9.4)	51(13)	18(10.1)	88(15.2)	0.035
Undernutrition, BMI < 21 kg/m <sup>2</sup>	1(0.2)	3(0.7)	4(2.2)	25(4.3)	0.0001
Osteoporosis	2(0.4)	23(5.8)	1(0.5)	35(6.0)	0.0001
Atheroma	20(4.6)	27(6.9)	9(5.0)	51(8.8)	0.048
Left cardiac insufficiency	8(1.8)	21(5.3)	12(6.7)	50(8.6)	0.0001
Vascular Stroke	16(3.7)	22(5.6)	5(2.8)	20(3.5)	0.26
Pulmonary hypertension	3(0.7)	7(1.8)	1(0.6)	26(4.5)	0.0002
Metabolic syndrome	10(2.3)	16(4.1)	5(2.8)	27(4.7)	0.21
Rhinitis/rhinosinusitis	8(3.2)	14(7)	7(7.8)	19(7.4)	0.16
Bronchiectasis	7(3.9)	31(4.1)	34(9.2)	6(6.4)	0.004

**Table 4** Comorbidities frequency (> = 2 vs 0–1) by COPD severity (statistical test to compare the distribution of comorbidity frequency in each severity stage)

Comorbidity frequency	GOLD 2011				p value
	A N = 435	B N = 391	C N = 178	D N = 580	
0–1864(54.55%)	281(32.5%)	194(22.4%)	117(13.5%)	272(31.4%)	< 0.0001
≥2720(45.45%)	154(21.3%)	197(27.3%)	61(8.4%)	308(42.7%)	

ischemic cardiopathy and heart rhythm disorder were more frequent in groups B and D than in groups A and C. The frequencies of all comorbidities for COPD stage are presented in Table 2.

We have identified 13 comorbidities which were more frequent in higher COPD stages (B,C,D) than in the mild stage (A). OAS was more frequent in mild stage (A) than in others.

Anxiety and depression was higher in groups D and B than in the other groups. Undernutrition was higher in group D. Osteoporosis was higher in groups B and D.

**Number of comorbidities**

In the group of patients with one comorbidity, 36.7% had hypertension, 11.8% had Obstructive Syndrome Apnea (OSA), 10.4% had depression and 10% had ischemic cardiopathy. The proportion of patients with two comorbidities was significantly higher ( $p < 0.001$ ) in GOLD B (50.4%) and D patients (53.1%) than in GOLD A (35.4%) and C ones (34.3%). The median of comorbidities was 1.6 (box plot) in the whole sample.

The proportion of GOLD B and D patients, increased significantly with the number of comorbidities, particularly among those with more than two comorbidities (Fig. 1). The number of comorbidities was higher in GOLD B and D patients (Table 3). The number of patients with two comorbidities or more was significantly higher in patients GOLD B and D patients than in GOLD A and C ones (Table 4).

**Management of COPD according to number of comorbidities**

In GOLD A and B patients, prescription of treatment as needed and regular treatment did not differ according to

the number of comorbidities (Table 5), unlike for GOLD C and D patients. In GOLD C patients, LABA were more frequently prescribed in those with comorbidities than in those without. In GOLD D patients, SABA was prescribed significantly more frequently in those with comorbidities. Pulmonary rehabilitation and vaccination were prescribed significantly more in GOLD B and D patients with comorbidities than in those with the same degree of severity but without comorbidities. Finally, smoking cessation was prescribed significantly more in GOLD C and D patients with comorbidities.

**Comorbidity clusters**

The cluster analysis showed five phenotypes of comorbidities: cluster 1 included cardiac profile; cluster 2 included less comorbidity; cluster 3 included metabolic syndrome, apnea and anxiety-depression; cluster 4 included cachectic and osteoporosis and cluster 5 included mainly bronchiectasis. The label of each cluster was given, comparing the prevalence of comorbidity in the whole population with prevalence of comorbidity in each cluster (Table 6).

The different clusters were distributed in the four stages of COPD severity, however cardiac cluster was more frequent in patients with GOLD B. Cluster with less comorbidity was more frequent in patient GOLD C and A. Metabolic syndrome was more frequent in GOLD C and D. Cachectic and osteoporotic profile were most frequent in GOLD B and D. Lastly, bronchiectasis profile was more frequent in patient GOLD D (Table 7).

**Discussion**

This study sought to determine whether comorbidities were associated with COPD severity in a clinic-based

**Table 3** Number of comorbidities n (%) by COPD severity (frequency of Number of comorbidities in each COPD severity stage)

Number of comorbidities	A N = 435	B N = 391	C N = 178	D N = 580	p
0	160 (36.7)	90(23)	64(35.9)	136(23.4)	< 0.0001
1	121(27.8)	104(26.6)	53(29.8)	136(23.4)	< 0.0001
2	94(21.6)	83(21.2)	33(18.5)	132(22.7)	< 0.0001
3	40(9.2)	60(15.3)	17(9.5)	75(12.9)	< 0.0001
4	16(3.7)	30(7.7)	5(2.8)	45(7.8)	< 0.0001
5	2(0.4)	21(5.4)	3(1.7)	34(5.9)	< 0.0001
6 and more	2(0.4)	3(0.7)	3(1.7)	22(2.1)	< 0.0001

**Table 5** COPD treatment by number of comorbidities and COPD severity (P value refers to compare each pharmacologic treatment according to number of comorbidities in each severity stage)

Number of comorbidities	A N = 435			B N = 391			C N = 178			D N = 580		
	0-1 N = 281	>=2 N = 154	p	0-1 N = 194	>=2 N = 197	p	0-1 N = 117	>=2 N = 61	p	0-1 N = 272	>=2 N = 308	p
SABA, n(%)	83(29.5)	42(27.2)	0.61	81(41.7)	92(46.7)	0.32	46(39.3)	20(32.8)	0.39	94(34.5)	179(58.2)	<0.0001
LABA or LAMA, n(%)	152(54.0)	79(51.3)	0.57	139(71.6)	143(72.6)	0.83	72(61.5)	47(77.0)	0.03	186(68.4)	227(73.7)	0.15
ICS and LABA, n(%)	53(18.9)	26(16.9)	0.60	55(28.3)	42(21.3)	0.10	31(26.5)	12(19.7)	0.31	102(37.5)	124(40.3)	0.49
LABA and LAMA, n(%)	30(10.7)	23(14.9)	0.02	26(13.4)	38(19.3)	0.01	18(15.4)	19(31.1)	0.02	45(16.5)	94(30.5)	0.01
LABA and LAMA and ICS, n(%)	1(0.35)	1(0.64)	0.35	2(1.03)	2(1.01)	0.35	2(1.7)	2(3.2)	0.35	8(2.9)	21(6.8)	0.02
Pulmonary Rehabilitation, n(%)	8(2.8)	4(2.6)	0.87	6(3.9)	17(8.6)	0.02	7(6)	3(5)	0.76	31(11.4)	66(21.4)	0.0012
Annual influenza vaccination, n(%)	105(37.4)	61(39.6)	0.64	95(49)	135(68.5)	<0.0001	60(51.3)	33(54.1)	0.72	167(61.4)	243(79)	<0.0001
Pneumococcal vaccination, n(%)	85(30.2)	56(36.3)	0.19	81(41.7)	123(62.4)	<0.0001	65(55.5)	35(57.3)	0.81	145(53.3)	234(76)	<0.0001
Smoking cessation, n(%)	24(8.5)	11(7.14)	0.60	13(6.7)	10(5.0)	0.49	13(11.1)	13(21.3)	0.06	17(6.2)	41(13.3)	0.0047

**Table 6** Prevalence of comorbidities in the five clusters

comorbidities (% in the whole population)	Cluster1 N = 360	Cluster2 N = 430	Cluster3 N = 233	Cluster4 N = 327	Cluster5 N = 234
OSA (17.6)	69(24.7)	36(12.9)	74(26.5)	60(21.5)	40(14.3)
Bronchiectasis (5.1)	17(21)	6(7.4)	20(24.7)	17(21)	21(25.9)
Left cardiac insufficiency (5.7)	19(20.9)	17(18.7)	23(25.3)	16(17.6)	16(17.6)
Hypertension (37.9)	130 (21.6)	98(16.3)	164 (27.3)	136 (22.6)	73(12.1)
Heart rhythm disorder (13.64)	48(22.2)	31(14.4)	54(25)	51(23.6)	32(14.8)
Atheroma (6.76)	23(21.5)	12(11.2)	35(32.7)	18(16.8)	19(17.8)
Pulmonary hypertension (2.3)	15(40.5)	4(10.8)	8(21.6)	5(13.5)	5(13.5)
Diabetes (12.5)	47(23.7)	26(13.1)	56(28.3)	51(25.8)	18(9.1)
Depression (13.95)	46(20.8)	33(14.9)	66(29.9)	44(19.9)	32(14.5)
Anxiety (17.5)	62(22.4)	15(5.4)	87(31.4)	67(24.2)	46(16.6)
Undernutrition (2)	10(30.3)	1(3)	8(24.2)	10(30.3)	4(12.1)
Ischemic cardiopathy (19.9)	61(19.4)	46(14.6)	83(26.3)	72(22.9)	53(16.8)
Osteoporosis (3.8)	13(21.3)	3(4.9)	14(23)	21(34.4)	10(16.4)

cohort of COPD patients, mostly GOLD A and B, followed up by pulmonologists. Only 28.4% of patients had no associated comorbidities. Fourteen comorbidities were significantly different with COPD severity. In this large population of patients, the median number of comorbidities was two.

Hypertension, OSA, dyslipidemia, ischemic cardiopathy and heart rhythm disorder were more frequent in GOLD B and D patients than in the other groups, as were anxiety and depression. Undernutrition was the most frequent in GOLD D patients and osteoporosis was the most frequent in GOLD B and D subjects. The number of comorbidities was the highest in GOLD B and D patients. Even when the severity of symptoms was similar, the management of COPD seemed to be different according to whether patients had comorbidities or not. Finally, five clusters of

comorbidities were established, the most frequent being the cluster with cardiovascular disease and obstructive apnea syndrome.

Our approach was to analyze the relationship between these comorbidities and COPD severity by using three different approaches: the impact of the number of comorbidities, the univariate association between the comorbidities and COPD severity and cluster analysis to determine the association between the comorbidities i.e. to establish the existence of different phenotypes. Our findings are consistent with previous publications reporting a high prevalence of comorbidity in COPD, particularly cardio-vascular disease. Chen et al. [7] in a large review reported that compared with the non-COPD population, patients with COPD were more likely to be diagnosed with cardiovascular disease (odds ratio [OR] 2.4; 95% CI 2.02–3.00;  $p < 0.0001$ ), including a two- to five-fold higher risk of

**Table 7** Distribution of comorbidities cluster by COPD severity

Clusters	GOLD 2011				p value
	A N = 435	B N = 391	C N = 178	D N = 580	
Cluster 1 Cardiac (22.7%)	107(24.6)	102(26)	33(18.5)	118(20)	< 0.0001
Cluster 2 less comorbidity (27%)	137(31.5)	96(24.5)	57(32)	140(24.1)	< 0.0001
Cluster 3 metabolic, apneic and anxiety-depression (14.7%)	56(12.8)	39(9.9)	41(23)	97(16)	< 0.0001
Cluster 4 Cachectic and osteoporosis (20.6%)	70(16)	101(25.8)	22(12.3)	134(23)	< 0.0001
Cluster 5 bronchiectasis (14.7%)	65(14.9)	53(13.5)	25(14)	91(15.6)	< 0.0001

ischemic heart disease, cardiac dysrhythmia, heart failure, diseases of the pulmonary circulation, and arterial disease. Additionally, patients with COPD reported hypertension more often (OR 1.3, 95% CI 1.1–1.5;  $p = 0.0007$ ), diabetes (1.3, 1.2–1.5;  $p < 0.0001$ ), and ever smoking (4.2, 3.2–5.6;  $p < 0.0001$ ). Divo et al. found in their cohort that cardiovascular disease was highly associated with the risk of mortality, but that the highest risk of mortality was associated with anxiety [3]. However, we found a high prevalence of OAS, probably owing to the overlap syndrome as reported by Soler et al. [14]. By contrast, OAS was the most frequent comorbidity in GOLD A and B patients although it seemed to be associated with moderate-to-severe COPD. It is essential to diagnose OAS in patients with COPD as patients with overlap syndrome who are not treated with CPAP have a higher mortality [15].

In our cohort, 14 comorbidities were significantly associated with COPD severity. Hypertension, OSA, dyslipidemia, ischemic cardiopathy, and heart rhythm disorder were more frequent in GOLD B and D compared to GOLD A and C. Anxiety and depression was higher in GOLD D and B, compared to the other groups of severity. These results are in line with the analysis performed in the Copenhagen cohort showing that GOLD B patients had more severe dyspnea and significantly poorer survival than group C ones, in spite of a higher FEV1 level [5]. The same trend concerned the number of comorbidities, with a prevalence of comorbidities (more than two) in GOLD B and D patients. At an equal level of severity, management of COPD seems to be different in severe COPD patients with comorbidities, with more LABA and SABA in severe COPD, suggesting that comorbidities could increase respiratory symptoms. Moreover, prevention of exacerbations requires interventions beyond the lungs, including treatment of comorbidities such as gastro-esophageal reflux disease, reduction of cardiovascular risks, and management of dyspnea and anxiety [16].

LABA were prescribed the most in GOLD A patients, which is not in agreement with the guidelines GOLD 2011. This could be due to the high prevalence of symptoms like cough in this group, as symptoms included in the GOLD classification are based on dyspnea and exacerbations but not on cough. We cannot rule out that symptomatic GOLD A patients could represent a specific phenotype. Recently, Woodruff et al. described a subgroup of symptomatic patients with no criteria for COPD regarding lung function [17]. In addition, management of COPD differed according to the comorbidities that patients had, even if those with or without had the same level of severity. This was particularly the case for rehabilitation and vaccination which were more prescribed in symptomatic GOLD B and D patients who had comorbidities than in those without.

We expected to have a gradient in COPD severity, perhaps patients GOLD B should be called differently, as they seemed to be more severe than GOLD C.

The cluster analysis revealed five clusters: The cluster analysis showed five phenotypes of comorbidities: cluster 1 included cardiac profile; cluster 2 included less comorbidities; cluster 3 included metabolic syndrome, apnea and anxiety-depression; cluster 4 included undernutrition and osteoporosis and cluster 5 included bronchiectasis. Vanfleteren found 13 comorbidities in a sample of 213 COPD patients [10] with five comorbid phenotypes: less comorbidity, cardiovascular, cachectic, metabolic, and psychological. Four of our clusters are concordant, i.e. cardiovascular, cachectic, metabolic and less comorbidities. Nevertheless, all the clusters were more significantly associated with GOLD D and in less manner with GOLD B. This finding could explain the higher risk of mortality in GOLD B and D patients, as previously reported elsewhere [5]. In the same way, Divo et al. [9] also identified a number of modules in the comorbidity network, including a cardiovascular one, and a module characterized by mild-moderate airflow limitation and metabolic syndrome with high BMI, these two modules are concordant with our findings.

Our results show that while comorbidity in COPD is a complex issue, comorbidities contributed prominently to the clinical severity of our patients, and that management of their severe COPD differed according to whether they had comorbidities or not, at the same level of obstruction.

Our study has some limitations. First, comorbidities were recorded by pulmonologists; previous studies showed that comorbidities are underdiagnosed in real life. This could also be the case in our study for most comorbidity except for OSA, as OSA was diagnosed by a polysomnography performed by the same pulmonologist who diagnosed COPD. However, we think that this bias is limited in our study, as we found a significant correlation between the comorbidities declared and compliance with the treatment given for them (data not shown). Second, we cannot generalize these findings to patients with COPD in the general population, as our population was managed both by a general practitioner and a pulmonologist. Third, we performed a multivariable exploratory analysis in order to better describe the associations between the different comorbidities [18]. This type of analysis uses a statistical method that processes a large amount of information from heterogeneous variables in homogeneous groups. It is well known that various factors can influence the analysis and therefore the results: the choice of the methods, hierarchical or nonhierarchical, the determination of the number of clusters before the analysis, the choice of the variables included in the analysis, the correlation between the selected variables and the clinical judgment of

the investigators. To limit the impact of the specific correlation between the variables, we first performed a principal component analysis. Then we used the scree plot of the eigenvalue, the Kaiser-Guttman criterion and the percentage of variance explained to determine the number of clusters.

Lastly, we cannot validate our clusters in terms of survival as the study was performed with inclusion criteria, or with systemic inflammation. Further analysis with survival data from these COPD patients would provide important information for validating these clusters of comorbidity.

## Conclusions

This study in a large clinic-based cohort shows that multimorbidity is common in patients with COPD, and that five comorbidity clusters can be identified. Patients GOLD B have more comorbidities than GOLD C.

The presence of comorbidities should therefore be included in any assessment of COPD severity. Further analysis is needed to validate these clusters in a future cohort.

## Additional file

**Additional file 1:** PALOMB English questionnaire (PDF 12 kb)

## Abbreviations

ATS/ERS: American Thoracic Society/European Respiratory Society; BODE: Bmi Obstruction Dyspnea Exercise; COPD: Chronic obstructive pulmonary disease; COTE index: Comorbidity index; FEV1: Forced expiratory Volume; GOLD: Global Obstructive Lung Disease Initiatives; ICS: Inhaled corticosteroids; LABA: Long acting bronchodilator; LAMA: Long acting muscarinic antagonist; OSA: Obstructive Syndrome Apnea; SABA: Short acting bronchodilator

## Acknowledgements

The authors dedicate this manuscript to the memory and the contribution of their dear co-author and friend François Pellet, MD, for his outstanding contribution to the Palomb project. They also thank R. Goin and C. Bousquet, A. Le-Leon (Bordeaux University Foundation) for their support, R. Cooke for copyediting the manuscript and E. Berteaud for data management. Inclusion centers: C. Roy, J. Moinard, Y. Daoudi, JM Dupis, E. Blanchard, H. Jungmann E. Monge, A. Prudhomme, M. Sapene, M. Sabatini.

## Funding

Funding (unrestricted grants): Bordeaux University Foundation, Novartis Pharma, Isis Medical, Boehringer Ingelheim, Glaxo-Smith Kline. The funding sources had no role in the design or conduct of the study, in the collection, management, analysis and interpretation of the data, or in the preparation, review or approval of the manuscript.

## Availability of data and materials

The data that support the findings of this study are available from INSERM U1219 Epicene Team but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of INSERM U1219 Epicene Team.

## Authors' contributions

Study concept and design: CR, AB, JC, CN, LF, FLG, LN, MM. Acquisition of data: all authors. Statistical analysis: CR, EO. Data access and responsibility: CR and EO had full access to the data of the study and take responsibility for

the integrity of the data and the accuracy of the data analysis. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

The CNIL (National Data Protection and Privacy Commission) and the CCTIRS (Advisory Committee for Data Processing in Health Research) approved the study, and informed verbal consent was obtained before enrollment. The authors had asked the local ethics committee for feedback regarding the need for ethical clearance for such a retrospective analysis, and were advised that this was not warranted.

## Consent for publication

Not applicable.

## Competing interests

Dr. Raheison reports grants from Bordeaux University Foundation, during the conduct of the study; personal fees from Astra Zeneca, personal fees from Chiesi, personal fees from ALK, personal fees from Boehringer Ingelheim, personal fees from Glaxo SmithKline, personal fees from MundiPharma, personal fees from Novartis, outside the submitted work; Dr. Nocent-Ejnnani has nothing to disclose. Dr. Molimard reports personal fee from the University of Bordeaux, during the conduct of the study, other from Novartis Pharma, GSK, MundiPharma outside the submitted work. Dr. Nguyen has nothing to disclose. Dr. Falque has nothing to disclose. Dr. Casteigt has nothing to disclose. Dr. Le Guillou has nothing to disclose. Dr. Ozier has nothing to disclose. Dr. Bernady has nothing to disclose. Mr. Ouailaya has nothing to disclose.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, team EPICENE, UMR 1219, F-33000 Bordeaux, France. <sup>2</sup>Pole cardiopulmonaire, Respiratory Diseases Department, CHU de Bordeaux, F-33000 Bordeaux, France. <sup>3</sup>Rehabilitation Center, Cambo-les-Bains, France. <sup>4</sup>Pneumology Clinic, St Medard en Jalles, France. <sup>5</sup>General Hospital, Bayonne, France. <sup>6</sup>Pneumology Clinic, Bordeaux, France. <sup>7</sup>Pneumology Clinic, La Rochelle, France. <sup>8</sup>Pneumology Clinic, St Augustin, Bordeaux, France. <sup>9</sup>U1219 Pharmaco-epidemiology, Bordeaux University, Bordeaux, France. <sup>10</sup>Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, team EPICENE, UMR 1219, 146 rue Leo Saignat, 33076 Cedex Bordeaux, France.

Received: 7 February 2018 Accepted: 5 July 2018

Published online: 16 July 2018

## References

1. Thun MJ, et al. 50-year trends in smoking-related mortality in the United States. *N Engl J Med*. 2013;368(4):351–64.
2. National Heart, L., and Blood Institute. World Health Organization., Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. NHLBI/WHO Workshop Report 2011. *Vestbo J*: National Institutes of Health; 2014. Available from: <https://goldcopd.org/gold-reports-2017/>.
3. Divo M, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;186(2):155–61.
4. Celli BR, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med*. 2004;350(10):1005–12.
5. Lange P, et al. Prediction of the clinical course of chronic obstructive pulmonary disease, using the new GOLD classification: a study of the general population. *Am J Respir Crit Care Med*. 2012;186(10):975–81.
6. Weinreich UM, et al. The effect of comorbidities on COPD assessment: a pilot study. *Int J Chron Obstruct Pulmon Dis*. 2015;10:429–38.
7. Chen W, et al. Risk of cardiovascular comorbidity in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Lancet Respir Med*. 2015;3(8):631–9.

8. Rennard SI, et al. Identification of five chronic obstructive pulmonary disease subgroups with different prognoses in the ECLIPSE cohort using cluster analysis. *Ann Am Thorac Soc*. 2015;12(3):303–12.
9. Divo MJ, et al. COPD comorbidities network. *Eur Respir J*. 2015;46(3):640–50.
10. Vanfleteren LE, et al. Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2013;187(7):728–35.
11. Celli BR, et al. An official American Thoracic Society/European Respiratory Society statement: research questions in COPD. *Eur Respir J*. 2015;45(4):879–905.
12. Celli BR, MacNee W, Force AET. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J*. 2004;23(6):932–46.
13. Gold PM. The 2007 GOLD guidelines: a comprehensive care framework. *Respir Care*. 2009;54(8):1040–9.
14. Soler X, et al. High prevalence of obstructive sleep apnea in patients with moderate to severe chronic obstructive pulmonary disease. *Ann Am Thorac Soc*. 2015;12(8):1219–25.
15. Marin JM, et al. Outcomes in patients with chronic obstructive pulmonary disease and obstructive sleep apnea: the overlap syndrome. *Am J Respir Crit Care Med*. 2010;182(3):325–31.
16. Vanfleteren LE, et al. Management of chronic obstructive pulmonary disease beyond the lungs. *Lancet Respir Med*. 2016;4(11):911–24.
17. Woodruff PG, et al. Clinical significance of symptoms in smokers with preserved pulmonary function. *N Engl J Med*. 2016;374(19):1811–21.
18. Burgel PR, Paillasseur JL, Roche N. Identification of clinical phenotypes using cluster analyses in COPD patients with multiple comorbidities. *Biomed Res Int*. 2014;2014:420134.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



### III. Les déterminants de la dyspnée par une régression pénalisée et l'analyse de sensibilité (n=1973)

Cette partie porte sur les résultats de l'étude des déterminants de la dyspnée (évalué par l'échelle mMRC) dans la cohorte PALOMB. Les résultats relatifs à cette étude font l'objet d'une publication sous forme d'article original dans la revue scientifique *Respiratory Medicine and Research* ; de deux communications orales aux journées de l'école doctorale « SP2 » en Mai 2018 à Bordeaux et au congrès international de l'ERS (European Respiratory Society Congress) en septembre 2018 à Paris ; et de deux communications affichées (Posters) à la conférence internationale de l'ATS (American Thoracic Society) en Mai 2019 à Dallas (USA) et au congrès international du CPLF (22<sup>ème</sup> édition du congrès de pneumologie de langue française) en janvier 2018 à Lyon. L'article est résumé dans l'encadré suivant, et est présenté dans son intégralité dans les pages suivantes. Le supplément de cet article est disponible en Annexe.

La dyspnée est un des symptômes cardinaux de la BPCO, et fait partie des éléments d'évaluation pour apprécier la sévérité de la BPCO et décider de la prise en charge thérapeutique. L'évaluation de la dyspnée en pratique clinique n'est pas chose aisée, et il est parfois difficile d'apprécier la part attribuable respiratoire de la dyspnée chez des patients ayant des comorbidités. L'objectif de ce travail est d'évaluer les déterminants de la dyspnée chez des patients BPCO, inclus dans le cadre de la cohorte PALOMB.

Les patients BPCO ont été inclus depuis le janvier 2014 par leur pneumologue par web questionnaire (cohorte PALOMB). Les variables suivantes ont été recueillies : âge, sexe, l'IMC, la sévérité de la BPCO (VEMS), VR, CI, CPT, CRF, la dyspnée (échelle mMRC), les comorbidités et la fréquence des exacerbations. Une analyse descriptive a été réalisée puis une analyse multivariée afin de déterminer les facteurs associés à la dyspnée définie par mMRC  $\geq 2$ . Le LASSO a été utilisé pour sélectionner les variables associées à l'échelle de dyspnée mMRC dans un sous-groupe (qui n'avait pas de valeurs manquantes) de 421 patients BPCO définis par les variables mentionnées précédemment.

Au total, 1973 patients ayant une BPCO (VEMS/CVF  $< 70$  % post-BD) ont été inclus ; 65,3 % d'hommes avec un âge moyen de 66 ans. Trente-huit pour cent sont encore fumeurs actifs, 59 % ont eu au moins une exacerbation au cours des 12 derniers mois. Selon GOLD 2017, la sévérité de la BPCO était principalement représentée par 3 groupes : 21,2 % (grade II, groupe A), 20 % (grade II, groupe B), et 12,9 % (grade III ; groupe B). En analyse univariée, la dyspnée était plus fréquente chez les femmes, chez les patients en surpoids ou obèses et chez ceux ayant des comorbidités cardiovasculaires. La dyspnée était également associée significativement à un VEMS altéré, à un VR ( $> 130$  %) et au nombre d'exacerbations ( $> 2$ ). L'analyse multivariée avec ajustement sur l'âge, le tabagisme, montre que les déterminants de la dyspnée ( $\geq 2$  mMRC) sont : être une femme OR : 0,71 [0,57–0,9] ; la sévérité du VEMS OR : 3,7 [2,8–4,8] ; l'HTA OR : 1,5 [1,2–1,9] ; l'anxiété OR : 2,5 [1,8–3,4] ; les DDB OR : 1,84 [1,01–1,81] ; le nombre d'exacerbations ( $\geq 2$ ) : OR : 1,41 [1,1–1,81] ; le surpoids OR : 1,33 [1,06–1,67] ; et la distension thoracique OR : 1,68 [1,34–2,1].

Nos résultats ont montré que la dyspnée est liée à la sévérité de la BPCO, au sexe, aux exacerbations, aux comorbidités et à la distension thoracique. Ces résultats illustrent dans une cohorte de vraie vie l'aspect plurifactoriel de la dyspnée chez les patients BPCO.

- **Article :**

El Hassane Ouaalaya, Laurent Falque, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Ejnaini, Frédéric Le Guillou, Mathieu Molimard, Chantal Raheison-Semjen. The determinants of dyspnoea evaluated by the mMRC scale: The French Palomb cohort. *Respir Med Res.* 2020 Nov 22; 79: 100803. DOI: <https://doi.org/10.1016/j.resmer.2020.100803>

- **Communication orale :**

El-Hassane Ouaalaya, Laurent Falque, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Ejnaini, Frédéric Le Guillou, Mathieu Molimard, Chantal Raheison-Semjen ; Quels sont les déterminants de la dyspnée chez les patients BPCO de la cohorte PALOMB ? [Présentation orale aux journées de l'EDSP2 ; PUSG Bordeaux Bastide ; 16 mai 2018. Sélection via résumé]

El-hassane Ouaalaya, Emilie Berteaud, Alain Bernardy, Jean-Michel Dupis, Marielle Sabatini, Laurent Falque, Frédéric Le Guillou, Cecilia Nocent-Ejnaini, Laurent Nguyen, Annaig Ozier, Julien Casteigt, Mathieu Molimard, Chantal Raheison-Semjen. Determinants of dyspnea in COPD patients: The French PALOMB cohort. *European Respiratory Journal*, November 19, 2018. vol. 52 no. Suppl 62; DOI: <https://doi.org/10.1183/13993003.congress-2018.PA757> ERS International Congress -Paris- France.

- **Communication affichée (Poster) :**

E.H. Ouaalaya, L. Falque, J.M. Dupis, M. Sabatini, C. Nocent-Ejnaini, A. Bernady, L. Nguyen, A. Ozier, F. Le Guillou, M. Molimard, C. Raheison-Semjen. The Resting Pulmonary Hyperinflation: A Determinant of the MMRC Dyspnea Scale in COPD Patients in a Real-Life Cohort? *American Thoracic Society 2019 International Conference -Dallas- USA. Conference Paper, May 21, 2019.* DOI: [https://doi.org/10.1164/ajrccm-conference.2019.199.1\\_MeetingAbstracts.A4888](https://doi.org/10.1164/ajrccm-conference.2019.199.1_MeetingAbstracts.A4888)

C. Raheison, E. Ouaalaya, A. Bernady, J. Casteigt, J. M. Dupis, C. Nocent-Ejnaini, M. Sabatini, L. Falque, F. Le Guillou, L. Nguyen, A. Ozier, M. Molimard. Quels sont les déterminants de la dyspnée chez les patients BPCO de la cohorte PALOMB ? *Revue des Maladies Respiratoires. Volume 35, Supplement, January 2018, Pages A10-A11.* DOI: <https://doi.org/10.1016/j.rmr.2017.10.024> [Présentation orale au CPLF - Lyon 2018 (22e édition du Congrès de Pneumologie de Langue Française) : sélection via résumé]



Available online at

ScienceDirect  
www.sciencedirect.com

Elsevier Masson France

EM|consulte  
www.em-consulte.com



Original article

## The determinants of dyspnoea evaluated by the mMRC scale: The French Palomb cohort



E.H. Ouaalaya<sup>a,\*</sup>, L. Falque<sup>c</sup>, J.M. Dupis<sup>d</sup>, M. Sabatini<sup>e</sup>, A. Bernady<sup>f</sup>, L. Nguyen<sup>g</sup>, A. Ozier<sup>g</sup>, C. Nocent-Ejnaini<sup>e</sup>, F. Le Guillou<sup>h</sup>, M. Molimard<sup>i</sup>, C. Raheison-Semjen<sup>a,b</sup>

<sup>a</sup> Bordeaux University, Inserm, Bordeaux Population Health Research Center, team: EPICENE, UMR1219, campus de Carreire, 146, rue Léo-Saignat, 33000 Bordeaux, France

<sup>b</sup> Respiratory Diseases Department, pôle cardiothoracique, CHU de Bordeaux, Bordeaux, France

<sup>c</sup> Cabinet médical, avenue Thiers, Bordeaux, France

<sup>d</sup> Cabinet médical Pessac, Pessac, France

<sup>e</sup> CHG de Côte-Basque, Bayonne, France

<sup>f</sup> Medical Center TOKI EDER, Cambo-les-bains, France

<sup>g</sup> Polyclinique Saint-Augustin, Bordeaux, France

<sup>h</sup> Cabinet médical, La Rochelle, France

<sup>i</sup> Université Bordeaux, Inserm, Bordeaux Population Health Research Center, team Pharmaco-epidemiology, UMR 1219, Bordeaux, France

### ARTICLE INFO

#### Article history:

Received 12 June 2020

Accepted 13 November 2020

Available online 22 November 2020

### ABSTRACT

**Introduction and objective.** – Dyspnoea is a major symptom in COPD patients, but the determinants that could be associated with a higher dyspnoea mMRC score in COPD patients remain unclear. Our research aimed to study the determinants of dyspnoea at the threshold of 1, 2, 3 and 4 mMRC.

**Patients and methods.** – Diagnosis of COPD was made using spirometry with post-bronchodilator  $\frac{FEV_1}{FVC} < 70\%$ . An online questionnaire has been employed by pulmonologists to recruit COPD patients. The following variables were collected: age, gender, BMI, FEV<sub>1</sub>, RV, IC, TLC, FRC, mMRC, frequency of exacerbations and comorbidities. The LASSO was used to select the variables associated with the mMRC dyspnoea scale in a subgroup (who had no missing IC, RV and FRC values) of 421 COPD patients defined by the previously mentioned variables.

**Results.** – One thousand nine hundred and seventy-three patients (65.3% males, average age = 66 ± 10, 38% current smokers) were included. Dyspnoea was correlated with a low FEV<sub>1</sub> and with the number of exacerbations in the past 12 months. Multivariate analysis showed that the determinants of dyspnoea (mMRC ≥ 2) are: FEV<sub>1</sub>: OR = 3.71 [2.86–4.82]; anxiety: OR = 2.52 [1.82–3.47]; cough: OR = 1.94 [1.57–2.40]; bronchiectasis: OR = 1.84 [1.03–3.29]; age: OR = 1.80 [1.45–2.24]; hyperinflation (RV/TLC): OR = 1.68 [1.34–2.11]; ischemic cardiopathy: OR = 1.63 [1.22–2.18]; hypertension: OR = 1.52 [1.21–1.91]; exacerbations (≥ 2): OR = 1.41 [1.10–1.81]; women: OR = 1.39 [1.10–1.74] and overweight: OR = 1.33 [1.06–1.67]. The subgroup analysis showed that: FEV<sub>1</sub>: OR = 3.47 [1.96–6.12]; exacerbations (≥ 2) OR = 2.31 [1.33–4.17] and hyperinflation (IC/TLC) OR = 0.57 [0.35–0.85] were associated with higher dyspnoea (mMRC ≥ 2).

**Conclusion.** – Our results showed that dyspnoea is related to the severity of airflow limitation, gender, exacerbations, comorbidities and hyperinflation.

**Abbreviations:** COPD, chronic obstructive pulmonary disease; mMRC, modified medical research council; FEV<sub>1</sub>, forced expiratory volume in one second; FVC, forced vital capacity; BMI, Body Mass index; RV, residual volume; IC, inspiratory capacity; TLC, total lung capacity; FRC, functional residual capacity; LASSO, least absolute shrinkage and selection operator.

\* Corresponding author.

E-mail address: el-hassane.ouaalaya@u-bordeaux.fr (E.H. Ouaalaya).

<https://doi.org/10.1016/j.resmer.2020.100803>

2590-0412/© 2020 SPLF and Elsevier Masson SAS. All rights reserved.

## 1. Introduction

Dyspnoea is the most dominant and defining symptom of COPD [1]. In many cases dyspnoea is linked to a worsening of COPD and an increased mortality risk [2]. Therefore, dyspnoea is of public health importance and needs more attention, especially since the degree of dyspnoea is an important and valid factor in the assessment of COPD severity. It is difficult to design methods capable of capturing the various dimensions of the complex pathophysiological and psychosocial etiology of dyspnoea. Moreover, there exists a scarcity of consensus on the use of the present tools for capturing the confused dyspnoea history, in clinical investigation context [3]. Furthermore, despite therapeutic management (use of long-acting muscarinic antagonists, long-acting beta-2 agonists, short-acting beta-2 agonists and inhaled corticosteroids), patients with advanced COPD usually suffer from dyspnoea. For better patient management, clinicians must evaluate patient's dyspnoea history and interactions on a regular basis. While there are many instruments available for evaluating dyspnoea, no classification system or instrument is deemed optimal for COPD patients.

According to GOLD recommendations, the mMRC scale was used by clinicians to assess dyspnoea. This scale measures breathing difficulty in two classes (dyspnoea mMRC 0-1 and dyspnoea mMRC  $\geq 2$ ). A basic inquiry regarding the existence or absence of higher dyspnoea will certainly not serve to drive therapy, as it would lack clarity on the advanced state of the disease and hence inadequate management of the disease [4]. Existing knowledge gaps such as medical dyspnoea norm, cognitive disability evaluation and determination of the quality of dyspnoea treatment in patients warrant a sensitivity analysis using different mMRC scale cut-off values to define higher dyspnoea [5]. Owing to the total absence of sensitivity analyses in this framework, the evaluation of the determinants of dyspnoea becomes very complicated. Hence, in our cohort, we successfully performed a sensitivity analysis for higher dyspnoea with different cut-off values: mMRC scale  $\geq 1$ ; mMRC scale  $\geq 2$ ; mMRC scale  $\geq 3$  and mMRC scale  $\geq 4$ .

Our research aimed to assess the determinants that might be associated with dyspnoea in COPD patients with a focus in the relationship in between mMRC scale and hyperinflation and propose a sensitivity analysis for higher dyspnoea.

## 2. Patients and methods

### 2.1. Study design and population

The Palomb project: a cohort of COPD patients enrolled in pulmonary clinics in south west of France in a real-life setting as an ongoing, prospective, multicenter and observational study.

The research duration began on January 2014 and ended on January 2018. 1,973 patients were enrolled in the study by pulmonologists: The data was obtained using a web-site questionnaire fulfilled by the pulmonologist on a secure platform and with specific agreement obtained for the storage of health data. The inclusion criterion was a diagnosis of COPD on the basis of a lung function test according to the ATS/ERS recommendations [6] and made using spirometry with post-bronchodilator (forced expiratory volume in one second/forced vital capacity < 70%) FEV1/FVC < 70%. Patients were excluded if they didn't have the lung function criteria of COPD.

A pulmonary function test (PFT) was performed for COPD patients by pulmonologists at a clinic visit; the pulmonologists/study coordinators enter patient information, lung function values and patient's mMRC (modified medical research council) scale at the same clinic visit. The first reported mMRC dyspnoea grade at inclusion was used to measure dyspnoea.

### 2.2. Measurements

The online questionnaire (additional file 1) included the subsequent sections: demographic and tobacco smoke information's, clinical symptoms, lung volumes, comorbidities and treatments.

#### 2.2.1. Dyspnoea scale

We used a simple, 5-point grading system (Table 1) called the mMRC scale, that tests breathing problems through everyday tasks [7]. The final value determines the extent of disability caused by dyspnoea. A pulmonologists coaching was provided to help patients to select a score on the mMRC scale. We used four cut-off values for a higher/lower dyspnoea group (Table 1).

#### 2.2.2. COPD severity and static hyperinflation

The 2017 GOLD [2] classification had changed, taking into account the GOLD grade [8] (I: FEV<sub>1</sub> > 80%; II: 80%  $\geq$  FEV<sub>1</sub>  $\geq$  50%; III: 50% > FEV<sub>1</sub>  $\geq$  30% and IV: 30% > FEV<sub>1</sub>), the mMRC dyspnoea grade (<2 or  $\geq 2$ ) and the number of exacerbations (<2 or  $\geq 2$ ) per year risk (refined ABCD tool) without comorbidities.

Recently, the ratio of residual volume to total lung capacity (RV/TLC) has been described as a distinct factor of poor prognosis in COPD patients [9,10]. In our study, RV, IC (inspiratory capacity), FRC (functional residual capacity) and TLC were measured. The RV/TLC ratio was used to set three cut-off values for static hyperinflation: patients without hyperinflation (RV/TLC < 130%); patients with moderate hyperinflation (130%  $\leq$  RV/TLC < 220%); and patients with severe hyperinflation (220%  $\leq$  RV/TLC). A variety of measures of hyperinflation have been used, but they all do not have the same meaning. [11] So, we decided to investigate the effects of other measures of hyperinflation [12] (FRC/TLC and IC/TLC ratio) by mMRC scale in the subgroups (who had no missing IC, RV and FRC values) of 421 patients defined by the previously mentioned variables.

#### 2.2.3. Comorbidities

Twenty-three comorbidities were systematically recorded by the pulmonologist (see annex). Comorbidity diagnosis was first confirmed by patient care report, and then by the analysis of the patient's medication history (data not shown). Patients with presence of cardiovascular comorbidities are patients who had at least one of the following comorbidities: hypertension, ischemic cardiopathy, arrhythmia, left heart failure, atheroma, stroke, pulmonary hypertension or lower extremity artery disease.

#### 2.2.4. Medications used for COPD

At the inclusion clinical visit, use of long-acting muscarinic antagonists (LAMAs), long-acting beta-2 agonists (LABAs), short-acting beta-2 agonists (SABAs) and inhaled corticosteroids (ICS) was collected.

### 2.3. Statistical analysis

Categorical variables are presented as proportions and continuous variables are presented as mean and SD. We performed a univariate logistic regression analysis between mMRC dyspnoea scale and FEV<sub>1</sub>, number of exacerbations, gender, static hyperinflation and comorbidities. Statistically significant differences between groups were tested by Wilcoxon test, chi-square test, or fisher's exact tests, as appropriate. We performed an in-depth cross-analysis of different factors, to describe the similarities/differences existing between dyspnoea groups. Relationships between mMRC dyspnoea grade and clinical variables were examined by both Pearson and spearman correlation coefficients. A logistic regression multivariate model with a backward elimination procedure

**Table 1**  
The dyspnoea mMRC scale utilized in GOLD guidelines.

GOLD classification	mMRC wording	Analysis groups: different cut off for a higher/lower dyspnoea group			
Less symptoms	0: I only get breathless with strenuous exercise	Lower dyspnoea (mMRC < 1)	Lower dyspnoea (mMRC < 2)	Lower dyspnoea (mMRC < 3)	Lower dyspnoea (mMRC < 4)
	1: I get short of breath when hurrying on level ground or walking up a slight hill	Higher dyspnoea (mMRC ≥ 1)			
More symptoms	2: on level ground, I walk slower than people of the same age because of breathlessness, or have to stop for breath when walking at my own pace		Higher dyspnoea (mMRC ≥ 2)		
	3: I stop for breath after walking about 100 yards or after a few minutes on level ground			Higher dyspnoea (mMRC ≥ 3)	
	4: I am too breathless to leave the house or I am breathless when dressing				Higher dyspnoea (mMRC ≥ 4)

was computed to identify factors associated with higher dyspnoea group. This model included the studied comorbidities, age, gender, hyperinflation, cough, sputum, the number of exacerbations and other parameters. Independent variables included in the model were also those showing a  $P < 0.05$  in bivariate analysis. We also performed a sensitivity analysis for higher dyspnoea with different cut-off values: mMRC scale  $\geq 1$ ; mMRC scale  $\geq 2$ ; mMRC scale  $\geq 3$  and mMRC scale  $\geq 4$ . The least absolute shrinkage and selection operator (LASSO [13]: a machine learning regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces) was used to select the variables associated with the mMRC dyspnoea scale in a subgroup of patients who had no missing RV, IC and FRC values. All analyses were performed with R version 3.3.2 and SAS software version 9.4.

### 3. Results

#### 3.1. Characteristics of the patients

One thousand nine hundred and seventy-three patients with COPD were included in the PALOMB cohort and their characteristics are summarized in Table 2. They were mostly men (65.33%), with a mean age (SD) of 66.11 (10.98) years, mean post-bronchodilator FEV<sub>1</sub> was 58.11% (19.5%). They were symptomatic COPD patients with mild to very severe airflow obstruction associated with frequent comorbidities. 59.35% of patients reported one or more exacerbations in the previous 12 months. 33.63% women had more than two exacerbations compared to 24.52% of men ( $P < 0.0001$ ). One thousand one hundred and fifty-eight patients were considered to have higher dyspnoea (mMRC grade  $\geq 2$ ) and 59.14% of patients suffer from the static hyperinflation (RV/TLC  $\geq 120\%$ ).

Stratification of the population according to the dyspnoea scale showed that was more frequent in women and overweight patients, and also correlated with the number of exacerbations. Dyspnoea was higher in patients with presence of cardiovascular comorbidities (60.45% Vs 39.55%), and more precisely in hypertensive patients (41.28%) and patients with ischemic cardiopathy (20.81%). Dyspnoea was also frequent in patients with presence of anxiety (21.85%), with 10 out of the 23 comorbidities being related to dyspnoea (Table 3).

#### 3.2. Relationship between mMRC grade and morphometric, clinical and functional data

According to Fig. 1, the GOLD grade (I, II, III and IV) was related to mMRC scale ( $P < 0.0001$ ). However, the correlation (Fig. 2) between mMRC and FEV<sub>1</sub> (% predicted) was low ( $r = 0.36$ ;  $P < 0.0001$ ). A

spearman non-parametric analysis (Fig. 2) found no correlations between mMRC grade and considered comorbidities. The correlation between depression and anxiety was low ( $r = 0.37$ ;  $P < 0.0001$ ). The presence of a chronic cough and sputum maintained a high correlation ( $r = 0.65$ ;  $P < 0.0001$ ). Furthermore, no significant correlation was found between mMRC grade and age, gender or BMI.

Univariate analysis was used to evaluate relationships between mMRC scale and hyperinflation. RV/TLC ratio was significantly different between mMRC scale 0–4 ( $P < 0.0001$ ). Dyspnoea was more frequent in patients with presence of hyperinflation (68.65%) than in patients with absence of hyperinflation (31.35%). more precisely, dyspnoea was more frequent (67.76%) in patients with presence of mild hyperinflation (Table S1).

#### 3.3. Independent factors associated with the mMRC dyspnoea scale

Multiple logistic regression analysis indicated that the presence of comorbidities such as Dyslipidemia, depression, asthma and left heart failure was not significantly different between the lower and the higher mMRC dyspnoea groups (0–1 Vs.  $\geq 2$ ).

Regardless of the cut-off values set to distinguish higher dyspnoea clusters, the severity of FEV<sub>1</sub> and anxiety remained linked with dyspnoea (Table 4). This data is also presented as a multiple forest plot diagram (Fig. S1) to make it easier to understand and interpret. The increase in the mMRC threshold (1; 2; 3 and 4) implies a substantial rise in the odds ratio of the severity of FEV<sub>1</sub> ( $\leq 50\%$ ): 2.93[1.92–4.49] vs. 3.71[2.86–4.82]; and vs. 4.30[3.30–5.60]; and vs. 4.66[2.96–7.31]. The results do not change substantially with added adjustment for factors such as: age, gender and FEV<sub>1</sub> (Table S2).

The information on medications used to treat COPD is important as they affect dyspnoea and could affect relationships among patient characteristics and mMRC scores. In our cohort, patients with mMRC  $\geq 2$ , compared with those with mild dyspnoea (mMRC = 1) or those with absence of dyspnoea (mMRC = 0) were more intensively treated with COPD inhaled treatments (Table 5).

#### 3.4. Subgroup analysis

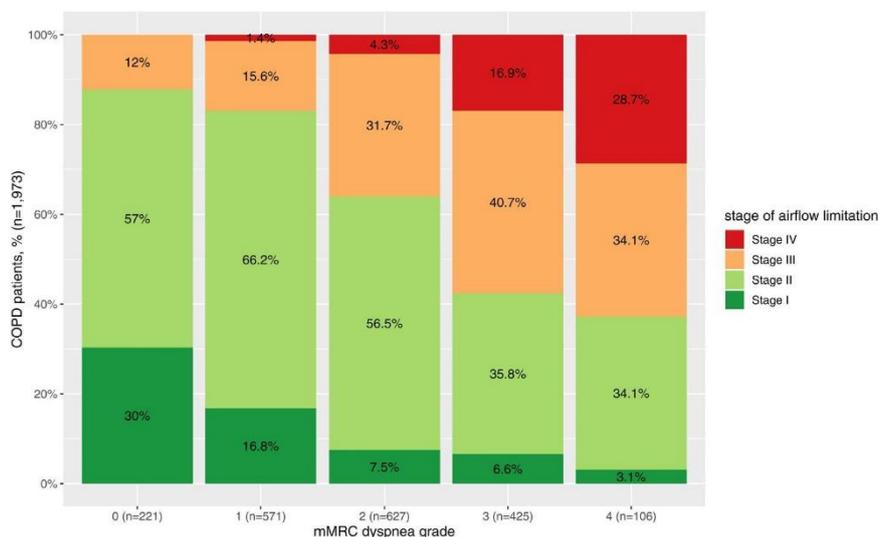
Box plot of the relationships between RV, IC, FRC, RV/TLC, IC/TLC and FRC/TLC ratio (% predicted), as indicators of static hyperinflation, and mMRC dyspnoea scale, as measure of dyspnoea, were assessed (Fig. 3). Increasing values of RV/TLC and FRC/TLC ratio (and decreasing value of IC/TLC ratio) were observed across increasing values of mMRC dyspnoea scale from 0 to 4. The LASSO was used to select the variables associated with the mMRC dyspnoea scale in a subgroup ( $n = 421$ ) of COPD patients who had no missing RV, IC

**Table 2**  
Summary of the characteristics of COPD patients (1,973) according to Dyspnoea grade mMRC.

Subjects characteristics, n (%)		Dyspnoea mMRC "0" n = 221	Dyspnoea mMRC "1" n = 571	Dyspnoea mMRC "2" n = 627	Dyspnoea mMRC "3" n = 425	Dyspnoea mMRC "4" n = 106	P value
Variable names	Data in all patients						
Males, n (%)	1289 (65.33)	166 (75.11)	369 (64.62)	401 (63.96)	265 (62.35)	71 (66.98)	0.0193
Age, yrs. mean (SD)	66.11 (10.98)	62.9 (10.20)	63.3 (10.40)	66.2 (10.80)	69.4 (10.61)	73.1 (10.71)	<0.0001
BMI, (kg/m <sup>2</sup> ), mean (SD)	26.69 (10.80)	26.05 (5.08)	26.3 (5.48)	27.2 (17.11)	26.55 (6.27)	27.59 (7.57)	0.4502
<21 (underweight), n (%)	331 (16.78)	32 (14.47)	93 (16.31)	107 (17.07)	76 (17.83)	20 (18.87)	0.3322
[21–26](normal), n (%)	683 (34.62)	86 (38.90)	196 (34.37)	216 (34.45)	147 (34.61)	28 (26.42)	
[26–29](overweight), n (%)	396 (20.07)	49 (22.17)	129 (22.63)	117 (18.66)	77 (18.12)	19 (17.92)	
≥ 29 (obese), n (%)	563 (28.54)	54 (24.43)	153 (26.80)	187 (29.82)	125 (29.41)	39 (36.79)	
FEV <sub>1</sub> , % pred, Mean (SD)	58.11 (19.50)	71.7 (16.70)	65.5 (15.90)	56.11 (17.71)	48.2 (18.70)	40.5 (20.11)	0.4504
Chronic cough, n (%)	1,124 (56.97)	97 (43.90)	273 (47.80)	383 (61.08)	288 (67.73)	74 (69.81)	<0.0001
Chronic sputum, n (%)	795 (40.30)	71 (32.13)	174 (30.47)	270 (43.06)	216 (50.82)	57 (53.77)	<0.0001
Current smokers, n (%)	793 (40.2)	107 (48.42)	238 (41.68)	253 (40.35)	159 (37.41)	30 (28.30)	0.0064
0–1 exacerbation, n (%)	1,427 (72.33)	183 (82.80)	465 (81.40)	464 (74)	243 (57.12)	56 (52.80)	<0.0001
≥ 2 exacerbations, n (%)	546 (27.67)	38 (17.20)	106 (18.61)	163 (26)	182 (42.91)	50 (47.22)	<0.0001
Pulmonary rehabilitation, n (%)	157 (8)	4 (1.80)	18 (3.15)	50 (7.93)	62 (14.63)	23 (21.71)	<0.0001
Smoking cessation, n (%)	155 (7.81)	15 (6.82)	52 (9.10)	44 (7.02)	37 (8.74)	6 (5.63)	0.5087
Hyperinflation (RV/TLC ≥ 120%), n (%)	1,167 (59.14)	79 (35.75)	277 (48.51)	402 (64.11)	313 (73.65)	80 (75.47)	<0.0001

**Table 3**  
Comorbidities associated with COPD according to dyspnoea grade (mMRC).

Comorbidities, n (%)	Dyspnoea mMRC "0" n = 221	Dyspnoea mMRC "1" n = 571	Dyspnoea mMRC "2" n = 627	Dyspnoea mMRC "3" n = 425	Dyspnoea mMRC "4" n = 106	P value
Hypertension, 734 (37.21)	75 (33.92)	167 (29.25)	248 (39.56)	171 (40.23)	59 (55.69)	<0.0001
Dyslipidemia, 446 (22.61)	47 (21.26)	106 (18.56)	123 (19.61)	128 (30.11)	34 (32.07)	<0.0001
Ischemic cardiopathy, 353 (17.97)	30 (13.57)	76 (13.34)	100 (15.94)	105 (24.78)	36 (33.96)	<0.0001
Apnea Syndrome, 340 (17.23)	59 (26.69)	112 (19.61)	80 (12.75)	67 (15.76)	19 (17.92)	<0.0001
Anxiety, 318 (16.13)	10 (4.52)	55 (9.63)	92 (14.67)	114 (26.82)	47 (44.33)	<0.0001
Depression, 282 (14.34)	22 (9.95)	58 (10.15)	78 (12.44)	89 (20.94)	33 (31.13)	<0.0001
Diabetes, 250 (12.67)	32 (14.47)	51 (8.93)	76 (12.12)	69 (16.23)	19 (17.92)	0.0037
Arrhythmia, 241 (12.24)	20 (9.04)	52 (9.13)	81 (12.91)	61 (14.35)	25 (23.58)	0.0002
Left heart failure, 100 (5.07)	3 (1.35)	17 (2.97)	26 (4.14)	32 (7.52)	20 (18.86)	<0.0001
Bronchiectasis, 92 (4.66)	6 (2.71)	12 (2.12)	42 (6.69)	27 (6.35)	5 (4.71)	0.0009
Cardiovascular comorbidities, 1087 (55.17)	109 (49.32)	262 (45.88)	353 (56.37)	269 (63.29)	78 (73.58)	<0.0001



**Fig. 1.** mMRC dyspnoea grade distribution by stage of airflow limitation.

and FRC values (A descriptive analysis was done to show the differences in between this subsample and the rest of the sample; Table S3). Thirty-two variables were subjected to lasso regression, and only 7 variables were selected: age, cough, the number

of exacerbations, BMI, FEV<sub>1</sub>, IC/TLC ratio and depression. Multiple logistic regression analysis was employed to examine the contribution of the variables selected by lasso to mMRC scale: FEV<sub>1</sub>: OR = 3.47[1.96–6.12]; exacerbations (≥ 2): OR = 2.31[1.33–4.17]

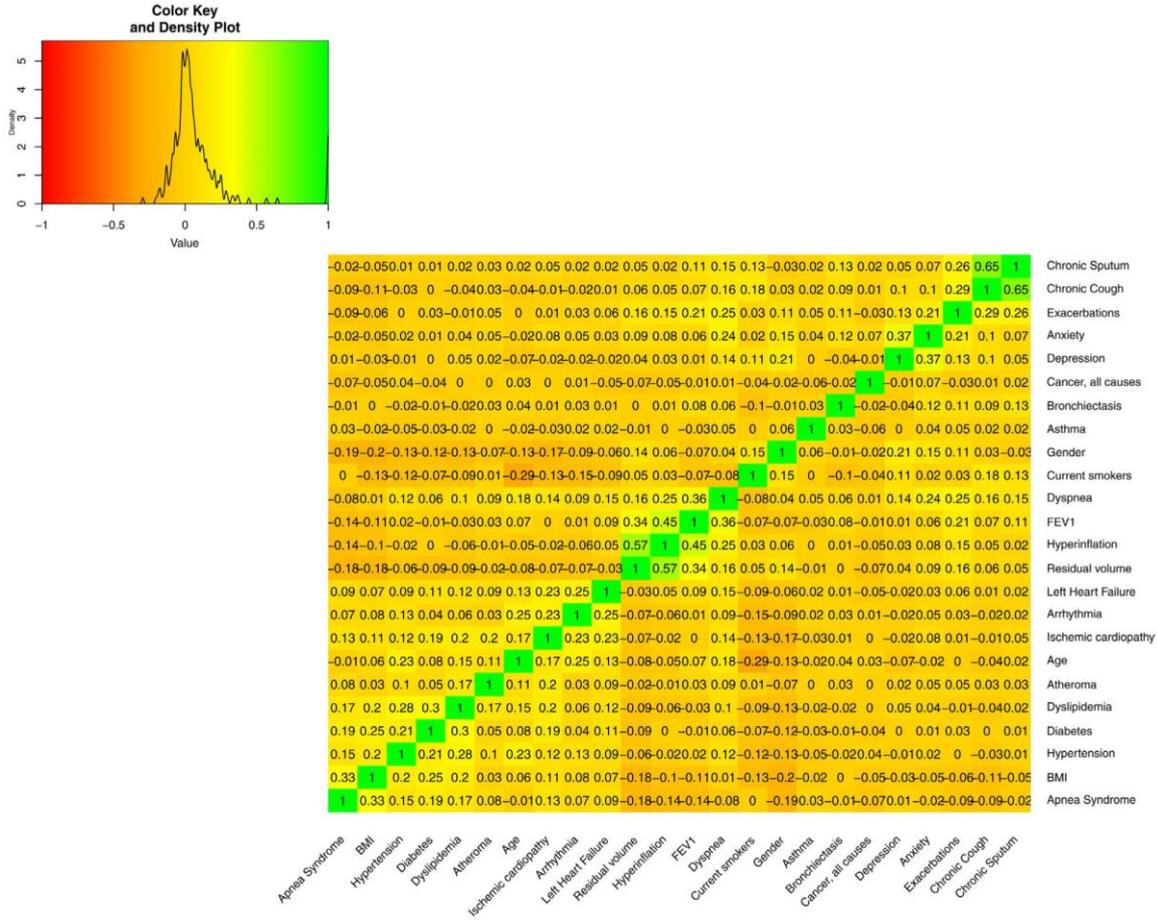


Fig. 2. Spearman correlations heat map for variables in dyspnoea model in the PALOMB cohort. Colours in each cell indicate the strength of the relationship.

Table 4

Multivariate analysis (in 1,973 COPD patients) indicating variables associated with higher dyspnoea, considering mMRC more than 1, 2, 3 and 4.

	OR [95% CI]			
	Higher dyspnoea (mMRC grade ≥ 1)	Higher dyspnoea (mMRC grade ≥ 2)	Higher dyspnoea (mMRC grade ≥ 3)	Higher dyspnoea (mMRC grade ≥ 4)
FEV <sub>1</sub> (% pred.)	2.93[1.92–4.49]	3.71[2.86–4.82]	4.30[3.30–5.60]	4.66[2.96–7.31]
Anxiety	4.09[2.12–7.89]	2.52[1.82–3.47]	2.47[1.82–3.37]	3.65[2.26–5.89]
Age	1.63[1.22–2.17]	1.80[1.45–2.24]	2.27[1.78–2.88]	2.85[1.75–4.64]
Cough	1.71[1.29–2.27]	1.94[1.57–2.40]	1.60[1.25–2.05]	
Hyperinflation	1.71[1.26–2.33]	1.68[1.34–2.11]	1.34[1.01–1.76]	
Gender (Women vs. Men)	1.76[1.26–2.46]	1.39[1.10–1.74]	1.30[1.00–1.68]	
BMI	1.76[1.28–2.41]	1.33[1.06–1.67]		
Ischemic cardiopathy		1.63[1.22–2.18]	1.90[1.41–2.54]	
Number of exacerbations during the past year		1.41[1.10–1.81]	1.84[1.43–2.36]	
Depression			1.67[1.19–2.34]	2.01[1.19–3.40]
Hypertension		1.52[1.21–1.91]		1.62[1.04–2.50]
Bronchiectasis		1.84[1.03–3.29]		
Dyslipidemia			1.81[1.38–2.38]	
Asthma			1.44[1.00–2.08]	
Left heart failure				3.35[1.84–6.11]

and hyperinflation (IC/TLC): OR=0.57[0.35–0.85] were associated with higher dyspnoea (mMRC ≥ 2) in this subgroup of patients.

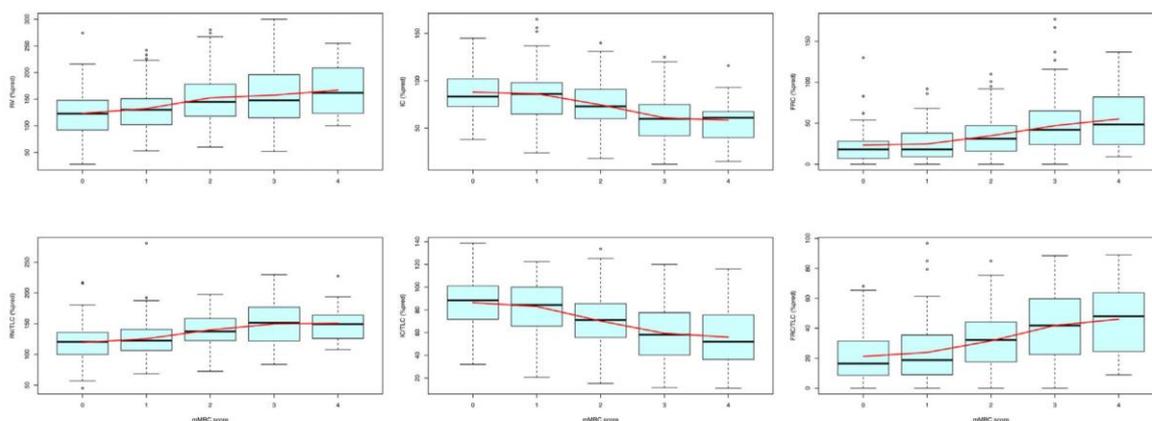
4. Discussion

The most interesting result in the current study of COPD patients is the specific relationship between the static

hyperinflation (defined by the IC/TLC ratio) and the mMRC dyspnoea scale. With a comparable degree of COPD severity the majority of hyperinflated patients had higher dyspnoea than those with lower hyperinflation. Dyspnoea is described as a physical perception with different degrees of severity of uneasy breathing [14] and a widely used parameter for assessment of COPD severity [15]. GOLD has recognized a need for a more rigorous evaluation of

**Table 5**  
COPD inhaled treatments according to dyspnoea grade (mMRC).

Medication, n (%)	Dyspnoea mMRC "0" n=221	Dyspnoea mMRC "1" n=571	Dyspnoea mMRC "2" n=627	Dyspnoea mMRC "3" n=425	Dyspnoea mMRC "4" n=106	P value
SABA, 749 (37.96)	41 (18.55)	185 (32.47)	243 (38.76)	214 (50.35)	60 (56.65)	<0.0001
LABA, 628 (31.83)	46 (20.81)	185 (32.46)	208 (33.17)	150 (35.29)	34 (32.08)	0.0042
LAMA, 936 (47.44)	53 (23.98)	234 (40.98)	327 (52.15)	255 (60)	57 (53.77)	<0.0001
LABA or LAMA, 1252 (63.46)	84 (38.01)	339 (59.37)	436 (69.54)	311 (73.18)	68 (64.15)	<0.0001
LABA and LAMA, 312 (15.81)	15 (6.79)	80 (14.01)	99 (15.79)	94 (22.12)	23 (21.74)	<0.0001
Inhaled corticosteroids, 101 (5.12)	2 (0.91)	15 (2.63)	30 (4.78)	40 (9.41)	13 (12.26)	<0.0001
Fixed combination (ICS + LABA), 532 (26.96)	28 (12.67)	110 (19.26)	184 (29.35)	161 (37.88)	42 (39.62)	<0.0001



**Fig. 3.** Box plot for RV, IC, FRC, RV/TLC, IC/TLC and FRC/TLC ratio by dyspnoea mMRC scale in the subgroup of 421 COPD patients.

symptoms, rather than a quick and easy breathlessness calculation [16]. However, many studies have highlighted the importance of dyspnoea in COPD patients such as the Spanish COPD study [17]. There is considerable disagreement about the appropriate methodology for measuring dyspnoea. We opted for the mMRC scale since it is a unidimensional quantitative measurement of severity of dyspnoea, and is used by pulmonologists worldwide.

A recent study showed that the number of comorbidities can predict dyspnoea [18]. Moreover, mMRC scores were significantly higher in patients with psychiatric co-morbidity, particularly anxiety and depression [19]. Dyspnoea is a potential predictor of depression in COPD [17,20]. This association however, has not been reflected in a few studies [21] possibly due to a catastrophic misinterpretation of the sensation of dyspnoea. The cause and effect relationship between anxiety and dyspnoea is, however, unclear. There are indications that dyspnoea is major cause of anxiety [22,23] while other studies indicate that anxiety intensify the subjective sensation of dyspnoea [24,25].

In our study 59.38% patients were considered to have higher dyspnoea (mMRC  $\geq 2$ ) and hypertension was the most frequent comorbidity in all dyspnoea grades. The observed association between dyspnoea and cardiovascular comorbidities can be explained partly by shared risk factors such as smoking, age, and inactivity. However, it is also thought that a systemic inflammatory process related to COPD may independently increase the risk of cardiovascular disease.

Despite difficulties in accurate quantification of dyspnoea symptoms due to self-reporting [14], we observed that dyspnoea was more frequent in women - a finding supported by several other studies [26]. We found a significant increase between dyspnoea and the mean age of our study. Previous studies show a consistent association between age and the mMRC dyspnoea score [26], also reflected in middle-aged participants in COPDGene study [27]. In our study, dyspnoea was more frequent in overweight patients

aligning with previous studies linking obesity to dyspnoea [28]. Contrastingly, in one study the level of dyspnoea increased as the BMI categories moved from underweight to overweight [29]. Exacerbations partly characterize a worsening of dyspnoea. Studies showed that the number of exacerbations correlates with mMRC grade and that the "frequent exacerbations" phenotype surpasses dyspnoea on the quality of life [30]. However, these differences are attributable to the COPD patients' perceptions of dyspnoea [31].

The CHAIN cohort established mMRC grade  $\geq 2$  as a good threshold through survival analysis [32]. The cut-off value of the dyspnoea mMRC scale suggested by the GOLD classification is the most commonly used. Lower mMRC scores indicate low dyspnoea. Hence, in our study, post dichotomization of mMRC, a response with mMRC more than 1, 2, 3 and 4 was employed to separate lower Vs. higher dyspnoea [2].

Static hyperinflation was also associated with dyspnoea [33] and may also be a predictor of mortality when expressed as either IC/TLC ratio [33], an increased RV [33]. RV increases owing to an extreme expiratory flow limitation at lung volume. As COPD progresses, the rise in RV is larger than the rise in TLC (higher RV/TLC ratio), resulting in a decrease in VC (vital capacity) and a consequent decline in FEV<sub>1</sub>. However, static hyperinflation cut-off values have never been standardized. Two studies [34,35] have suggested that hyperinflation should be defined as FRC (functional residual capacity) instead of increased RV or RV/TLC ratio [36] but these studies [34,35], use RV to analyse different treatments for emphysema. Many research [37,38], have centred on the usage of IC during clinical practice for assessment of static hyperinflation. Unfortunately, we could not analyse the FRC, the IC or the IC/TLC ratio because of their limitations in clinical practice and their missing values in our data. That is why we opted for a subgroup (n = 421) of COPD patients who had no missing IC and FRC values, to investigate the correlations between mMRC scale and FRC, FRC/TLC or IC/TLC ratio, to evaluate the effects of other

measures of hyperinflation by mMRC scale in the subgroups of 421 patients defined by the previously mentioned variables. FEV<sub>1</sub> OR = 3.47 [1.96–6.12]; exacerbations ( $\geq 2$ ) OR = 2.31 [1.33–4.17] and hyperinflation (IC/TLC) OR = 0.57 [0.35–0.85] were associated with higher dyspnoea (mMRC  $\geq 2$ ) in the subgroup of patients.

In the French “Initiatives BPCO” cohort of COPD patients [39] with a wide range of air-flow limitation, mMRC dyspnoea grades and baseline dyspnea index (BDI) scores were correlated at the group level. These two measurements shared most, but not all, of their determinants, which explained a moderate proportion of their variations. This study [39] suggested that mMRC and BDI scores explore quite differently the dyspnoea intensity in COPD patients and thus should ideally be used in combination. Consistent with our result, Thierry Perez et al. [39] found that FEV<sub>1</sub>, hyperinflation, depression, and comorbidities are all major determinants of mMRC and BDI dyspnea scores.

Our analysis has a range of advantages and disadvantages. Firstly, pulmonologists reported the patient comorbidities, although comorbidities were underdiagnosed in previous research—representing real life. The apnea syndrome was diagnosed by a polysomnography performed by the same pulmonologist who diagnosed COPD. We find a strong association between the comorbidities recorded and their care adherence. Secondly, we cannot generalize the findings in COPD patients to the general population, as our COPD patients have been managed by a general practitioner and a pulmonologist. Thirdly, the GOLD classification would differ according to the symptom scale, CAT (COPD Assessment Test) vs mMRC, but in our cohort, we focused on the mMRC grade by developing more sensitivity analysis. It's well founded that multiple considerations, such as the choice of the variables, the association between the factors chosen and the scientific judgement of the investigators can influence the analysis and therefore the results. To limit the impact of the specific correlation between variables, we first performed an in-depth cross-analysis of different factors; independent variables included in the model were also those showing a  $P < 0.05$  in bivariate analysis. Lastly, most of PALOMB patients are covered by health insurance, meaning that there exists an unrecognized population of uninsured patients who probably have a higher prevalence of frequent dyspnoea.

Despite its limitations, this is one of the few studies that have investigated the association of higher dyspnoea and several determinants in COPD patients.

## 5. Conclusions

Regardless of the cut-off values set to identify higher dyspnoea, our results suggested that dyspnoea is related to the severity of airflow limitation, gender, exacerbations, comorbidities and hyperinflation. Revealing that the hyperinflation defined by an IC/TLC ratio was a better determinant of the mMRC dyspnoea scale in comparison to both RV/TLC and FRC/TLC ratios. These cohort results illustrate in real-life the multifactorial aspect of dyspnoea in COPD patients. The development of a predictive score by assigning a weight to each determinant category would provide interesting information. Risk factors associated with higher dyspnoea should be used to develop and externally validate a new higher dyspnoea risk score in other cohorts. Future studies are needed for validation of this prediction rule and its value in clinical practice.

**Authorship statement**All authors contributed to the acquisition of the data. CR and E-HO had full access to all of the data in the study and takes responsibility for the integrity of the data and accuracy of the data analysis. CR, E-HO, AB, CN, LF, FLG, LN, JMD, MM and AO contributed to the conception and design of the study. All authors read and approved the final manuscript.

## Funding sources

Funding (unrestricted grants): Supported by Bordeaux university foundation, Novartis Pharma, Isis Medical, Glaxo SmithKline, Boehringer Ingelheim and Chiesi. The funding sources had no role in the design or conduct of the study, in the collection, management, analysis and interpretation of the data, or in the preparation, review or approval of the manuscript.

## Disclosure of interest

Mr. Ouaalaya, Dr. Falque, Dr. Dupis, Dr. Sabatini, Dr. Bernady, Dr. Nguyen, Dr. Ozier, Dr. Nocent-Eijnani declare that they have no competing interest. Dr. Le Guillou reports personal fees from GSK, personal fees from Menarini, personal fees from Pfizer, personal fees from Philips, personal fees from Astra Zeneca, personal fees from Boehringer Ingelheim, personal fees from Chiesi, other from Alize santé, other from SOS oxygène, other from Bastide Médical, other from Boehringer Ingelheim, outside the submitted work; Dr. Molimard report personal fees from University of Bordeaux during the conduct study, other from Novartis Pharma, Boehringer Ingelheim, GSK outside the submitted work; Dr. Raheison reports grants from Bordeaux University Foundation, during the conduct of the study; personal fees from Astra Zeneca, personal fees from ALK, personal fees from Boehringer Ingelheim, personal fees from Glaxo SmithKline, personal fees from MundiPharma, personal fees from Novartis, personal fees from Chiesi, outside the submitted work.

## Inclusion centres

E. Blanchard, H. Jungmann, M. Sapène, J. Casteigt, J. Moinard, M. Staali, S. Morgado, J. Truchon, E. Chabry, E. Iglesias, C. Roy, Y. Daoudi, A. Prud'homme, E. Monge.

## Availability of data and materials

The data that support the findings of this study are available from INSERM U1219 Epicene Team but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of INSERM U1219 Epicene Team.

## Ethics approval and consent to participate

The CNIL (National Data Protection and Privacy Commission) and the CCTIRS (Advisory Committee for Data Processing in Health Research) approved the study, and informed consent was taken before enrolment. The authors had asked the local ethics committee for feedback regarding the need for ethical clearance for such a retrospective analysis, and were advised that this was not warranted.

## Consent for publication

Consent for publication statement is not applicable as no personal information is provided in this manuscript.

## Acknowledgments

The authors dedicate this manuscript to the memory and the contribution of their dear co-author and friend François Pellet, MD, for his outstanding contribution to the Palomb project. They also thank E. Berteaud for data management; C. Bousquet and A. Le-

leon from Bordeaux University foundation for their support. All members of the study team are greatly acknowledged.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.resmer.2020.100803>.

#### References

- Joshi M, Joshi A, Bartter, Thaddeus. Symptom burden in chronic obstructive pulmonary disease and cancer. *Curr Opin Pulm Med* 2012;18(2):97–103, <http://dx.doi.org/10.1097/MCP.0b013e32834fa84c>.
- From the global strategy for the diagnosis management prevention of COPD. Global Initiative for Chronic Obstructive Lung Disease (GOLD); 2017 [Internet, cited 2017 May 1, <http://www.goldcopd.org>].
- Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax* 1999;54(7):581–6.
- Richard A, Mularski, Margaret L, Campbell, Steven M, Asch, et al. A review of quality of care evaluation for the palliation of dyspnea. *Am J Respir Crit Care Med* 2010;181:534–8, <http://dx.doi.org/10.1164/rccm.200903-0462PP>.
- Moreira GL, Donaria L, Furlanetto KC, Paes T, Sant'Anna T, Hernandes NA, et al. GOLD B–C–D groups or GOLD II–III–IV grades: which one better reflects the functionality of patients with chronic obstructive pulmonary disease? *Chron Respir Dis* 2015;12(2):102110, <http://dx.doi.org/10.1177/1479972315573528>.
- Celli BR, MacNee W, committee members. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J* 2004;23:932–46, <http://dx.doi.org/10.1183/09031936.04.00014304>.
- Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. *Chest* 1988;93:580–6, <http://dx.doi.org/10.1378/chest.93.3.580>.
- Sonia Coton, William M, Vollmer, Bateman E, Guy B, Marks, et al. Severity of airflow obstruction in chronic obstructive pulmonary disease (COPD): proposal for a new classification. *COPD* 2017;14(5):469–75, <http://dx.doi.org/10.1080/15412555.2017.1339681>.
- Tae Rim S, Yeon-Mok O, Joo Hun P, Keu Sung L, Sunghee O, Dae Ryoung K, et al. The prognostic value of residual volume/total lung capacity in patients with chronic obstructive pulmonary disease. *J Korean Med Sci* 2015;30(10):1459–65, <http://dx.doi.org/10.3346/jkms.2015.30.10.1459>.
- Budweiser S, Harlacher M, Pfeifer M, Jörres RA. Co-morbidities and hyperinflation are independent risk factors of all-cause mortality in very severe COPD. *COPD* 2014;11(4):388–400, <http://dx.doi.org/10.3109/15412555.2013.836174>.
- Smith BM, Hoffman EA, Basner RC, et al. Not all measures of hyperinflation are created equal. Lung structure and clinical correlates of gas trapping and hyper-expansion in COPD: the multi-ethnic study of atherosclerosis (MESA) COPD study. *Chest* 2014;145(6):1305–15, <http://dx.doi.org/10.1378/chest.13-1884>.
- Pierantonio L, Albuquerque A, Aliverti A, Babb T, Barreiro E, Dres M, et al. ERS statement on respiratory muscle testing at rest and during exercise. *Eur Respir J* 2019;53(6):1801214, <http://dx.doi.org/10.1183/13993003.01214-2018>.
- Goto T, Jo T, Matsui H, Fushimi K, Hayashi H, Yasunaga H. Machine learning-based prediction models for 30-day readmission after hospitalization for chronic obstructive pulmonary disease. *COPD* 2019;16(5–6):338–43, <http://dx.doi.org/10.1080/15412555.2019.1688278>.
- Parshall MB, Schwartzstein RM, Adams L, Robert B, et al. An official American thoracic society statement: update on the mechanisms, assessment, and management of dyspnea. *Am J Respir Crit Care Med* 2012;185(4):435–52, <http://dx.doi.org/10.1164/rccm.201111-2042ST>.
- Lavolette L, Laveneziana P. Dyspnoea: a multidimensional and multidisciplinary approach. *Eur Respir J* 2014;43:1750–62, <http://dx.doi.org/10.1183/09031936.00092613>.
- Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2013;187(4):347–65, <http://dx.doi.org/10.1164/rccm.201204-0596PP>.
- Almagro P, Yun S, Sangil A, Rodríguez-Carballeira M, Meritxell M, Landete P, et al. Palliative care and prognosis in COPD: a systematic review with a validation cohort. *Int J Chron Obstruct Pulmon Dis* 2017;12:1721–9, <http://dx.doi.org/10.2147/COPD.S135657>.
- Putcha N, Puhana MA, Drummond BM, Han MK, Regan EA, Hanania NA, et al. A simplified score to quantify comorbidity in COPD. *PLoS One* 2014;9(12):e114438, <http://dx.doi.org/10.1371/journal.pone.0114438>.
- Ruchi D, Anindya D, Abhishek K, Sandeep K, Mayank M, Kapil S. Association of comorbid anxiety and depression with chronic obstructive pulmonary disease. *Lung India* 2018;35(1):31–6, <http://dx.doi.org/10.4103/lungindia.lungindia.537.16>.
- Martinez Rivera C, Costan Galicia J, Alcazar Navarrete B, Garcia-Polo C, Ruiz Iturriaga LA, Herrejon A, et al. Factors associated with depression in COPD: a multicenter study. *Lung* 2016;194(3):335–43, <http://dx.doi.org/10.1007/s00408-016-9862-7>.
- Sanchez O, Caumont-Prim A, Gillet-Juvin K, Callens É, Graba S, Essalhi M, et al. Activity-related dyspnea is not modified by psychological status in people with COPD, interstitial lung disease or obesity. *Respir Physiol Neurobiol* 2012;182(1):18–25, <http://dx.doi.org/10.1016/j.resp.2012.02.004>.
- Xu W, Collet J-P, Shapiro S, Lin Y, Yang T, Robert W, et al. Independent effect of depression and anxiety on chronic obstructive pulmonary disease exacerbations and hospitalizations. *Am J Respir Crit Care Med* 2008;178(9):913–20, <http://dx.doi.org/10.1164/rccm.200804-6190C>.
- van Manen JG, Bindels PJ, Dekker FW, Ijzermans CJ, van der Zee JS, Schade E. Risk of depression in patients with chronic obstructive pulmonary disease and its determinants. *Thorax* 2002;57:412–6.
- Tiller J, Pain M, Biddle N. Anxiety disorder and perception of inspiratory resistive loads. *Chest* 1987;91:547–51.
- Martinez-Moragon E, Perpina M, Belloch A, de Diego A, Martinez-Frances M. Determinants of dyspnea in patients with different grades of stable asthma. *J Asthma* 2003;40:375–82.
- Rune G, Vollmer WM, Hardie JA, Inga Sif O, et al. Predictors of dyspnoea prevalence: results from the BOLD study. *Eur Respir J* 2014;43(6):1610–20, <http://dx.doi.org/10.1183/09031936.00036813>.
- Martinez CH, Diaz AA, Parulekar AD, Rennard SI, Kanner RE, Hansel NN, et al. Age-related differences in health-related quality of life in COPD, an analysis of the COPDGen and SPIROMICS Cohorts. *Chest* 2016;149(4):927–35, <http://dx.doi.org/10.1016/j.chest.2015.11.025>.
- Lambert AA, Putcha N, Drummond MB, Boriek AM, Hanania NA, Kim V, et al. Obesity Is Associated with Increased Morbidity in Moderate to Severe COPD. *Chest* 2017;151(1):68–77, <http://dx.doi.org/10.1016/j.chest.2016.08.1432>.
- Jeong Uk L, Jae Ha L, Ju Sang K, Yong Il H, Tae-Hyung K, Seong Yong L, et al. Comparison of World Health Organization and Asia-Pacific body mass index classifications in COPD patients. *Int J Chron Obstruct Pulmon Dis* 2017;12:2465–75, <http://dx.doi.org/10.2147/COPD.S141295>.
- Mascaró O, Serra J, Lucchetti G, Gimeno X, Ruiz F. COPD phenotypes and comorbidities. *Eur Respir J* 2017;50(suppl 61):PA3610, <http://dx.doi.org/10.1183/1393003.congress-2017.PA3610>.
- Scioscia G, Blanco I, Arismendi E, Burgos F, Concepción G, María Pia Foschino B, et al. Different dyspnoea perception in COPD patients with frequent and infrequent exacerbations. *Thorax* 2017;72(2):117–21, <http://dx.doi.org/10.1136/thoraxjnl-2016-208332>.
- Ciro C, Marin JM, Martínez-González C, de Lucas-Ramos P, Mir-Viladrich I, Cosío B, et al. Differential effect of modified medical research council dyspnea, COPD assessment test, and clinical COPD questionnaire for symptoms evaluation with the new GOLD staging and mortality in COPD. *Chest* 2015;148(1):159–68, <http://dx.doi.org/10.1378/chest.14-2449>.
- Ciro C, Cote C, de Torres JP, Aguirre-Jaime A, Marin JM, Pinto-Plata V, et al. Inspiratory-to-total lung capacity ratio predicts mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2005;171:591–7, <http://dx.doi.org/10.1164/rccm.200407-867OC>.
- Hartman JE, ten Hacken NHT, Klooster K, Marike Boezen H, de Greef MHG, Slebos D-J. The minimal important difference for residual volume in patients with severe emphysema. *Eur Respir J* 2012;40(5):1137–41, <http://dx.doi.org/10.1183/09031936.00219111>.
- Fiorelli A, D'Andrilli A, Bezzi M, Ibrahim M, Anile M, Diso D, et al. Complications related to endoscopic lung volume reduction for emphysema with endobronchial valves: results of a multicenter study. *J Thorac Dis* 2018;10(27):S3315–25, <http://dx.doi.org/10.21037/jtd.2018.06.69>.
- Rossi A, Aisanov ZR, Avdeev S, Di Maria G, Donner CF, Izquierdo JL, et al. Mechanisms, assessment and therapeutic implications of lung hyperinflation in COPD. *Respir Med* 2015;109(7):785–802, <http://dx.doi.org/10.1016/j.rmed.2015.03.010>.
- Duranti R, Filippelli M, Bianchi R, Romagnoli I, Pellegrino R, Brusasco V, et al. Inspiratory capacity and decrease in lung hyperinflation with albuterol in COPD. *Chest* 2002;122(6):2009–14, <http://dx.doi.org/10.1378/chest.122.6.2009>.
- Maria A, Ramon, Jaume Ferrer, Elena Gimeno-Santos, David Donaire-Gonzalez, Esther Rodriguez, et al. Inspiratory capacity-to-total lung capacity ratio and dyspnoea predict exercise capacity decline in COPD. *Respirology* 2016;21(3):476–82, <http://dx.doi.org/10.1111/resp.12723>.
- Perez T, Borge PR, Paillassier JL, Caillaud D, Deslée G, Chanez P, et al. Modified Medical Research Council scale vs Baseline Dyspnea Index to evaluate dyspnea in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2015;10:1663–72, <http://dx.doi.org/10.2147/COPD.S82408>.

## IV. Les phénotypes et les déterminants d'exacerbations identifiés par une classification non-supervisée (n=932)

Ce travail a fait l'objet d'une publication sous forme d'article original dans la revue scientifique *Respiratory Medicine*, d'une communication orale au congrès international du CPLF (24<sup>ème</sup> édition du congrès de pneumologie de langue française) en janvier 2020 à Paris et d'une communication affichée (Poster) au congrès international de l'ERS (European Respiratory Society Congress) en septembre 2019 à Madrid (Espagne). L'article est résumé dans l'encadré suivant, et est présenté dans son intégralité dans les pages suivantes. Le supplément de cet article est disponible en Annexe.

La BPCO est une affection fréquente, responsable d'une morbi-mortalité importante, due à la sévérité de l'obstruction bronchique et aux comorbidités cardiovasculaires mais aussi à la fréquence des exacerbations. L'objectif de notre travail était de déterminer les facteurs associés au phénotype exacerbateur fréquent avec un suivi de 3 ans.

Les patients BPCO ont été inclus depuis le janvier 2014 par leur pneumologue par web questionnaire (cohorte PALOMB). Nous avons analysé la fréquence et les associations d'exacerbation chez les patients BPCO suivis sur une période de 3 ans. Les données ont été analysées à l'aide de la méthode Kml conçue pour regrouper les données longitudinales et l'analyse ROC (receiver operating characteristic) conçue pour évaluer l'exactitude des prévisions d'un modèle afin de déterminer le meilleur seuil pour attribuer les patients aux clusters identifiés. Des analyses univariées et multivariées ont été réalisées pour étudier les caractéristiques associées aux différents phénotypes.

Au total, 932 patients ayant une BPCO (VEMS/CVF < 70 % post-BD) ont été inclus ; 446 patients (65,68 % d'hommes, 35,59 % de fumeurs actuels) ont été suivis durant 3 ans sans aucune donnée manquante. 239 (28,15 %) patients (avec un âge moyen de 65,83 ; 44,62 % de femmes) avaient deux exacerbations ou plus au cours de l'année précédant leur inclusion (exacerbations fréquentes). Parmi eux 142 (16,68 %) ont également eu des exacerbations fréquentes au cours de la première année de l'étude et 69 (8,10 %) qui sont restés des exacerbateurs fréquents au cours de la deuxième année. Sur la base de notre hypothèse, nous avons pu déterminer quatre phénotypes : le phénotype A (Exacerbateurs non fréquents), B (exacerbations fréquentes chez les patients en sous poids), C (transitoires) et D (exacerbations fréquentes chez les patients obèses). Après ajustement sur les variables de confusion, l'analyse multivariée a montré que les patients qui avaient des exacerbations fréquentes l'année précédant leur inclusion (HR [IC95 %] : 3,72 [2,53–5,49]), une anxiété (HR [IC95 %] : 2,03 [1,24–3,31]) et l'absence de vaccination antigrippale annuelle (HR [IC95 %] : 1,97[1,20–3,24]) restent associées aux phénotypes d'exacerbateurs fréquents.

Ces résultats montrent que les déterminants les plus importants du phénotype exacerbateur fréquent sont les antécédents d'exacerbations, l'absence de vaccination antigrippale annuelle et l'anxiété. Ces analyses confirment l'existence et la pertinence clinique d'un phénotype d'exacerbateurs fréquents de patients BPCO et le seuil actuellement utilisé pour définir ce phénotype.

- **Article :**

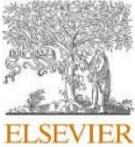
El Hassane Ouaalaya, Laurent Falque, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Ejnaini, Frédéric Le Guillou, Mathieu Molimard, Maéva Zysman, Chantal Raheison-Semjen. Susceptibility to frequent exacerbation in COPD patients: Impact of the exacerbations history, vaccinations and comorbidities? *Respir Med.* 2020 May 11; 169: 106018. DOI: <https://doi.org/10.1016/j.rmed.2020.106018>

- **Communication orale :**

E. H. Ouaalaya, A. Bernady, J. Casteigt, J. M. Dupis, C. Nocent-Eijnaini, M. Sabatini, L. Falque, F. Le Guillou, L. Nguyen, A. Ozier, M. Molimard, C. Raheison-Semjen. Susceptibilité aux exacerbations fréquentes chez les patients BPCO : impact des vaccinations et des comorbidités ? *Revue des Maladies Respiratoires Actualités.* Volume 12, numéro 1, janvier 2020, page 24. DOI: <https://doi.org/10.1016/j.rmra.2019.11.036> [CPLF - Paris 2020 (24e édition du Congrès de Pneumologie de Langue Française) : sélection via résumé]

- **Communication affichée (Poster) :**

El-hassane Ouaalaya, Laurent Falque, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Ejnaini, Frédéric Le Guillou, Mathieu Molimard, Chantal Raheison-Semjen. Characterization of the “frequent exacerbator phenotype” over a period of 3 years: Palomb cohort. *European Respiratory Journal* (November 21, 2019) vol. 54 no. suppl 63; DOI: <https://doi.org/10.1183/13993003.congress-2019.PA4432> ERS International Congress -Madrid- Spain.



Contents lists available at ScienceDirect

## Respiratory Medicine

journal homepage: <http://www.elsevier.com/locate/rmed>

## Susceptibility to frequent exacerbation in COPD patients: Impact of the exacerbations history, vaccinations and comorbidities?

El Hassane Ouaalaya<sup>a,\*</sup>, Laurent Falque<sup>c</sup>, Jean Michel Dupis<sup>d</sup>, Marielle Sabatini<sup>e</sup>,  
Alain Bernady<sup>f</sup>, Laurent Nguyen<sup>g</sup>, Annaig Ozier<sup>g</sup>, Cécilia Nocent-Ejnaini<sup>e</sup>, Frédéric Le Guillou<sup>h</sup>,  
Mathieu Molimard<sup>i</sup>, Maéva Zysman<sup>j</sup>, Chantal Raheison-Semjen<sup>a,b</sup>

<sup>a</sup> Bordeaux University, INSERM, Bordeaux Population Health Research Center, team: EPICENE, UMR1219, Bordeaux, France<sup>b</sup> CHU de Bordeaux, Pole Cardiothoracique, Respiratory Diseases Department, Bordeaux, France<sup>c</sup> Polyclinique Bordeaux Rive Droite, Lormont, France<sup>d</sup> Cabinet Medical Pessac, Pessac, France<sup>e</sup> CHG Côte Basque, Bayonne, France<sup>f</sup> Medical Center TOKI EDER, Cambo les bains, France<sup>g</sup> Polyclinique Saint-Augustin, Bordeaux, France<sup>h</sup> Cabinet Medical, la Rochelle, France<sup>i</sup> Bordeaux University, INSERM, Bordeaux Population Health Research Center, team Pharmaco-epidemiology, UMR 1219, Bordeaux, France<sup>j</sup> Bordeaux University, INSERM, Cardiothoracic Research Center of Bordeaux, team: 2, UMR1045, Bordeaux, France

## ARTICLE INFO

## Keywords:

COPD Phenotypes  
Frequent exacerbator  
Cluster analysis  
Longitudinal analysis

## ABSTRACT

**Introduction:** Exacerbations are key events in the natural history of COPD, but our understanding of their longitudinal determinants remains unclear. We used data from a large observational study to test the hypothesis that vaccination status and comorbidities could be associated with the occurrence of exacerbations profile.

**Methods:** Diagnosed COPD patients have been included by their pulmonologists, with up to 3 years of follow-up. Data were analyzed using the Kml method designed to cluster longitudinal data and receiver operating characteristic curve analysis to determine the best threshold to allocate patients to identified clusters.

**Results:** 932 COPD patients were included since January 2014, 446 patients (65.68% males, 35.59% current smokers) were followed over a period of 3 years with complete data. 239(28.15%) patients reported two or more exacerbations in the year before enrolment (frequent exacerbations). Among them 142(16.68%) also had frequent exacerbations in the first year of the study, and 69(8.10%) who remained frequent exacerbators in the second year. Based on our hypothesis, we were able to determine four phenotypes: A (infrequent), B (frequent in underweight patients), C (transient), and D (frequent in obese patients). Frequent exacerbators had more airflow limitation and symptoms. Irrespective of cut-offs set to define the optimal number of clusters, a history of exacerbations OR: 3.72[2.53–5.49], presence of anxiety OR: 2.03[1.24–3.31] and absence of the annual influenza vaccination OR: 1.97[1.20–3.24] remained associated with the frequent exacerbator phenotypes.

**Conclusions:** The most important determinants of frequent exacerbations are a history of exacerbations, anxiety and unvaccinated against influenza.

## 1. Introduction

Chronic obstructive pulmonary disease (COPD) is a chronic inflammatory disease of the airways characterized by an airflow limitation that is not fully reversible [1] and remains a highly prevalent [2] disease. Several phenotypes have been described with a focus on frequent

exacerbator profile. Exacerbations become increasingly frequent and severe as COPD progresses [3], impacting lung function, morbidity, and mortality [4–7]. The causes of COPD exacerbations are not well known, and they can occur across all stages of airflow limitation.

Several models predicting exacerbations in COPD patients have been published reporting combinations of information from patient history, clinical characteristics and lung function. Though, it is still unclear

\* Corresponding author. El Hassane OUAALAYA Bordeaux University - ISPED Case 11, Team "Epidemiology of cancer and environmental exposures" (EPICENE), Inserm U1219- Bordeaux Population Health Research Center (BPH), 146 Rue Léo Saignat, Bordeaux Cedex, 33076, France.

E-mail address: [el-hassane.ouaalaya@u-bordeaux.fr](mailto:el-hassane.ouaalaya@u-bordeaux.fr) (E.H. Ouaalaya).

<https://doi.org/10.1016/j.rmed.2020.106018>

Received 30 January 2020; Received in revised form 1 May 2020; Accepted 8 May 2020

Available online 11 May 2020

0954-6111/© 2020 Elsevier Ltd. All rights reserved.

Abbreviation			
COPD	Chronic Obstructive Pulmonary Disease	BMI	Body Mass Index
CNIL	Commission nationale de l'informatique et des libertés National (Data Protection and Privacy Commission)	GOLD	Global Initiative for Chronic Obstructive Lung Disease
CCTIRS	Advisory Committee for Data Processing in Health Research	mMRC	modified Medical Research Council
FEV1	Forced Expiratory Volume in 1 s	CAT	COPD Assessment Test
FVC	Forced vital capacity	ATS	the American Thoracic Society
RV	Residual Volume	ERS	the European Respiratory Society
TLC	Total Lung Capacity	PFT	pulmonary function test
EELV	The end-expiratory lung volume	LABA	long-acting $\beta$ -agonist
VC	Vital Capacity	LAMA	long-acting muscarinic antagonist
IC	Inspiratory Capacity	ICS	inhaled corticosteroid
FRC	Functional residual Capacity	SABA	short-acting $\beta$ -agonist
		RNIPP	Répertoire national d'Identification des personnes physiques (National identification register of private individuals)
		ROC	receiver operating characteristic

which prediction model predicts exacerbations most accurately and applicable in daily practice.

Previous work on risk factors of COPD exacerbations confirms that a recent history of exacerbations is the strongest risk predictor of future exacerbations. However, this theory has not been adequately investigated because our current understanding of COPD exacerbations and their relationship to disease severity is based on large intervention studies or multiple smaller studies that have used varying definitions of exacerbation.

The ECLIPSE investigators suggested that individuals with two or more exacerbations in a given year represent a distinct frequent exacerbator phenotype [3]. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) guide [1] to COPD diagnosis and management uses a threshold of two or more acute exacerbations in the previous year, or at least one hospital admission related to an acute exacerbation, to identify individuals at high risk of future events (groups C and D). However, intra-subject variability of annual exacerbation frequency limits the ability of these criteria to predict the future risk of exacerbations at an individual level [3,8].

The French cohort (Initiatives-BPCO) identified four phenotypes from which phenotype 1 was characterized by 4 exacerbations per year [9]. The SPIROMICS study reported a markedly lower proportion of stable frequent exacerbators over 3 years (2.00%) compared to the ECLIPSE cohort (12.00%) and the French cohort (EXACO) (7.00%) [2,3,10]. Some studies suggested a third intermediate phenotype called infrequent exacerbators, which includes patients with one exacerbation per year [3,8,11–13].

Most COPD patients with frequent exacerbations have comorbidities. A recent study demonstrates a salient influence of the cardiovascular comorbidities on the appearance of COPD-related respiratory exacerbations [14].

Psychiatric comorbidities such as anxiety and depression have a negative impact on exacerbation frequency [15]. A Chinese prospective cohort study report that the total number of exacerbations were higher in COPD patients with bronchiectasis [16]. The retrospective observational Korean study (KNHANES) showed that the metabolic syndrome increased the frequency of exacerbations in women with early COPD [17]. In addition, some comorbidities may mimic an exacerbation of COPD (e.g. ischemic heart disease) [15].

No definitive evidence exists about the impact of prevention of COPD exacerbations in reducing mortality, and current pharmacological treatments decrease exacerbations but do not abolish them [18–20]. Inhaled treatment aims to reduce exacerbation frequency in COPD but its impact is usually not taken into account in observational studies. Another aspect of management is vaccination status, recommended in clinical practice to reduce the number of exacerbations, however the impact of vaccinations on exacerbations rates in a real-life study is

unknown: Two systematic Cochrane reviews showed a reduction in exacerbations in COPD patients for annual influenza and pneumococcal vaccinations versus control [21,22]. The authors of a recent systematic review reported the acceptable safety profile of influenza vaccines in COPD patients [23]. Also Influenza vaccination has been shown to reduce the incidence of COPD exacerbations [24]. Descriptive population-based and cohort studies have shown that influenza vaccination significantly reduces hospitalizations and mortality in COPD patients [25–27]. Recently, Exacerbation rates were associated with influenza vaccination in the multicenter retrospective Turkish cohort [28].

We used data from a large observational study to test the hypothesis that vaccination status and comorbidities could be associated with the occurrence of exacerbations profile, taking into account inhaled treatment.

## 2. Methods

### 2.1. Study design and population

The Palomb project is an ongoing, prospective, multicenter, observational cohort of COPD patients recruited in pulmonary clinics in South-West of France and followed up for 3 years. The study period started from 1<sup>st</sup> January 2014 (cohort entry date) and finished at 1<sup>st</sup> January 2018 (death, left practice or end of follow-up).

932 patients were enrolled and followed up by pulmonologists. Data was obtained using a web-site questionnaire filled by the pulmonologists on a secure platform and with specific agreement obtained for the storage of health data. The inclusion criterion was a diagnosis of COPD: The diagnosis was made by a pulmonologist based on symptoms, smoking history and confirmed by lung function, according to the ATS/ERS standards [1]; spirometry with post-bronchodilator FEV<sub>1</sub>/FVC < 70.00%. A pulmonary function test (PFT) was performed for COPD patients by pulmonologists at a clinic visit.

The CNIL (National Data Protection and Privacy Commission) and the CCTIRS (Advisory Committee for Data Processing in Health Research) approved the study, and informed verbal consent was obtained before enrollment. The authors had asked the local ethics committee for feedback regarding the need for ethical clearance for such a retrospective analysis and were advised that this was not warranted.

### 2.2. Definition of outcomes

The online questionnaire included many fields such as demographic criteria, smoking habits, symptoms, severity of airflow obstruction [29], BMI, medications, and comorbidities.

The case definition of an exacerbation was based on the decision by

the pulmonologist and defined by an increase of cough and/or phlegm and/or dyspnea lead to antibiotics and/or systemic corticosteroids. The exacerbations data in the year before enrolment was collected at the baseline visit and in the follow up visits. The frequency of exacerbations was classified as none, one, and two or more exacerbations.

Static hyperinflation was defined by a baseline RV/TLC >120.00% of the predicted value [30–32]. Twenty-three comorbidities were systematically recorded by the pulmonologist (see annex). Comorbidity diagnosis was confirmed by the patient's report, and reviewing their medication list (Data not shown). At the annual clinical visit, use of long-acting muscarinic antagonists (LAMAs), long-acting beta-2 agonists (LABAs), short-acting beta-2 agonists (SABAs) and inhaled corticosteroids (ICS) was collected, along with pneumococcal and influenza vaccinations.

On the 1<sup>st</sup> January 2018, vital status was obtained for each participant through the national database that annually records the death of all French citizens, living inside or outside the French Territory (National Identification Register of Private Individuals, Répertoire National d'Identification des Personnes Physiques (RNIPP)) [33]. The information obtained by the RNIPP database was compared to the data recorded in the civil registration and less than 1% of discrepancies were noted. All dates of death were obtained for participants who died during the follow-up. Authorizations to use these data were obtained in accordance with the French law (CNIL).

### 2.3. Data analysis

The first analysis, done at inclusion, allowed us to describe and characterize frequent exacerbation patients ( $\geq 2$ ). The second, called the extended analysis, was a confirmatory analysis using complete information on annual exacerbations, during each of the 3 years of follow-up. We used KML analysis, which is an implementation of a nonparametric algorithm designed to work specifically on longitudinal data [34]. It identified clusters of patients with similar evolution of exacerbation frequency over the 3 years. Five different indexes were used to choose the optimal number of clusters: The Calinski and Harabatz index (version 1, 2 and 3), the Ray and turi index, and the Davies and Bouldin index [34]. Receiver operating characteristic (ROC) curve analysis was performed to identify the best threshold to separate clusters.

Statistically significant differences between patients and differences between groups were examined using student's t-test, chi-squared ( $\chi^2$ ), or fisher's exact tests, as appropriate. Independent variables included in the model were those showing a  $P < 0.05$  in bivariate analysis. A logistic regression multivariate model with a backward elimination procedure was computed to identify factors associated with the frequent exacerbations groups. The model included the studied comorbidities and was adjusted for treatments and other parameters. All analyses were performed with SAS software version 9.4 and R version 3.3.2. Data were analyzed using packages KML, version 2.4.1, pROC, version 1.8, and dplyr version 0.5.0 [34,35].

## 3. Results

### 3.1. Characteristics of the patients

932 patients with COPD were included in the PALOMB cohort (Fig. 1) and their characteristics are summarized in Table 1. They were mostly men (64.91%) who were current smokers with symptomatic COPD and mild to very severe airflow obstruction associated with frequent comorbidities. 252 (27.04%) patients reported two or more exacerbations in the previous 12 months. 44.44% women had more than two exacerbations compared to 55.56% of men ( $P = 0.0003$ ). 546 patients were considered to have higher dyspnea (defined by mMRC scale  $\geq 2$ ).

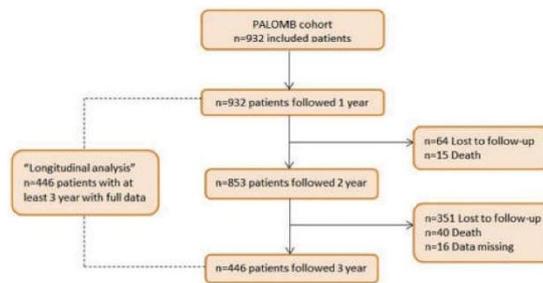


Fig. 1. Flow diagram of participants through each stage of the study.

**Table 1**  
Baseline characteristics of the patients stratified by infrequent (<2/year) or frequent exacerbators ( $\geq 2$ /year).

	All COPD, n = 932	Infrequent exacerbators (<2), n = 680 (72.96)	Frequent exacerbators ( $\geq 2$ ), n = 252 (27.04)	P Value
Males, n (%)	605 (64.91)	465 (68.38)	140 (55.56)	0.0003
Age, yrs. Mean (SD)	66.22 (10.25)	66.33 (10.17)	65.94 (10.47)	0.6105
BMI, (kg/m <sup>2</sup> ), mean (SD)	27.14 (14.52)	26.94 (6.25)	27.69 (26.01)	0.4871
<21 (underweight), n (%)	163 (17.49)	108 (15.88)	55 (21.83)	0.1344
[21-26 [(normal), n (%)	303 (32.51)	220 (32.35)	83 (32.94)	
[26-29 [(overweight), n (%)	170 (18.24)	126 (18.53)	44 (17.46)	
$\geq 29$ (obese), n (%)	296 (31.76)	226 (33.24)	70 (27.78)	
FEV1, % pred, Mean (SD)	57.53 (20.73)	59.24 (19.36)	52.92 (23.46)	0.0002
GOLD stage I, >80%, n (%)	130 (13.95)	104 (15.29)	26 (10.32)	0.0003
GOLD stage II, 50-80%, n (%)	449 (48.18)	341 (50.15)	108 (42.86)	
GOLD stage III, 30-50%, n (%)	282 (30.26)	196 (28.82)	86 (34.13)	
GOLD stage IV, <30%, n (%)	71 (7.62)	39 (5.74)	32 (12.70)	
Chronic cough, n (%)	521 (55.90)	341 (50.15)	180 (71.43)	<0.0001
Chronic sputum, n (%)	398 (42.70)	256 (37.65)	142 (56.35)	<0.0001
Current smokers, n (%)	328 (36.12)	237 (35.85)	91 (36.84)	0.7828
Dyspnea, 0-1 mMRC, n (%)	386 (41.42)	321 (47.21)	65 (25.79)	<0.0001
Dyspnea, $\geq 2$ mMRC, n (%)	546 (58.58)	359 (52.79)	187 (74.21)	
Pulmonary rehabilitation, n (%)	62 (6.65)	23 (3.38)	39 (15.48)	<0.0001
Smoking cessation, n (%)	90 (9.66)	64 (9.41)	26 (10.32)	0.6776
RV, Mean (SD)	147.27 (53.6)	142.02 (52.38)	161.44 (54.37)	<0.0001
TLC, Mean (SD)	105.02 (20.70)	103.34 (20.82)	109.56 (19.70)	<0.0001
RV/TLC, % pred, Mean(SD)	137.84 (33.00)	135.15 (33.07)	145.12 (31.73)	<0.0001
Hyperinflation (RV/TLC $\geq 120\%$ ), n (%)	659 (70.71)	465 (68.38)	194 (76.98)	0.0104

3.2. Factors associated with exacerbations

The frequent exacerbations ( $\geq 2$ ) were higher in men with COPD, women with COPD, GOLD stage III, GOLD stage IV, chronic cough, chronic sputum, current smokers, higher dyspnea, and hyperinflation. FEV<sub>1</sub>, RV, TLC and RV/TLC ratio were related to the frequent exacerbations ( $P < 0.0001$ ).

Considering comorbidities, hypertension and anxiety were higher in symptomatic COPD patients whereas apnea syndrome was more frequent in mild and asymptomatic COPD. The frequent exacerbations were higher in patients with anxiety (23.81%), depression (25.00%), diabetes (15.48%), undernutrition (5.16%) and osteoporosis (3.97%) (See Fig. 2). In our cohort, patients with the frequent exacerbations (two or more exacerbations), compared with those with infrequent exacerbations (zero or one exacerbation) were more intensively treated with COPD medications (See Table 2). Dead patients were older and exhibited more severe airflow limitation. Longitudinal analysis was performed in 446 patients with at least 3 years of follow-up and complete data (See Fig. 1).

Multivariate analysis revealed that the presence of comorbidities such as anxiety, diabetes and depression was significantly different between the groups and strongly related to the frequent exacerbations group (Table 3).

3.3. Stability of the Frequent-Exacerbation Phenotype

Among patients with 3-year data, 3.13% had one exacerbation per year, 28.25% had no exacerbation per year and 3.13% had two or more exacerbations per year (See Fig. 3). Among the 121 patients with two or more exacerbations during the first year of the study, 27.27% and 11.57% persisted in having two or more exacerbations during the second and third years, respectively. Similarly, among the 121 patients with one exacerbation during the first year of the study, 28.92% and 11.57% still had one exacerbation during the second and third following years, respectively. Among the 204 patients with zero exacerbation during the first year of the study, 76.47% and 61.76% still had zero exacerbation during the second and third following years, respectively.

The Kml analysis was used in the sub group of 446 patients. According to all criteria, the optimal number of clusters to separate patients into groups with homogeneous numbers of exacerbations over time was four (See Fig. 4a). The first cluster A consisted of 261 patients (58.52%) who experienced zero or one exacerbation ( $0.30 \pm 0.46$  exacerbations per patient per year), the second cluster B comprised 85 patients (19.05%) who experienced two or more exacerbations ( $2.68 \pm 1.01$  exacerbations per patient per year), the third cluster C consisted of

82 patients (18.38%) who experienced zero, one or two exacerbations ( $1.02 \pm 0.83$  exacerbations per patient per year) and the fourth cluster D consisted of 18 patients (4.03%) who experienced more than two exacerbations ( $3.55 \pm 1.19$  exacerbations per patient per year) over the 3 years of the study (See Fig. 4b).

The ROC curve analysis found good accuracy, with multi-class area under the curve (AUC) of 0.88 (95% CI, 0.82–0.94) and a best threshold to maximize (sensitivity + specificity) of at least two exacerbations in the first year (sensitivity, 91%; specificity, 79%).

3.4. Factor associated with the frequent exacerbators phenotypes

Frequent exacerbators (cluster B and D) had more female patients. BMI proved to be significantly different between all the clusters. Noticeably, cluster B had the highest percentage of underweight ( $< 21 \text{ kg/m}^2$ ) patients while contrastingly obese ( $\geq 29 \text{ kg/m}^2$ ) patients were higher in cluster D. Severe airflow limitation was seen in cluster D patients while cluster C (transient phenotype) encompassed the highest number of GOLD stage III patients. Chronic cough, sputum and dyspnea along with pulmonary rehabilitation were evidently higher in frequent exacerbators. These findings were further supported by higher functional measurement (RV/TLC ration) in these phenotypes (See Table 4). Determinants of frequent exacerbators (phenotype B or D) included higher levels of anxiety, chronic sputum, treatment with LABA, and unvaccinated against influenza (See Table 5).

4. Discussion

We examined the frequency of exacerbations among 932 patients with mild, moderate, severe and very severe COPD. They were mostly men who were current smokers with symptomatic COPD associated with frequent comorbidities. 27.04% of the 932 patients reported two or more exacerbations in the previous 12 months. A range of variables showed an association with exacerbation frequency: GOLD stage III and IV, chronic cough, chronic sputum, current smokers, higher dyspnea and hyperinflation. The Kml method identified four clusters of patients based on longitudinal 3-year follow-up of exacerbation frequency. The multivariate analysis showed the determinants of frequent exacerbators (phenotype B or D): anxiety, chronic sputum, and unvaccinated against influenza.

At the end of our study, only 3.13% of COPD participants had two or more exacerbations compared with the 2.00% in SPIROMICS study, 12.00% in the ECLIPSE cohort and 7.00% in the French cohort (EXACO) [3,8,10]. We considered all stages of COPD severity, as opposed to the severe form in the above mentioned studies which accounts for the

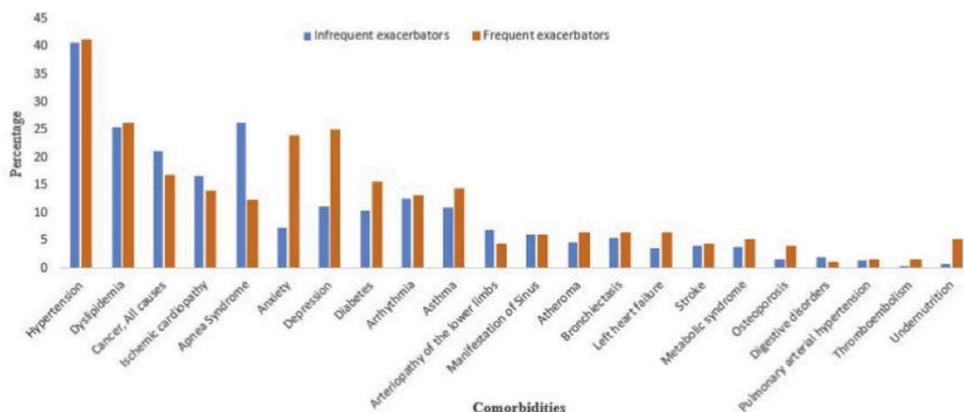


Fig. 2. Prevalence of the comorbidities in the frequent and in-frequent exacerbators.

**Table 2**  
Medications used for treating the 932 COPD patients stratified by infrequent or frequent exacerbators.

COPD Treatment	All COPD, n = 932	Infrequent exacerbators (<2), n = 680 (72.96)	Frequent exacerbators (≥2), n = 252 (27.04)	P Value
SABA, n (%)	369 (39.59)	248 (36.47)	121 (48.02)	0.0014
LABA, n (%)	292 (31.33)	215 (31.62)	77 (30.56)	0.7562
LAMA, n (%)	494 (53.00)	341 (50.15)	153 (60.71)	0.0041
LABA or LAMA, n (%)	621 (66.63)	441 (64.85)	180 (71.43)	0.0586
LABA and LAMA, n (%)	165 (17.70)	115 (16.91)	50 (19.84)	0.2980
Inhaled corticosteroids, n (%)	54 (5.79)	27 (3.97)	27 (10.71)	<0.0001
Fixed combination (ICS + LABA), n (%)	264 (28.33)	174 (25.59)	90 (35.71)	0.0023
LABA and LAMA and ICS, n (%)	29 (3.11)	11 (1.62)	18 (7.14)	<0.0001
Annual influenza vaccination, n (%)	331 (35.52)	197 (28.97)	134 (53.17)	<0.0001
Pneumococcal vaccination, n (%)	406 (43.56)	255 (37.50)	151 (59.92)	<0.0001
Oxygen therapy, n (%)	88 (9.44)	46 (6.76)	42 (16.67)	<0.0001
Monotherapy SABA, n (%)	47 (5.04)	39 (5.74)	8 (3.17)	0.1126
LABA, n (%)	64 (6.87)	54 (7.94)	10 (3.97)	0.0332
LAMA, n (%)	64 (6.87)	47 (6.91)	17 (6.75)	0.9292
Inhaled corticosteroids, n (%)	2 (0.21)	2 (0.29)	0 (0.00)	0.3888
Dual Therapy SABA + LABA, n (%)	49 (5.26)	35 (5.15)	14 (5.56)	0.8040
SABA + LAMA, n (%)	152 (16.31)	100 (14.71)	52 (20.63)	0.0295
SABA + ICS, n (%)	7 (0.75)	3 (0.44)	4 (1.59)	0.0719
LABA + LAMA, n (%)	89 (9.55)	69 (10.15)	20 (7.94)	0.3078
LABA + ICS, n (%)	6 (0.64)	4 (0.59)	2 (0.79)	0.7276
Triple Therapy SABA + LABA + LAMA, n (%)	47 (5.04)	35 (5.15)	12 (4.76)	0.8114
SABA + LABA + ICS, n (%)	5 (0.54)	4 (0.59)	1 (0.40)	0.7224
SABA + LAMA + ICS, n (%)	5 (0.54)	3 (0.44)	2 (0.79)	0.5129
LABA + LAMA + ICS, n (%)	13 (1.39)	5 (0.74)	8 (3.17)	0.0048

**Table 3**  
The determinants of frequent exacerbators (≥2/year) at inclusion in the 932 COPD patients.

	Odds ratio (95% CI)
Pulmonary rehabilitation (Yes vs No)	3.68 [2.03 – 6.67]
Anxiety (Presence vs Absence)	2.12 [1.30 – 3.43]
Diabetes (Presence vs Absence)	1.93 [1.20 – 3.09]
Depression (Presence vs Absence)	1.90 [1.22 – 2.95]
Dyspnea (mMRC 0-1 Vs mMRC ≥2)	1.62 [1.14 – 2.29]
Chronic sputum (Presence vs Absence)	1.60 [1.07 – 2.39]
Chronic cough (Presence vs Absence)	1.59 [1.04 – 2.42]
Residual volume (RV ≤ 140% Vs RV>140%)	1.53 [1.10 – 2.13]

differences in the results.

Between years 1 and 2, 85.00% of patients changed from frequent exacerbators to infrequent exacerbators (0–1 exacerbation), while 30.00% changed from infrequent exacerbators to frequent exacerbators. This intra-subject variability of annual exacerbation frequency, limits the ability of these criteria to predict the future risk of exacerbations, at an individual level. It is possible, that a period longer than 3 years of follow-up might reduce the captured variability of exacerbation rate over time, allowing for a better predictive efficacy by the model.

Most exacerbations are treated in an outpatient setting using SABA, LABA, LAMA, Inhaled and oral corticosteroids with or without antibiotics [1,36]. Pneumococcal and annual influenza vaccinations are also often used to treat COPD exacerbations as influenza is a frequent cause of COPD exacerbations: A recent study showed that influenza viruses are detected in between 2.50% and 11.60% of COPD exacerbations [37], which is consistent with previously published reports documenting influenza detection in 5.00–22.00% of patients hospitalized with exacerbations [38–41]. Immunization against influenza is recommended for all patients with COPD by the World Health Organization (WHO), the US Centers for Disease Control and Prevention, the European Centre for Disease Control and Prevention (ECDC), numerous national agencies [42–44], and the Advisory Committee on Immunization Practices (ACIP) in USA [45]. Current guidelines recommend influenza and pneumococcal vaccination for COPD patients [1].

The originality of our study was to analyze the impact of annual influenza vaccination, pneumococcal vaccination status, and all prescribed COPD treatments on the longitudinal analysis of exacerbations frequency. We were able to show a strong association between unvaccinated against influenza and the frequent exacerbator phenotypes. Despite longstanding recommendations to vaccinate vulnerable high-risk groups against seasonal influenza, vaccination rates remain sub-optimal in this population.

Our study identified two stable frequent exacerbator phenotypes which differed on the basis of BMI. Phenotype D (frequent in obese patients) had a higher association with anxiety. Hence, in COPD patients with frequent exacerbations, adequate management of anxiety should be given due importance. Patient education about influenza vaccination might contribute towards decrease in exacerbations frequency.

Furthermore, studies have demonstrated an impact of pertussis vaccination in reducing frequency of exacerbations and could be an interesting consideration for future studies.

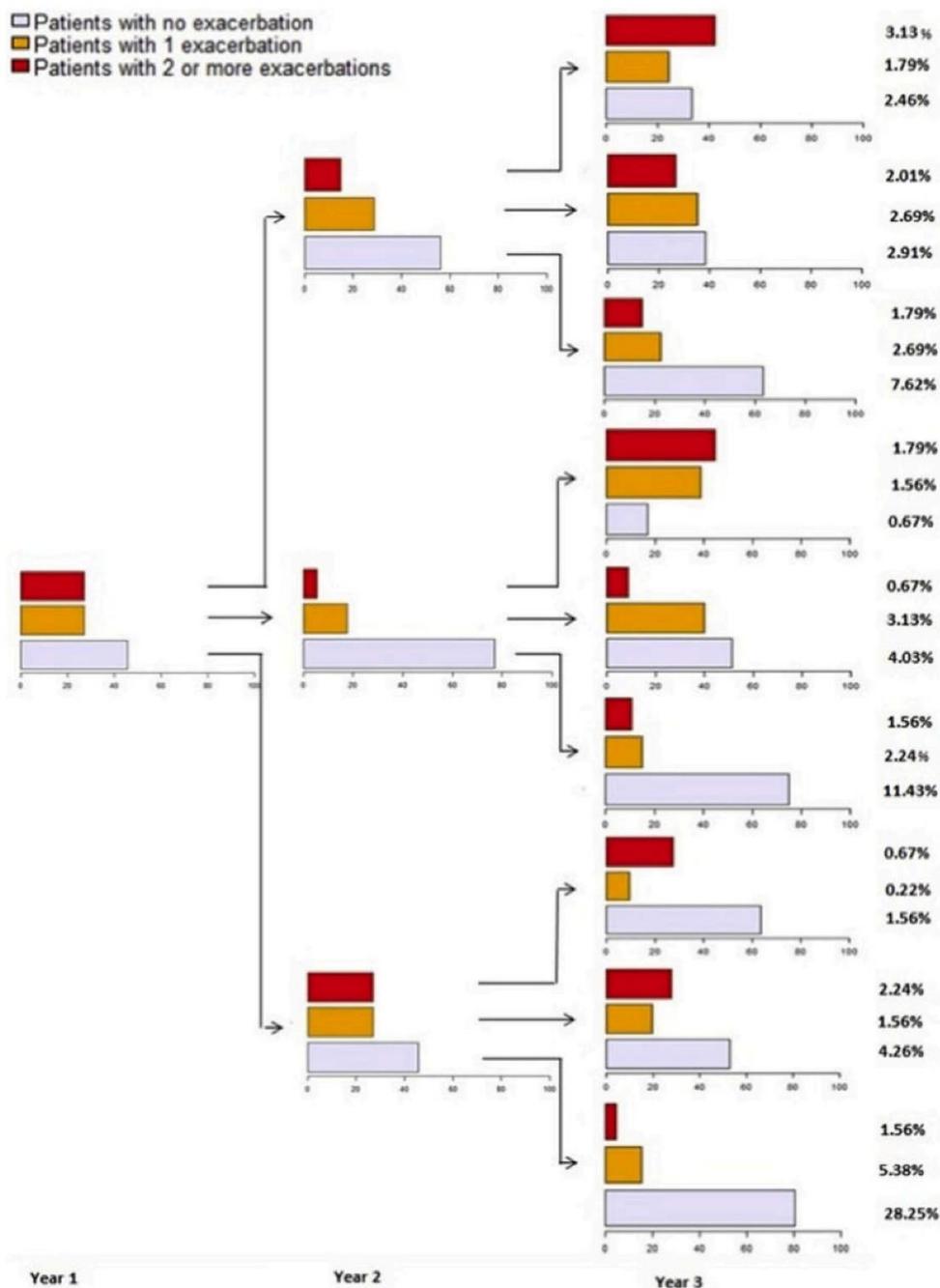
## 5. Strengths and limitations

Our study has several strengths and limitations: 486 patients lost to follow-up were asymptomatic patients. There is a selection bias in our study but it is reflective of the complicated pathology of COPD. These results confirm that patients with COPD managed by pulmonologist are more severe than those managed in primary care. As more severe COPD patients were followed by pulmonologists, we cannot formally exclude selection bias with overestimation of frequent exacerbation phenotype.

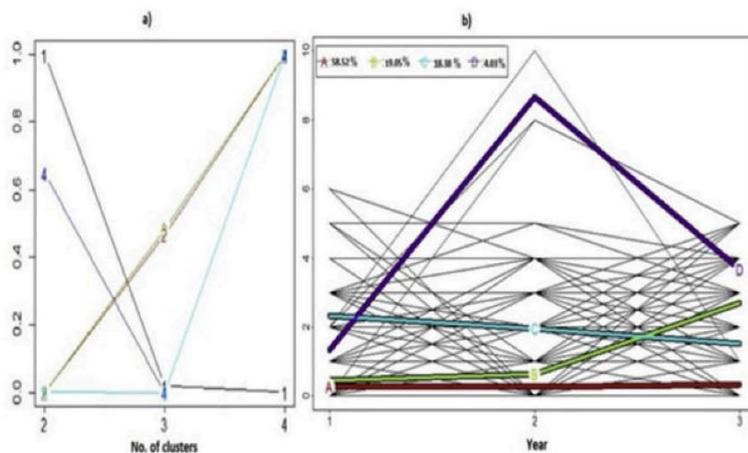
We found four clusters with two frequent exacerbating phenotypes with a difference based on BMI. Adherence to COPD treatment was included in our analysis which has not been previously taken into account.

Comorbidities were recorded by pulmonologists, whereas in previous studies comorbidities remained underdiagnosed – reflective of real life. The recording of comorbidities might be a potential bias. However, this bias is limited, as we found a significant correlation between the reported comorbidities and compliance with their given treatment (data not shown).

Despite its limitations, this is one of the few studies that have investigated the association of frequent exacerbations and risk factors in COPD patients, taking into account therapeutic management: unvaccinated against influenza is an important determinant of frequent exacerbations and comorbidity of anxiety should be adequately treated.



**Fig. 3.** Stability of the Frequent-Exacerbation Phenotype in the 446 COPD patients who completed the study.  
 \*\*The bars at the left show the proportion of patients with no exacerbations, one exacerbation, or two or more exacerbations in year 1  
 \*\*The bars in the middle show the respective incidence of exacerbations for these patients in year 2  
 \*\*The bars at the right show the respective incidence in year 3  
 \*\*The percentages at right denote the proportions of all patients with no exacerbations, one exacerbation, or two or more exacerbations.



**Fig. 4.** Kml clustering analysis of prospectively recorded annual exacerbation frequency. **a):** Criteria applied to choose the optimal number of clusters including the first Calinski and Harabatz index (1, Blood red line), the second version of Calinski and Harabatz index (2, red line), the third version of Calinski and Harabatz index (3, green line), the Ray and Turi index (4, blue line), and the Davies and Bouldin index (5, sky blue line). These criteria are normalized to be mapped together using the Kml package in R. Scores range from 0 to 1 for each tested number of clusters ranging from 2 to 4. The highest score corresponds to the most appropriate number of clusters. **b):** Individual trajectories of annual exacerbation frequency during each year of the study. Red corresponds to cluster A (58.52% of patients), green corresponds to cluster B (19.05% of patients), sky blue corresponds to cluster C (18.38% of patients) and purple corresponds to cluster D (4.03% of patients). Thick lines labeled with the cluster letter represent the mean exacerbation frequency for each cluster. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

**Table 4**  
Baseline characteristics of patients with 3-year follow-up separated into Clusters A, B, C and D.

	COPD patients followed for 3 years, n = 446	A n = 261 (58.50%) Infrequent	B n = 85 (19.05%) Frequent	C n = 82 (18.38%) Transient	D n = 18 (4.03%) Frequent	P value
Males, n (%)	302(67.71)	194(74.33)	47(55.29)	51(62.20)	10(55.56)	0.0035
Age, yrs. Mean (SD)	66.95 (10.04)	66.79 (9.98)	66.05 (10.22)	69.17 (9.86)	63.44 (9.70)	0.0866
BMI, (kg/m <sup>2</sup> ), mean (SD)	27.11(6.29)	27.84 (6.17)	25.94 (6.28)	25.91 (6.60)	27.47 (5.26)	0.0058
<21 (underweight), n (%)	76(17.04)	28(10.73)	23(27.06)	22(26.83)	3(16.67)	0.0033
[21-26 [(normal), n (%)	134(30.04)	77(29.50)	25(29.41)	27(32.93)	5(27.78)	
[26-29 [(overweight), n (%)	85(19.06)	61(23.37)	12(14.12)	10(12.20)	2(11.11)	
≥29 (obese), n (%)	151(33.86)	95(36.40)	25(29.41)	23(28.05)	8(44.44)	
FEV1, % pred, Mean (SD)	58.84 (19.98)	60.97 (19.36)	56.27 (21.28)	56.12 (19.43)	52.32 (22.18)	0.0399
GOLD stage I, >80%, n (%)	69(15.47)	43(16.48)	13(15.29)	10(12.20)	3(16.67)	0.038
GOLD stage II, 50-80%, n (%)	212 (47.53)	137(52.49)	34(40.00)	35(42.68)	6(33.33)	
GOLD stage III, 30-50%, n (%)	136 (30.49)	69(26.44)	30(35.29)	32(39.02)	5(27.78)	
GOLD stage IV, <30%, n (%)	29 (6.50)	12(4.60)	8(9.41)	5(6.10)	4(22.22)	
Chronic cough, n (%)	253(56.73)	133(50.96)	61(71.76)	47(57.32)	12(66.67)	0.007
Chronic sputum, n (%)	201(45.07)	101(38.70)	46(54.12)	44(53.66)	10(55.56)	0.0159
Current smokers, n (%)	153(34.30)	87(33.33)	32(37.65)	27(32.93)	7(38.89)	0.8572
Dyspnea, 0-1 mMRC, n (%)	189(42.38)	129(49.43)	23(27.06)	33(40.24)	4(22.22)	0.0008
Dyspnea, ≥ 2 mMRC, n (%)	257(57.62)	132(50.57)	62(72.94)	49(59.76)	14(77.78)	
Pulmonary rehabilitation, n (%)	32(7.17)	5(1.92)	14(16.47)	9(10.98)	4(22.22)	<0.0001
Smoking cessation, n (%)	48(10.76)	27(10.34)	10(11.76)	10(12.20)	1(5.56)	0.8448
RV, Mean (SD)	143.17 (54.01)	136.51 (53.15)	153.54 (54.21)	150.82 (55.11)	155.94 (50.03)	0.0045
TLC, Mean (SD)	103.98 (21.22)	101.57 (20.65)	107.16 (20.06)	107.82 (24.35)	106.5 (15.33)	0.0135
RV/TLC, % pred, Mean (SD)	135.28 (34.73)	132.16 (37.03)	140.9 (31.09)	137.12 (28.41)	145.58(39.26)	0.0153
Hyperinflation, n (%)	296 (66.37)	160(61.30)	63(74.12)	59(71.95)	14(77.78)	0.058

**Table 5**  
Factors associated with the frequent exacerbator cluster (B) or (D) in 446 patients with full 3-year follow-up.

	Odds ratio (95% CI)
Anxiety (Presence vs Absence)	3.25 [1.67–6.32]
LABA (Presence vs Absence)	2.00 [1.22–3.28]
Annual influenza vaccination (Absence vs presence)	1.97 [1.20–3.24]
Chronic sputum (Presence vs Absence)	1.75 [1.08–2.82]
BMI (underweight, normal vs. overweight, obese)	1.67 [1.03–2.69]

**6. Conclusions**

Our study shows that the most important determinants of frequent exacerbations are: history of exacerbations, anxiety and unvaccinated against influenza. These analyses show the existence and clinical relevance of two frequent exacerbators' phenotypes of patients with COPD and the currently used threshold to define this phenotype.

**Author contributions**

All authors contributed to the acquisition of the data. CR-S and E-HO had full access to all of the data in the study and takes responsibility for the integrity of the data and accuracy of the data analysis. CR-S, AB, CN-E, LF, FLG, LN, JMD, MS, MZ, MM and AO contributed to the conception

and design of the study. All authors read and approved the final manuscript. CR-S is the guarantor. The corresponding author (E-HO) attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

#### Funding (unrestricted grants)

Supported by Bordeaux university foundation, Novartis Pharma, Isis Medical, GlaxoSmithKline, Boehringer Ingelheim and Chiesi. The funding sources had no role in the design or conduct of the study, in the collection, management, analysis and interpretation of the data, or in the preparation, review or approval of the manuscript.

#### Declaration of competing interest

All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare: Mr. Ouaalaya has nothing to disclose; Dr. Falque has nothing to disclose; Dr. Bernady has nothing to disclose; Dr. Nguyen has nothing to disclose; Dr. Ozier has nothing to disclose; Dr. Nocent-Eijnani has nothing to disclose; Dr. Zysman has nothing to disclose; Dr. Le Guillou reports personal fees from GSK, personal fees from Menarini, personal fees from Pfizer, personal fees from Philips, personal fees from Astra Zeneca, personal fees from Boehringer Ingelheim, personal fees from Chiesi, other from Alize santé, other from SOS oxygène, other from Bastide Médical, other from Boehringer Ingelheim, outside the submitted work; Dr. Molimard report personal fees from University of Bordeaux during the conduct study, other from Novartis Pharma, Boehringer Ingelheim, GSK outside the submitted work; Dr. Raherison reports grants from Bordeaux University Foundation, during the conduct of the study; personal fees from Astra Zeneca, personal fees from ALK, personal fees from Boehringer Ingelheim, personal fees from Glaxo SmithKline, personal fees from Mundi-Pharma, personal fees from Novartis, personal fees from Chiesi, outside the submitted work.

#### Acknowledgments

The authors dedicate this manuscript to the memory and the contribution of their dear co-author and friend François Pellet, MD, for his outstanding contribution to the Palomb project. They also thank E. Berteaud for data management; C. Bousquet and A. Le-leon, Barbara Cloute from Bordeaux University foundation for their support. All members of the study team are greatly acknowledged. Inclusion centers: E. Blanchard, H. Jungmann, M. Sapène, J. Casteigt, J. Moinard, M. Staali, S. Morgado, J. Truchon, E. Chabry, E. Iglesias, C. Roy, Y. Daoudi, A. Prud'homme, E. Monge.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rmed.2020.106018>.

#### Availability of data and materials

The data that support the findings of this study are available from INSERM U1219 Epicene Team, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of INSERM U1219 Epicene Team.

#### Transparency

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been

omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

#### The list of studied comorbidities

Hypertension, dyslipidemia, ischemic cardiopathy, apnea syndrome, anxiety, cancer, depression, diabetes, arrhythmia, asthma, left heart failure, bronchiectasis, atheroma, stroke, pulmonary arterial hypertension, lower limb arteriopathy, embolic thromboembolism, metabolic syndrome, undernutrition, osteoporosis, and sinus manifestation.

#### References

- [1] Dave Singh, Alvar Agusti, Antonio Anzueto, et al., Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease: the GOLD science committee report 2019, *Eur. Respir. J.* 53 (2019) 1900164, <https://doi.org/10.1183/13993003.00164-2019>.
- [2] GBD 2015 Mortality and Causes of Death Collaborators, Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015, *Lancet* 388 (2016) 1459-1544, [https://doi.org/10.1016/S0140-6736\(16\)31012-1](https://doi.org/10.1016/S0140-6736(16)31012-1).
- [3] J.R. Hurst, J. Vestbo, A. Anzueto, et al., Susceptibility to exacerbation in chronic obstructive pulmonary disease, *N. Engl. J. Med.* 363 (12) (2010) 1128-1138, <https://doi.org/10.1056/NEJMoa0909883>.
- [4] I.D. Pavord, P.W. Jones, P.-R. Burgel, et al., Exacerbations of COPD, *Int. J. Chronic Obstr. Pulm. Dis.* 11 (2016) 21-30, <https://doi.org/10.2147/COPD.S85978>.
- [5] M.T. Dransfield, K.M. Kunisaki, M.J. Strand, et al., Acute exacerbations and lung function loss in smokers with and without chronic obstructive pulmonary disease, *Am. J. Respir. Crit. Care Med.* 195 (2017) 324-330, <https://doi.org/10.1164/rccm.201605-1014OC>.
- [6] S. Suissa, S. Dell'Aniello, P. Ernst, Longterm natural history of chronic obstructive pulmonary disease: severe exacerbations and mortality, *Thorax* 67 (11) (2012) 957-963, <https://doi.org/10.1136/thoraxjnl-2011-201518>.
- [7] Y.S. Punekar, A. Shukla, H. Müllerova, COPD management costs according to the frequency of COPD exacerbations in UK primary care, *Int. J. Chronic Obstr. Pulm. Dis.* 9 (2014) 65-73, <https://doi.org/10.2147/COPD.S54417>.
- [8] M.K. Han, P.M. Quilbrera, E.E. Carretta, et al., Frequency of exacerbations in patients with chronic obstructive pulmonary disease: an analysis of the SPIROMICS cohort, *Lancet Respir. Med.* 5 (8) (2017) 619-626, [https://doi.org/10.1016/S2213-2600\(17\)30207-2](https://doi.org/10.1016/S2213-2600(17)30207-2).
- [9] P.-R. Burgel, J.-L. Paillasseur, D. Caillaud, et al., Clinical COPD phenotypes: a novel approach using principal component and cluster analyses, *Eur. Respir. J.* 36 (2010) 531-539, <https://doi.org/10.1183/09031936.00175109>.
- [10] Le Rouzic Olivier, Nicolas Roche, Alexis B. Cortot, et al., Defining the "frequent exacerbator" phenotype in COPD, *Chest* 153 (5) (2018) 1106-1115, <https://doi.org/10.1016/j.chest.2017.10.009>.
- [11] K.M. Beeh, T. Glaab, S. Stowasser, et al., Characterisation of exacerbation risk and exacerbator phenotypes in the POETCOPD trial, *Respir. Res.* 14 (2013) 116, <https://doi.org/10.1186/1465-9921-14-116>.
- [12] H. Yang, P. Xiang, E. Zhang, et al., Predictors of exacerbation frequency in chronic obstructive pulmonary disease, *Eur. J. Med. Res.* 19 (2014) 18, <https://doi.org/10.1186/2047-783X-19-18>.
- [13] R. Tomioka, T. Kawayama, M. Suetomo, et al., "Frequent exacerbator" is a phenotype of poor prognosis in Japanese patients with chronic obstructive pulmonary disease, *Int. J. Chronic Obstr. Pulm. Dis.* 11 (2016) 207-216, <https://doi.org/10.2147/COPD.S98205>.
- [14] M. Zielinski, M. Gasior, D. Jastrzebski, A. Desperak, D. Ziara, Influence of gaseous pollutants on COPD exacerbations in patients with cardiovascular comorbidities, *Adv. Exp. Med. Biol.* 1114 (2018) 11-17, [https://doi.org/10.1007/5584\\_2018\\_206](https://doi.org/10.1007/5584_2018_206).
- [15] C. Viniol, C.F. Vogelmeier, Exacerbations of COPD, *Eur. Respir. Rev.* 27 (2018) 170103, <https://doi.org/10.1183/16000617.0103-2017>.
- [16] Y.L. Huang, B. Mao, J. Mim, et al., The associations between comorbidity and exacerbation risk in patients with chronic obstructive pulmonary disease, *Zhonghua Jiehe He Huxi Zazhi* 41 (5) (2018 May 12) 349-354, <https://doi.org/10.3760/cma.j.issn.1001-0939.2018.05.009>.
- [17] H.S. Choi, C.K. Rhee, Y.B. Park, K.H. Yoo, S.Y. Lim, Metabolic syndrome in early chronic obstructive pulmonary disease: gender differences and impact on exacerbation and medical costs, *Int. J. Chronic Obstr. Pulm. Dis.* 14 (2019 Dec 10) 2873-2883, <https://doi.org/10.2147/COPD.S228497>.
- [18] T.M. Wilkinson, G.C. Donaldson, J.R. Hurst, et al., Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease, *Am. J. Respir. Crit. Care Med.* 169 (2004) 1298-1303, <https://doi.org/10.1164/rccm.200310-1443OC>.
- [19] J.A. Wedzicha, D. Banerji, K.R. Chapman, et al., Indacaterol-glycopyrronium versus salmeterol-fluticasone for COPD, *N. Engl. J. Med.* 374 (2016) 2222-2347, <https://doi.org/10.1056/NEJMoa1516385>.
- [20] F.J. Martinez, J. Vestbo, J.A. Anderson, et al., Effect of fluticasone furoate and vilanterol on exacerbations of COPD in patients with moderate airflow obstruction, *Am. J. Respir. Crit. Care Med.* 195 (2017) 881-888, <https://doi.org/10.1164/rccm.201607-1421OC>.

- [21] P.J. Poole, E. Chacko, R.W.B. Wood-Baker, et al., Influenza vaccine for patients with chronic obstructive pulmonary disease, *Cochrane Database Syst. Rev.* (2006), <https://doi.org/10.1002/14651858.CD002733.pub2>.
- [22] J.A. Walters, J.N.Q. Tang, P. Poole, et al., Pneumococcal vaccines for preventing pneumonia in chronic obstructive pulmonary disease, *Cochrane Database Syst. Rev.* (2017), <https://doi.org/10.1002/14651858.CD001390.pub4>.
- [23] R. Bekkat-Berkani, T. Wilkinson, P. Buchy, G. Dos Santos, D. Stefanidis, J. M. Devaster, et al., Seasonal influenza vaccination in patients with COPD: a systematic literature review, *BMC Pulm. Med.* 17 (1) (2017) 79, <https://doi.org/10.1186/s12890-017-0420-8>.
- [24] S. Sehatzadeh, Influenza and pneumococcal vaccinations for patients with chronic obstructive pulmonary disease (COPD): an evidence-based review, *Ont Health Technol. Assess Ser.* 12 (2012) 1–64. Available from: [www.hqontario.ca/en/mas/tech/pdfs/2012/rev\\_COPD\\_Vaccinations\\_March.pdf](http://www.hqontario.ca/en/mas/tech/pdfs/2012/rev_COPD_Vaccinations_March.pdf).
- [25] C.S. Wang, S.T. Wang, C.T. Lai, L.J. Lin, P. Chou, Impact of influenza vaccination on major cause-specific mortality, *Vaccine* 25 (2007) 1196–1203, <https://doi.org/10.1016/j.vaccine.2006.10.015>.
- [26] K.L. Nichol, L. Baken, A. Nelson, Relation between influenza vaccination and outpatient visits, hospitalization, and mortality in elderly persons with chronic lung disease, *Ann. Intern. Med.* 130 (1999) 397–403, <https://doi.org/10.7326/0003-4819-130-5-199903020-00003>.
- [27] R. Garrastazu, J.L. Garcia-Rivero, M. Ruiz, et al., Prevalence of influenza vaccination in chronic obstructive pulmonary disease patients and impact on the risk of severe exacerbations, *Arch. Bronconeumol.* 52 (2016) 88–95, <https://doi.org/10.1016/j.arbres.2015.09.001>.
- [28] T. Ozlu, Y. Bulbul, D. Aydin, et al., Immunization status in chronic obstructive pulmonary disease: a multicenter study from Turkey, *Ann. Thorac. Med.* 14 (1) (2019 Jan-Mar) 75–82, [https://doi.org/10.4103/atm.ATM\\_145\\_18](https://doi.org/10.4103/atm.ATM_145_18).
- [29] Bartolome R. Celli, Claudia G. Cote, Jose M. Marin, et al., The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease, *N. Engl. J. Med.* 350 (2004) 1005–1012, <https://doi.org/10.1056/NEJMoa021322>. PMID:14999112.
- [30] B.R. Celli, W. MacNee, A. Agusti, et al., Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper, *Eur. Respir. J.* 23 (2004) 932–946, <https://doi.org/10.1183/09031936.04.00014304>.
- [31] Donald A. Mahler, Carolyn K. Wells, Evaluation of clinical methods for rating dyspnea, *Chest* 93 (1988 Mar) 580–586, <https://doi.org/10.1378/chest.93.3.580>. PMID : 3342669.
- [32] Sonia Coton, William M. Vollmer, Eric Bateman, et al., Severity of airflow obstruction in chronic obstructive pulmonary disease (COPD): proposal for a new classification, *COPD* 14 (5) (2017 Aug 11) 469–475, <https://doi.org/10.1080/15412555.2017.1339681>.
- [33] CESP (Centre de recherche en Épidémiologie et Santé des Populations), INSERM – université Paris Sud. Mise en œuvre du décret n° 98–37 autorisant l'accès aux données relatives au décès des personnes inscrites au Répertoire National d'Identification des Personnes Physiques (RNIPP) dans le cadre des recherches dans le domaine de la santé, Accessed, <http://cesp.vjf.inserm.fr/svcd>. (Accessed 27 May 2015).
- [34] C. Genolini, X. Alcaque, M. Sentenac, C. Arnaud, kml and kml3d: R packages to cluster longitudinal data, *J. Stat. Software* 65 (4) (2015) 1–34, <https://doi.org/10.18637/jss.v065.i04>.
- [35] X. Robin, N. Turck, A. Hainard, et al., pROC: an open-source package for R and Sp to analyze and compare ROC curves, *BMC Bioinf.* 12 (2011) 77, <https://doi.org/10.1186/1471-2105-12-77>.
- [36] J.A. Wedzicha, M. Miravittles, J.R. Hurst, et al., Management of COPD exacerbations: a European respiratory society/American thoracic society guideline, *Eur. Respir. J.* (2017), <https://doi.org/10.1183/13993003.00791-2016>.
- [37] K. Kostikas, A. Clemens, F. Patalano, Prediction and prevention of exacerbations and mortality in patients with COPD, *Expert Rev. Respir. Med.* 10 (7) (2016) 739–753, <https://doi.org/10.1080/17476348.2016.1185371>.
- [38] J.D. Beckham, A. Cadena, J. Lin, et al., Respiratory viral infections in patients with chronic, obstructive pulmonary disease, *J. Infect.* 50 (2005) 322–330, <https://doi.org/10.1016/j.jinf.2004.07.011>.
- [39] F.W. Ko, M. Ip, P.K. Chan, et al., Viral etiology of acute exacerbations of COPD in Hong Kong, *Chest* 132 (2007) 900–908, <https://doi.org/10.1378/chest.07-0530>.
- [40] T.E. McManus, A.M. Marley, N. Baxter, et al., Respiratory viral infection in exacerbations of COPD, *Respir. Med.* 102 (2008) 1575–1580, <https://doi.org/10.1016/j.rmed.2008.06.006>.
- [41] G. Rohde, A. Wiethage, I. Borg, et al., Respiratory viruses in exacerbations of chronic obstructive pulmonary disease requiring hospitalisation: a case-control study, *Thorax* 58 (2003) 37–42, <https://doi.org/10.1136/thorax.58.1.37>.
- [42] European Centre for Disease Prevention and Control (ECDC), Seasonal Influenza Vaccination in Europe. Overview of Vaccination Recommendations and Coverage Rates in the EU Member States for the 2012–13 Influenza Season, 2015.
- [43] Vaccines against Influenza WHO position paper - november 2012, *Wkly. Epidemiol. Rec.* 87 (2012) 461–476. PMID: 23210147.
- [44] Centers for Disease Control and Prevention (CDC), People at High Risk of Developing Flu-Related Complications, 2015. Accessed on, [http://www.cdc.gov/flu/about/disease/high\\_risk.htm](http://www.cdc.gov/flu/about/disease/high_risk.htm). (Accessed 24 February 2016).
- [45] L.A. Grohskopf, L.Z. Sokolow, K.R. Broder, et al., Prevention and control of seasonal influenza with vaccines, *MMWR Recomm. Rep.* 65 (2016) 1–54, <https://doi.org/10.15585/mmwr.rrr6505a1>.

## V. L'analyse de mortalité et les phénotypes cliniques des patients BPCO identifiés par l'analyse en cluster la plus optimale (n=2653 et n=2968)

Cette partie porte sur les résultats de l'étude de l'analyse de mortalité et l'identification des phénotypes cliniques au sein la cohorte PALOMB. Les résultats relatifs à cette étude font l'objet d'un article en cours d'évaluation dans la revue scientifique Plos One ; d'une communication orale au congrès international du CPLF (24<sup>ème</sup> édition du congrès de pneumologie de langue française) en janvier 2020 à Paris ; de deux communications affichées (Posters) à la conférence internationale de l'ATS (American Thoracic Society) en Mai 2020 à Philadelphia (USA) et au congrès international de l'ERS (European Respiratory Society Congress) en septembre 2020 à Vienne (Autriche). Ces résultats seront également présentés sous forme d'une communication orale au congrès international du CPLF (25<sup>ème</sup> édition du congrès de pneumologie de langue française) en janvier 2021 à Lille.

- **Article en cours de publication :**

El Hassane Ouaalaya, Maéva Zysman, Laurent Falque, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Ejnaini, Frédéric Le Guillou, Mathieu Molimard, Chantal Raheison-Semjen. Main determinants associated with mortality in COPD patients. Plos One.

- **Communication orale :**

E. H. Ouaalaya, A. Bernady, J. Casteigt, J. M. Dupis, C. Nocent-Eijnaini, M. Sabatini, L. Falque, F. Le Guillou, L. Nguyen, A. Ozier, M. Molimard, C. Raheison-Semjen. Le cancer et les comorbidités cardiovasculaires sont associés à la mortalité chez les patients BPCO. Revue des Maladies Respiratoires Actualités. Volume 12, numéro 1, janvier 2020, pages 41-42. DOI : <https://doi.org/10.1016/j.rmra.2019.11.070> [CPLF - Paris 2020 (24e édition du Congrès de Pneumologie de Langue Française) : sélection via résumé]

- **Communication affichée (Poster) :**

E.H. Ouaalaya, L. Falque, J.M. Dupis, M. Sabatini, A. Bernady, L. Nguyen, A. Ozier, C. Nocent-Ejnaini, F. Le Guillou, M. Molimard, C. Raheison-semjen. Cancer and Cardiovascular Comorbidities Are Associated with COPD Mortality: The French PALOMB Cohort. American Thoracic Society 2020 International Conference -Philadelphia- USA. Conference Paper, May 15-20, 2020. DOI: [https://doi.org/10.1164/ajrccm-conference.2020.201.1\\_MeetingAbstracts.A7162](https://doi.org/10.1164/ajrccm-conference.2020.201.1_MeetingAbstracts.A7162)

El-hassane Ouaalaya, Maéva Zysman, Émilie Berteaud, Laurent Falque, Laurent Nguyen, Jean Michel Dupis, Marielle Sabatini, Alain Bernady, Annaig Ozier, Cécilia Nocent-Ejnaini, Christophe Roy, Frédéric Le Guillou, Emmanuel Monge, Mohammed Aliati, Anne Prudhomme, Laura Petrov, Marie Line Quinquenel, Mohammed Staali, Frédéric Pilard, Marc Sapène, Julien Casteigt, Jean Moinard, Yannick Daoudi, Esther Iglesias, Élodie Blanchard, Julie Macey, Rémi Veillon, Xavier Demant, Claire Bon, Leo Grassion, Mathieu Molimard, Chantal Raheison-Semjen. COPD phenotypes and 5-years mortality risk prediction: PALOMB Cohort. European Respiratory Journal 2020; [The virtual ERS International Congress 7-9 September, 2020]

## V.1. Les caractéristiques initiales des patients inclus dans cette étude

Nous rapportons ici le suivi longitudinal des 2653 patients BPCO inclus dans cette étude observationnelle transversale. Il s'agissait principalement d'un échantillon caractérisé par (un âge moyen de 66 ans, 64% d'hommes et un taux de mortalité de 12,8%), les maladies cardiovasculaires, le diabète, la dépression et les cancers ont été identifiés respectivement chez 48,10%, 12,44%, 13,80% et 18,43% des patients. La médiane de l'indice BOD était est de 2 [1-3]. La médiane du temps de suivi pour tous ces patients BPCO était de 29 mois (22 mois chez les patients décédés vs 30 mois chez les patients vivants,  $p < 0,0001$ ) (**Tableau supplémentaire 1 disponible en Annexe**). Les caractéristiques initiales sont résumées dans le **tableau 12**.

**Tableau 12.** Association des caractéristiques de 2653 patients BPCO à l'inclusion selon le statut vital.

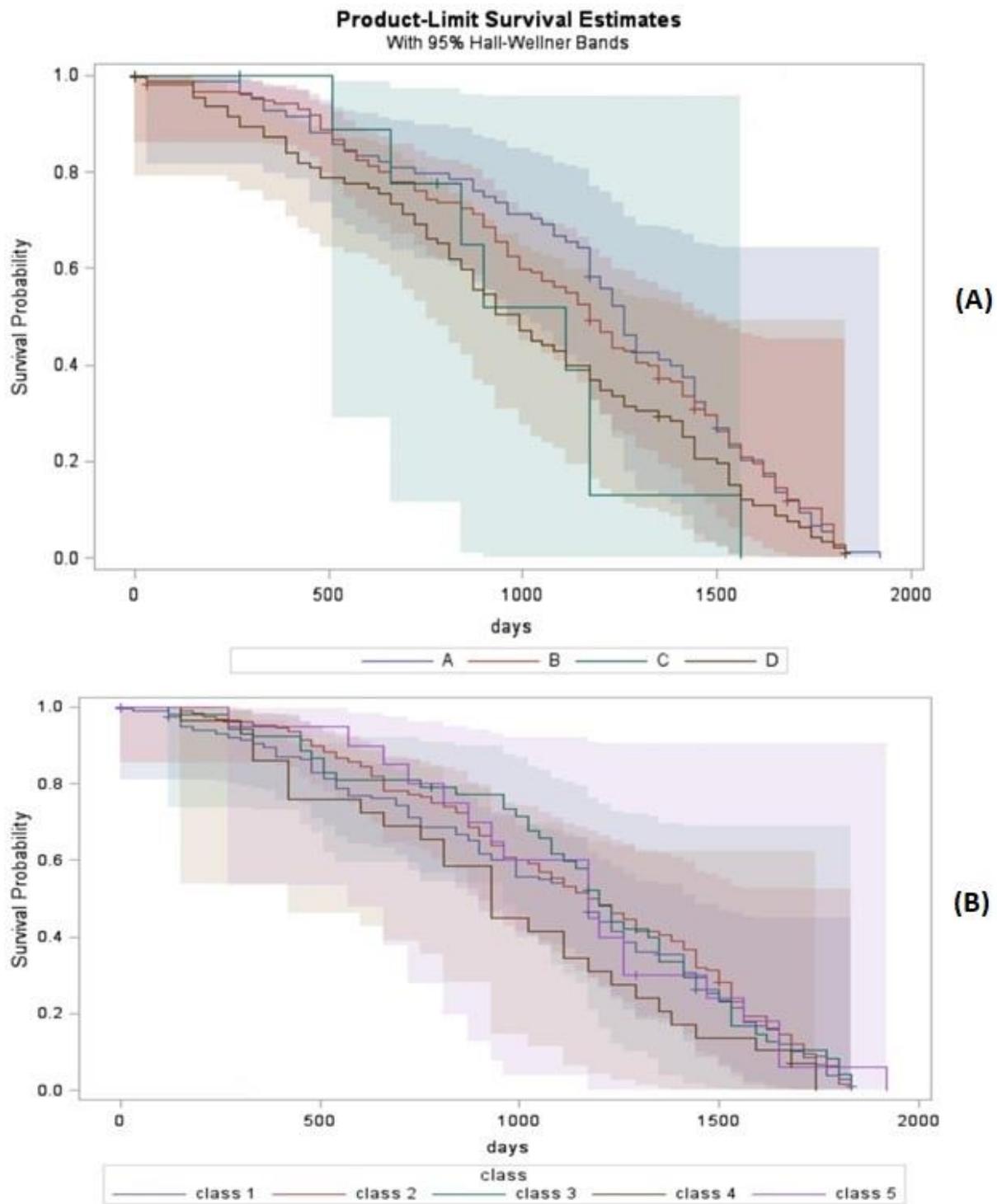
	All COPD, n=2,653	Deceased patients, n=339 (12.78)	Live patients, n=2,314 (87.22)	P Value
Sex (male), n (%)	1,705 (64.27)	254 (74.93)	1,451 (62.71)	<0.0001
Age, yrs. mean(SD)	66.12 (10.87)	71.3 (10.21)	65.37 (10.76)	<0.0001
BMI, (kg/m <sup>2</sup> ). mean(SD)	26.6 (11.11)	25.74 (6.13)	26.73 (11.66)	0.1231
FEV <sub>1</sub> , % pred, mean (SD)	58.32 (40.26)	49.11 (17.24)	59.68 (42.44)	<0.0001
Chronic cough, n (%)	1,537 (57.93)	218 (64.31)	1,319 (57)	0.0109
Chronic sputum, n (%)	1,105 (41.65)	155 (45.72)	950 (41.05)	0.1035
Current smokers, n (%)	1,030 (40.23)	113 (34.45)	917 (41.08)	0.0222
Dyspnea, ≥ 2 mMRC, n (%)	1,533 (57.78)	251 (74.04)	1,282 (55.4)	<0.0001
Frequent exacerbations (≥2), n (%)	682 (25.71)	102 (30.09)	580 (25.06)	0.0481
Pulmonary rehabilitation, n (%)	186 (7.01)	38 (11.21)	148 (6.40)	0.0012
Smoking cessation, n (%)	235 (8.86)	13 (3.83)	222 (9.59)	0.0005
RV, mean (SD)	147.3 (51.42)	156.4 (56.98)	146.02 (50.43)	0.0016
TLC, mean (SD)	105.8 (20.32)	104.4 (21.94)	106.04 (20.07)	0.1649
RV/TLC, % pred, mean (SD)	137.4 (37.87)	146.8 (30.44)	136.1 (38.65)	<0.0001
Hyperinflation (RV/TLC≥120%), n (%)	1,902 (71.69)	280 (82.60)	1,622 (70.10)	<0.0001

## V.2. Les déterminants de la mortalité

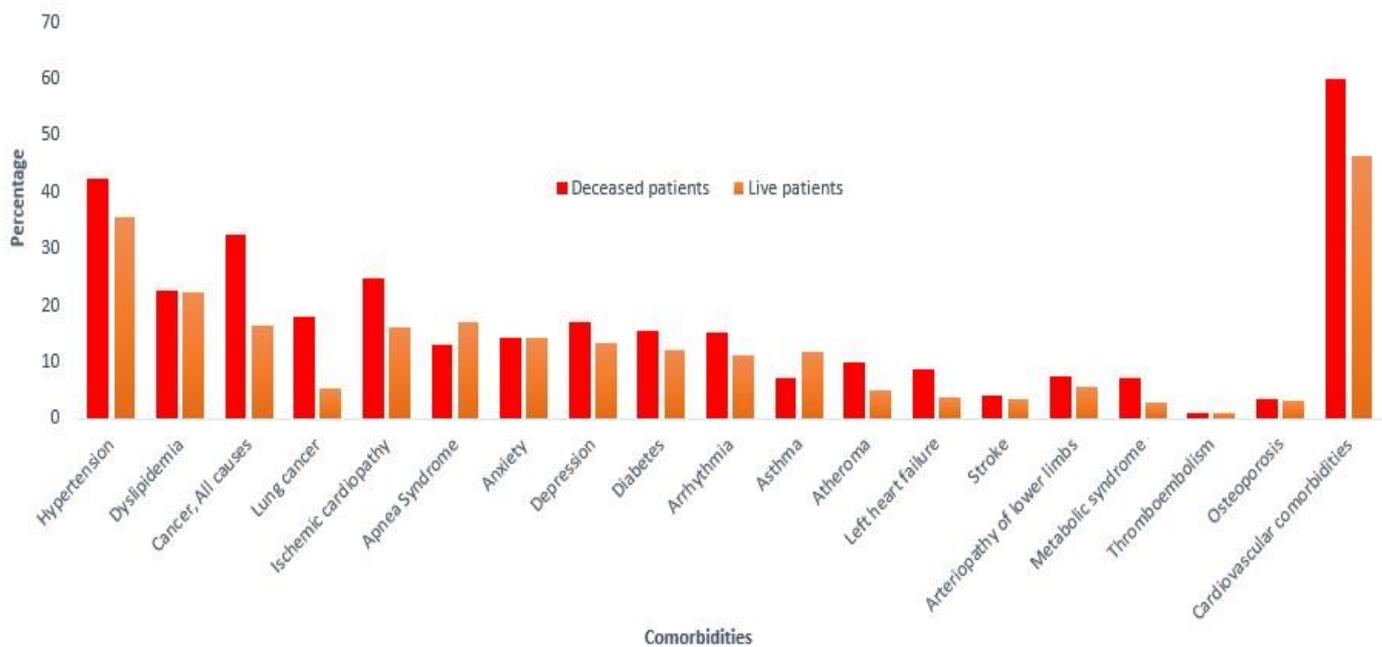
Le statut vital des patients BPCO a été systématiquement collecté en avril 2019. Au cours de la période de suivi, 339 (12,8%) sujets sont décédés. Nous avons comparé les caractéristiques cliniques et fonctionnelles des sujets décédés au cours du suivi par rapport aux patients vivants. Plusieurs caractéristiques cliniques et fonctionnelles étaient associées à la mortalité : sexe masculin, âge avancé, la dyspnée élevée (mMRC  $\geq 2$ ), les exacerbations fréquentes, un indice BOD plus élevée, un VEMS bas et la distension thoracique (évaluée par le % prédit du rapport VR/CPT) (**Voir Tableau 12**). La survie était significativement différente selon : la classification GOLD ABCD (Log Rank,  $p = 0,0036$ ) (**voir Figure 20**) et la distension thoracique (Log Rank,  $p = 0,0189$ ) (**voir Fig. supplémentaire 1 disponible en Annexe**).

Par ailleurs, chez les patients décédés, certaines comorbidités étaient significativement plus fréquentes telles que la cardiopathie ischémique, le trouble du rythme, le syndrome métabolique, la dépression, le diabète et l'athérome (**voir Figure 21**). Les cancers étaient également plus fréquents chez les patients décédés (32,5% vs 16,4%,  $p < 0,0001$ ) en particulier le cancer du poumon, étant significativement plus fréquents chez les patients décédés (16,5% vs 4,9%,  $p < 0,0001$ ), alors que les autres cancers ne l'étaient pas (sein, cancer urogénital, laryngé). Le nombre de comorbidités augmentait avec la sévérité de la BPCO, la médiane du nombre de comorbidités était de deux. En outre, chez le groupe de patients ayant deux comorbidités ou plus, plus d'hommes que de femmes sont décédés (49,8% vs 15,3%,  $p = 0,0095$ ). Les comorbidités cardiovasculaires étaient plus fréquentes chez les hommes alors que l'anxiété-dépression et l'ostéoporose étaient plus fréquentes chez les femmes. Dans le groupe de patients ayant des comorbidités cardiovasculaires, plus d'hommes que de femmes sont décédés (**voir tableau supplémentaire 2 disponible en Annexe**).

L'analyse multivariée ajustée sur l'âge a montré que les facteurs significativement associés à la mortalité étaient le cancer du poumon, l'absence du sevrage tabagique, le syndrome métabolique, le sexe masculin, la toux chronique, l'athérome et l'hyperinflation (**Tableau 13**). Même après ajustement sur l'indice de BOD (**Tableau 14**) ou sur le sexe (**voir tableau supplémentaire 3 disponible en Annexe**), le cancer du poumon reste le facteur le plus important parmi les déterminants de la mortalité.



**Figure 17.** L'analyse de Kaplan-Meier pour évaluer la mortalité toutes causes confondues à 5 ans dans la cohorte PALOMB BPCO : **(A)** selon les outils d'évaluation ABCD de GOLD. **(B)** selon les cinq classes identifiées par l'algorithme CART.



**Figure 18.** Prévalence des comorbidités chez les patients BPCO vivants et décédés.

**Tableau 13.** Les déterminants de la mortalité ajustés sur l'âge chez 2653 patients BPCO.

	Estimated coefficients and 95% Cis for the adjusted ORs
<b>Lung cancer</b> (Presence Vs Absence)	4.81 [3.36 – 6.89]
<b>Smoking cessation</b> (Absence Vs Presence)	2.42 [1.32 – 4.40]
<b>Metabolic syndrome</b> (Presence Vs Absence)	2.47 [1.45 – 4.22]
<b>FEV<sub>1</sub>, % pred</b> (<50% Vs ≥50%)	1.85 [1.41 – 2.42]
<b>Depression</b> (Presence vs Absence)	1.81 [1.26 – 2.59]
<b>Dyspnea</b> (mMRC ≥2 Vs mMRC 0-1)	1.73 [1.30 – 2.30]
<b>Male Gender</b>	1.68 [1.26 – 2.24]
<b>Atheroma</b> (Presence vs Absence)	1.60 [1.03 – 2.48]
<b>Hyperinflation</b> (RV/TLC≥120% Vs RV/TLC<120%)	1.55 [1.11 – 2.16]
<b>Ischemic cardiopathy</b> (Presence vs Absence)	1.37 [1.01 – 1.86]

**Tableau 14.** Les déterminants de la mortalité ajustés sur l'indice BOD chez 2653 patients BPCO.

	Estimated coefficients and 95% Cis for the adjusted ORs
<b>Lung cancer</b> (Presence Vs Absence)	3.02 [1.89 - 4.83]
<b>Smoking cessation</b> (Absence Vs Presence)	2.38 [1.29 – 4.36]
<b>Metabolic syndrome</b> (Presence Vs Absence)	2.10 [1.23 – 3.59]
<b>Age</b> ( $\geq 66$ Vs $< 66$ )	1.97 [1.52 – 2.55]
<b>Cancer, Other causes</b> (Presence Vs Absence)	1.80 [1.26 – 2.58]
<b>Atheroma</b> (Presence Vs Absence)	1.70 [1.09 – 2.65]
<b>Male Gender</b>	1.63 [1.23 – 2.16]
<b>Hyperinflation</b> (RV/TLC $\geq 120\%$ Vs RV/TLC $< 120\%$ )	1.56 [1.13 – 2.17]
<b>Chronic cough</b> (Presence Vs Absence)	1.31 [1.01 – 1.70]

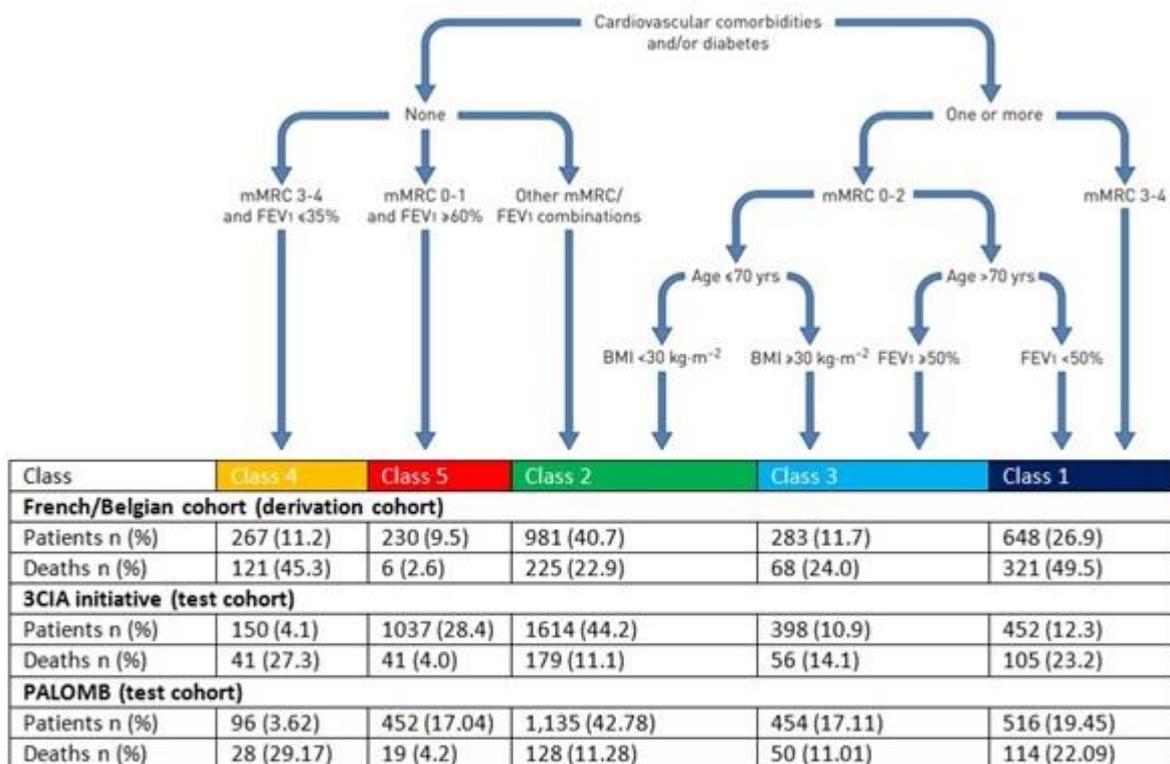
### V.3. L'analyse de déclin du VEMS : Le sous-groupe

Dans un sous-groupe (n = 770) de patients BPCO ayant eu au moins une visite de suivi avec des données complètes sur le VEMS, le déclin du VEMS était associé à une mortalité plus élevée sans atteindre la significativité statistique (**voir tableau supplémentaire 4 disponible en Annexe**). Une analyse descriptive a été effectuée pour montrer les différences entre ce sous-échantillon et le reste de l'échantillon (**voir tableau supplémentaire 5 disponible en Annexe**). La distribution de la variation du VEMS exprimée en pourcentage de la valeur prédite selon la classification GOLD ABCD était plus manifestement normale (**voir Fig. 2 supplémentaire disponible en Annexe**).

### V.4. L'analyse CART (Arbre de classification et de régression)

L'analyse CART a fourni un algorithme qui a facilité l'affectation des patients aux sous-groupes identifiés par l'analyse en clusters (**Figure 22**). Les caractéristiques cliniques des patients, selon les cinq clusters obtenues en appliquant cet algorithme, sont présentées dans le **tableau E6 disponible en Annexe**. Le premier groupe de mauvais pronostic (CART 1) était caractérisé par des taux élevés de

cancer, d'hypertension, de cardiopathie ischémique, de diabète, d'arythmie, d'athérome et d'insuffisance cardiaque gauche (**voir tableau supplémentaire 7 disponible en Annexe**) alors que l'anxiété, la dépression, l'asthme et la dénutrition étaient plus fréquente chez les patients du deuxième groupe de mauvais pronostic (CART 4). La médiane de l'indice de BOD chez les patients de CART 1 était de 4 [3-5], tandis que cette médiane s'élevait à 6 [5-6] chez les patients de CART 4 (**voir tableau supplémentaire 8 disponible en Annexe**). La mortalité était plus élevée dans les groupes de mauvais pronostic CART 1 et 4 avec respectivement 22,1% et 29,2% (**Figure 22**). Le groupe CART 4 était bien séparé des autres, comme décrit en utilisant l'analyse Kaplan Meier, mais le groupe CART 1 ne l'était pas ( $p < 0,0001$ ). Cependant, aucune différence significative n'a été observée entre les autres phénotypes, indiquant des taux de mortalité comparables pendant la période de suivi relativement courte (**Figure 22. B**). En conclusion, la classification CART ne permet pas une meilleure discrimination en termes de pronostic comparativement à l'outil d'évaluation ABCD de GOLD (**Figure 22. A et Figure 22. B**).



**Figure 19.** L'algorithme développé par CART pour la classification des patients BPCO. Application aux cohortes franco-belge, 3CIA et PALOMB.

## V.5. Les phénotypes cliniques des patients BPCO identifiés par l'analyse en cluster la plus optimale (n=2968)

Nous disposons, dans notre étude, de 4 variables quantitatives parmi nos variables d'intérêt (âge, IMC, VEMS et le rapport VR/CPT). Pour chaque sous-groupe, vis-à-vis des classifications de la sévérité, nous conduisons un test statistique de Shapiro sur ces 4 variables cliniques. Nous obtenons des p-values < 0,05 et nous rejetons donc l'hypothèse de normalité. Pour les 140 patients ayant des valeurs manquantes, 56,42% sont des hommes avec une moyenne d'âge de 68,67 ans ( $\pm 11,32$ ) et un IMC de 26,04 kg/m ( $\pm 6,03$ ). Pour les stades "L", "M", "S" et "TS", de la sévérité par spirométrie, ils enregistrent des fréquences de 17,85%, 32,14%, 35% et 15%, respectivement. De plus, 20% de ces patients sont des exacerbateurs fréquents et 67,14% ont une dyspnée élevée (mMRC  $\geq 2$ ). Concernant les comorbidités, 46,42% ont au moins une maladie pleuropulmonaire. Les maladies cardiovasculaires sont présentes chez 49,28% de ces patients.

Les 2968 patients BPCO, n'ayant pas de valeurs manquantes, ont une fréquence d'hommes de 63,74% avec une moyenne d'âge de 70 ans ( $\pm 10,89$ ) et un IMC de 26,3 kg/m ( $\pm 5,84$ ). Seulement 5% de la population n'a jamais fumé, les anciens fumeurs, quant à eux, sont majoritaires (56,3%). Le taux de mortalité global est de 11,3%.

Concernant les mesures spirométriques et les indicateurs de la sévérité, les patients BPCO ont, en moyenne, un VEMS de 61,31% ( $\pm 20,08$ ) de la valeur prédite et une distension thoracique (VR/CPT) de 137,17% ( $\pm 35,73$ ). Plus précisément, la classification GOLD 2007 indique que 51,7 % des patients sont en stade "Modéré" de la maladie, contre 30,10% en stades "Sévère" et "Très Sévère".

Quant aux évaluations symptomatiques, seulement 25,95% des patients sont des exacerbateurs fréquents. Ils sont, néanmoins, 56,36% à avoir une dyspnée élevée (mMRC  $\geq 2$ ). Les stades sévères ("B" et "D") de GOLD 2017, constituent 38,37% et 18%, respectivement, de la population de notre étude. L'évaluation des comorbidités montre que seulement 19,67% n'ont pas de comorbidités et seulement 6% ont plus de 6 comorbidités en même temps. En particulier, pour les comorbidités pleuropulmonaires, 56,5% des patients ont au moins une comorbidité. Par ailleurs, les maladies cardiovasculaires constituent 46,5%. Et enfin, 64% des patients ont au moins une comorbidité métabolique.

- Remarque : La comparaison entre les patients ayant des valeurs manquantes et ceux ayant les données complètes, montre une ressemblance entre les deux populations. Nous nous contenterons de l'échantillon complet et nous ne procéderons pas à une imputation des valeurs manquantes, dans le cadre de cette étude.

**Les tableaux 15 et 16**, ci-dessous, présentent les caractéristiques des 2968 patients selon la classification GOLD 2007. Plusieurs variables cliniques montrent des différences significatives vis-à-vis de la classification de la sévérité par spirométrie (GOLD 2007). Comparativement aux groupes

de patients ("L" et "M"), les patients les plus sévères ("S" et "TS") se caractérisent par une population plus âgée (72 ans et 69 ans respectivement) et un taux de mortalité plus élevé (16,20% et 21,30% respectivement). Ils présentent un taux important d'indicateurs symptomatiques (i.e. Toux, Expectoration, mMRC, Exacerbations). En effet, cela est confirmé par la corrélation (p-value < 0,001) entre la variable d'évaluation symptomatique de la sévérité GOLD 2017 - l'outil ABCD et celle de l'évaluation spirométrique de la sévérité GOLD 2007.

L'asthme, le cancer (sauf cancer de poumon), l'athérome, l'AVC, l'artériopathie, la dyslipidémie, le syndrome métabolique, les manifestations sinusiennes, les maladies thromboembolique et la dépression ne montrent pas une différence significative selon GOLD 2007. Les comorbidités cardiovasculaires sont plus fréquentes dans les groupes sévères ("S", "TS"). En particulier, le groupe "TS" est aussi plus caractérisé par la dénutrition, les problèmes digestifs, ainsi que les maladies ostéo-articulaires (ostéoporose et ostéodensitométrie).

Tableau 15. Description des variables cliniques vis-à-vis de la Classification GOLD 2007.

Variables cliniques	Classification GOLD 2007				p-value
	L   N=540 18,20%	M   N=1535 51,70%	S   N=715 24,10%	TS   N=178 6%	
<b>Hommes, 1892 (63.74%)</b>	307 (56.9)	981 (63.9)	480 (67.1)	124 (69.7)	0.001***
<b>Âge, 70.07 (10.89)</b>	66.5 (10.7)	70.5 (11.0)	72.0 (10.5)	69.3 (9.66)	<0.001***
<b>Décès, 335 (11.30%)</b>	21 (3.89)	160 (10.4)	116 (16.2)	38 (21.3)	<0.001***
<b>IMC (kg/m), 26.30 (5.84)</b>	25.6 (4.88)	27.1 (5.99)	25.9 (5.95)	23.2 (5.26)	<0.001***
<b>Tabagisme, N (%) :</b>					
<b>Jamais, 148 (5%)</b>	19 (3.52)	76 (4.95)	43 (6.01)	10 (5.62)	<0.001***
<b>Ancien Fumeur, 1671 (56.30%)</b>	267 (49.4)	855 (55.7)	427 (59.7)	122 (68.5)	
<b>Fumeur Actif, 1149 (38.70%)</b>	254 (47.0)	604 (39.3)	245 (34.3)	46 (25.8)	
<b>Toux, 1722 (58.02%)</b>	289 (53.5)	889 (57.9)	432 (60.4)	112 (62.9)	0.047*
<b>Expectoration, 1251 (42.15%)</b>	206 (38.1)	619 (40.3)	337 (47.1)	89 (50.0)	0.001***
<b>mMRC, N (%) :</b>					
<b>0-1   1295 (42.15%)</b>	368 (68.1)	747 (48.7)	162 (22.7)	18 (10.1)	<0.001***
<b>≥ 2   1673 (56.36%)</b>	172 (31.9)	788 (51.3)	553 (77.3)	160 (89.9)	
<b>Exacerbations, N (%) :</b>					
<b>0-1   2198 (74.05%)</b>	439 (81.3)	1186 (77.3)	473 (66.2)	100 (56.2)	<0.001***
<b>≥ 2   770 (25.95%)</b>	101 (18.7)	349 (22.7)	242 (33.8)	78 (43.8)	
<b>VR/CPT, 137.17 (36.73)</b>	113 (54.4)	131 (22.4)	158 (25.2)	182 (28.8)	<0.001***
<b>VEMS, 61.31 (20.08)</b>	90.7 (9.33)	64.7 (8.27)	41.1 (5.69)	24.4 (3.82)	0.000***
<b>GOLD 2017, N (%)</b>					
<b>A   1059 (35.68%)</b>	308 (57.0)	612 (39.9)	124 (17.3)	15 (8.43)	<0.001***
<b>B   1139 (38.37%)</b>	131 (24.3)	574 (37.4)	349 (48.8)	85 (47.8)	
<b>C   236 (7.95%)</b>	60 (11.1)	135 (8.79)	38 (5.31)	3 (1.69)	
<b>D   534 (18%)</b>	41 (7.59)	214 (13.9)	204 (28.5)	75 (42.1)	

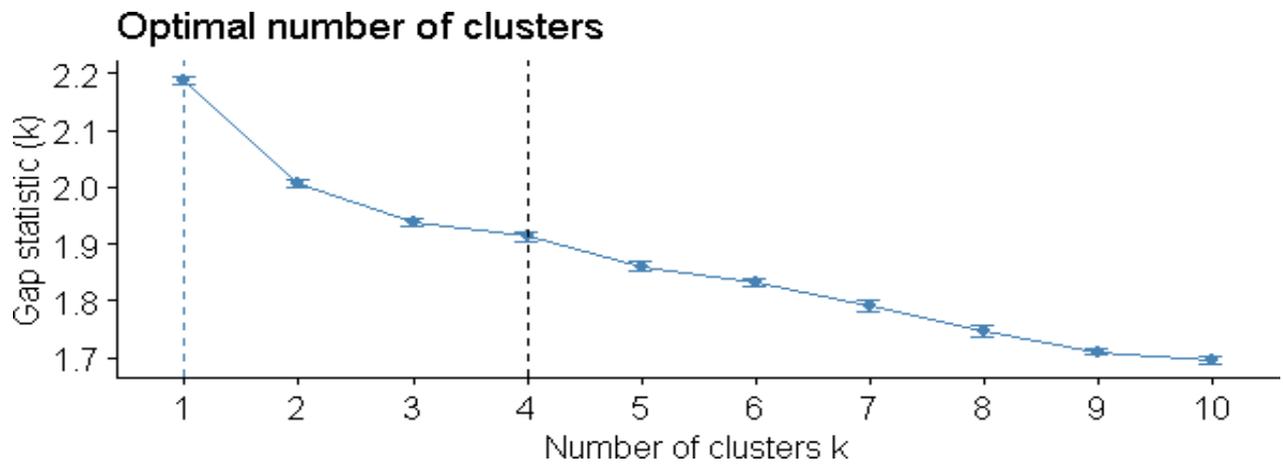
Tableau 16. Distribution des comorbidités vis-à-vis de la Classification GOLD 2007.

Comorbidités	Classification GOLD 2007				p-value
	L   N=540 18,2%	M   N=1535 51,7%	S   N=715 24,10%	TS   N=17 86%	
<b>SAS, 490 (16.51%)</b>	114 (21.1)	278(18.1)	85 (11.9)	13(7.3)	<0.001
<b>Asthme, 324 (11%)</b>	66 (12.2)	167(10.9)	75 (10.5)	16(8.99)	0.623
<b>DDB, 145 (4.88%)</b>	25 (4.6)	58(3.7)	53 (7.41)	9(5.06)	0.003**
<b>Cancer Poumon, 204 (6.87%)</b>	29 (5.3)	120(7.8)	50 (6.99)	5(2.81)	0.034*
<b>Autre Cancer, 351 (11.82%)</b>	94 (17.4)	305(19.9)	134 (18.7)	22(12.4)	0.083
<b>Hypertension, 1090 (36.72%)</b>	148 (27.4)	593(38.6)	275 (38.5)	74(41.6)	<0.001***
<b>Cardiopathie Ischémique, 502 (17%)</b>	71 (13.1)	284(18.5)	111 (15.5)	36(20.2)	0.014*
<b>Trouble du rythme, 346 (11.65%)</b>	39 (7.22)	191(12.4)	89 (12.4)	27(15.2)	0.003**
<b>Insuffisance Cardiaque Gauche, 121 (4.07%)</b>	9 (1.6)	55(3.5)	41 (5.73)	16(8.9)	<0.001***
<b>Athérome, 170 (5.72%)</b>	20 (3.7)	88(5.7)	50 (6.99)	12(6.7)	0.087
<b>AVC, 99 (3.33%)</b>	15 (2.7)	55(3.5)	23 (3.22)	6(3.3)	0.839
<b>HTAP, 64 (2.15%)</b>	6 (1.1)	17(1.1)	22 (3.08)	19(10.7)	<0.001***
<b>Artériopathie, 190 (6.40%)</b>	24 (4.4)	100(6.5)	49 (6.85)	17(9.5)	0.083
<b>Diabète, 367 (12.36%)</b>	42 (7.7)	219(14.3)	87 (12.2)	19(10.7)	0.001***
<b>Dyslipidémie, 673 (22.67%)</b>	127 (23.5)	358(23.3)	153 (21.4)	35(19.7)	0.538
<b>Syndrome Métabolique, 105 (3.53%)</b>	12 (2.2)	58(3.7)	29 (4.06)	6(3.37)	0.311
<b>Dénutrition, 59 (17%)</b>	3 (0.5)	13(0.8)	25 (2)	18(10.1)	<0.001***
<b>Manifestation Sinusiennes, 160 (5.40%)</b>	40 (7.4)	70(4.5)	40 (5.5)	10(5.6)	0.091
<b>Problèmes Digestifs, 75 (2.52%)</b>	12 (2.2)	28(1.8)	31 (4.3)	4(2.25)	0.008**
<b>Maladie Thromboembolique, 36 (1.21%)</b>	6 (1.1)	20(1.3)	6 (0.8)	4(2.25)	0.442
<b>Ostéoporose, 97 (3.26%)</b>	14 (2.5)	33(2.1)	37 (5.1)	13(7.3)	<0.001***
<b>Ostéodensitométrie, 97 (3.26%)</b>	14 (2.5)	33(2.1)	37 (5.1)	13(7.3)	<0.001***
<b>Anxiété, 430 (14.48%)</b>	47 (8.7)	211(13.7)	120 (16.8)	52(29.2)	<0.001***
<b>Dépression, 382 (12.87%)</b>	70 (13)	191(12.4)	87 (12.2)	34(19.1)	0.081

### V.5.1 Partitionnement en k-Médoïdes – Choix du nombre optimal de clusters

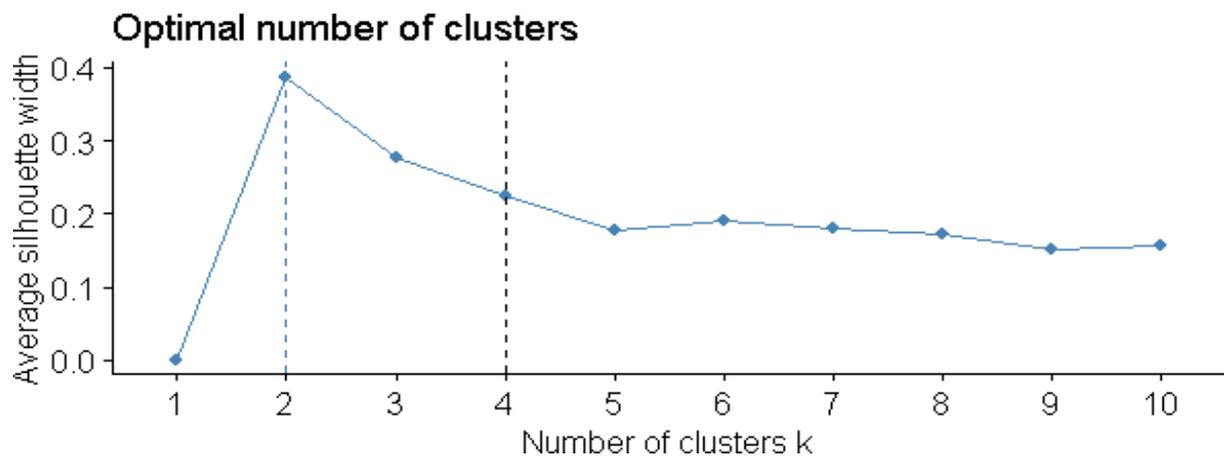
Ci-dessous, nous avons la représentation des résultats des méthodes utilisées pour le choix du nombre optimal de classes, pour l'initialisation du partitionnement en k-médoïdes. Nous retrouvons  $k = 4$  (le même résultat que celui des méthodes de clustering : k-means et HCPC).

- **Gap Statistic :**



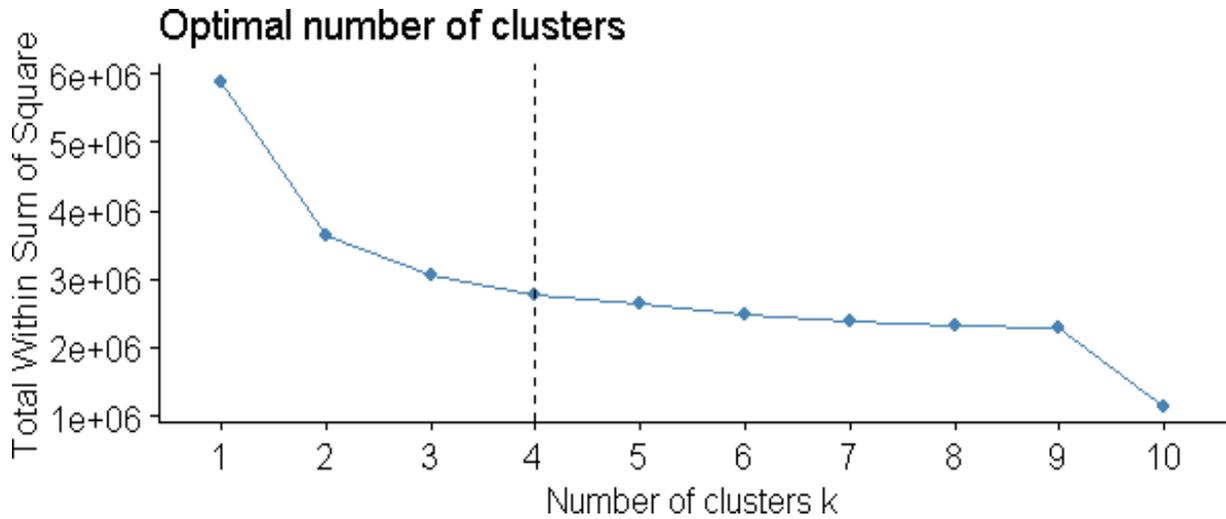
**Figure 20.** Choix du nombre de classes - Gap Statistic

- **Statistique de Silhouette :**



**Figure 21.** Choix du nombre de classes - Statistique de Silhouette

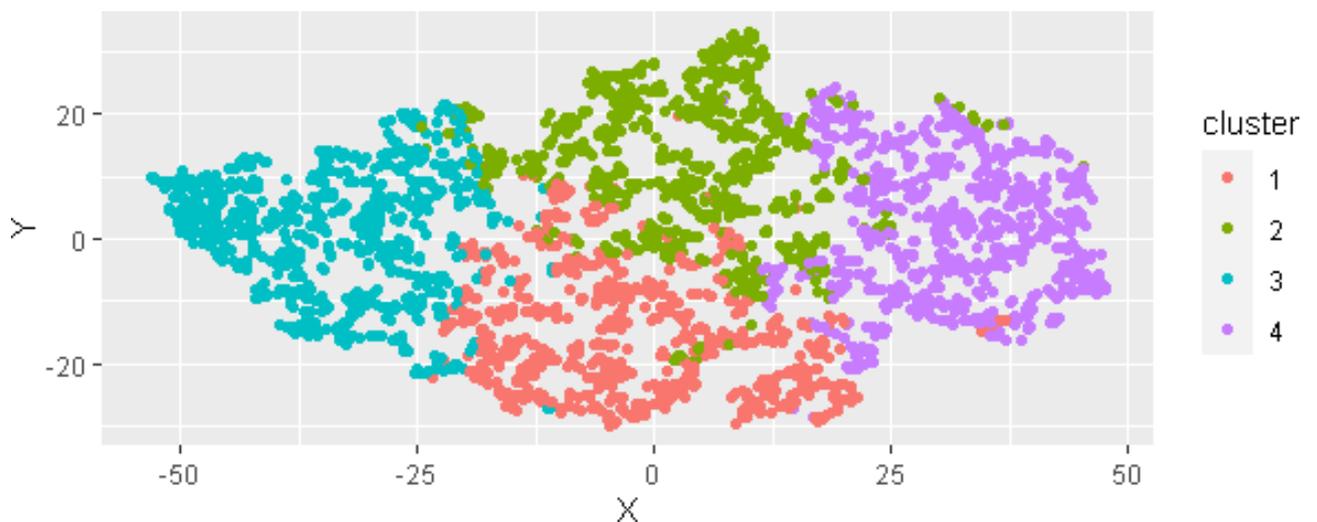
- Méthode du coude :



**Figure 22.** Choix du nombre de classes - Méthode du coude

### V.5.2 Résultats de la classification PAM

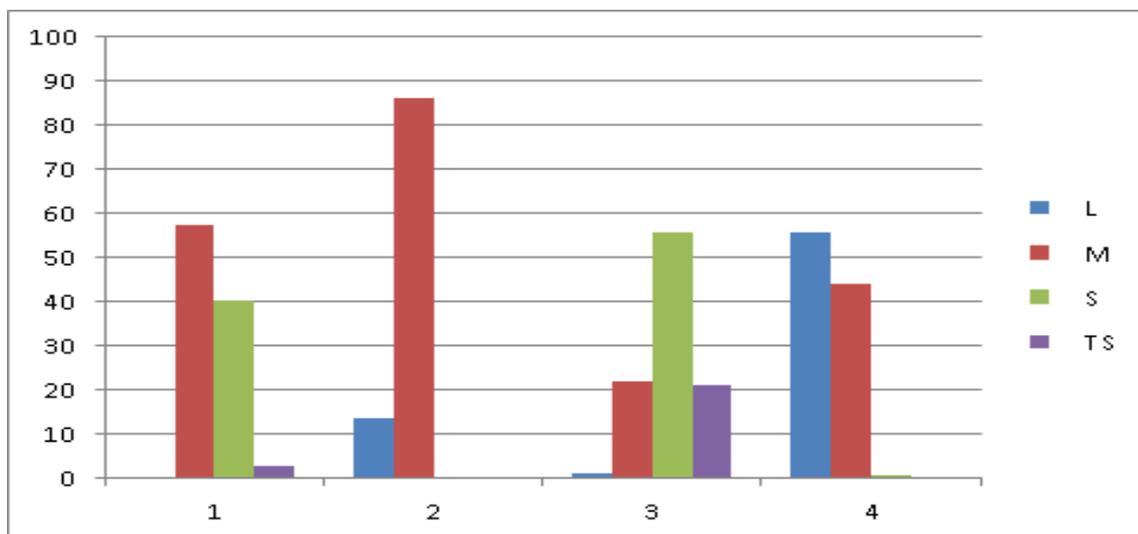
Une projection du nuage des individus (**Figure 26**), suivant l'appartenance aux classes retrouvées, montre un résultat satisfaisant au sens des critères de la qualité de la classification (i.e. "compacité" et "séparation").



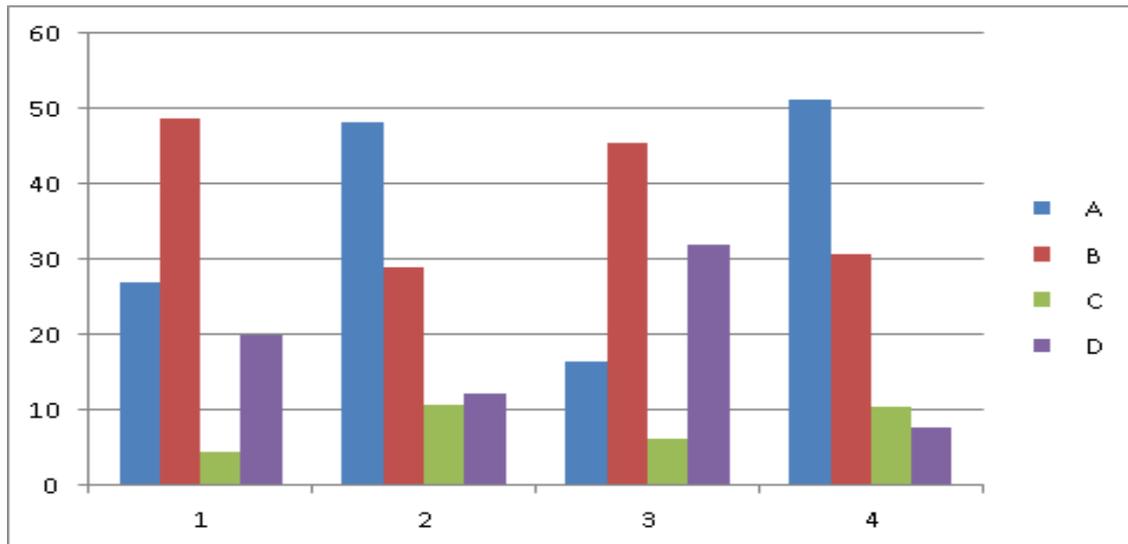
**Figure 23.** Distribution des patients par rapport au PAM

Les histogrammes (**Figures 27 et 28**) montrent une meilleure distinction des groupes identifiés par PAM par rapport à la sévérité, que la méthode k-Means. En effet, nous retrouvons un groupe 3 avec une fréquence élevée pour les stades "Sévère" et "Très sévère" de GOLD 2007, ainsi que les stades "B" et "D" suivant GOLD 2017.

Le groupe 1 est un groupe d'individus en stade plutôt "Modéré"/"Sévère" pour les mesures spirométriques, mais très symptomatique au sens de GOLD 2017. Les groupe 2 et 4, quant à eux, ont des fréquences plus importantes des stades "Léger", "Modéré" et "A", contre des fréquences faibles pour les stades sévères.



**Figure 24.** Fréquence des classes GOLD 2007 dans chaque cluster PAM



**Figure 25.** Fréquence des classes GOLD 2007 dans chaque cluster de PAM

**Les tableaux (17 et 18)** présentés ci-dessous, décrivent les 2968 patients selon les résultats de la classification en k-Médoïdes. L'ensemble des variables cliniques et comorbidités sont significativement corrélées avec les groupes obtenus, à l'exception de l'asthme, le cancer (tout type), l'athérome, l'AVC, le syndrome métabolique et les maladies thromboemboliques.

Les résultats du PAM se distinguent des autres méthodes (k-means et HCPC) par les proportions quasi-uniformes des groupes obtenus (24.22%, 23.59%, 25.57% et 26.24% respectivement).

Le groupe 3 est le plus sévère au sens des mesures spirométriques, suivi du groupe 1. Ils se caractérisent, en effet, par un taux de mortalité très grand par rapport aux autres groupes, ainsi qu'une plus grande sévérité au sens de l'évaluation symptomatique. Le groupe 1, en particulier, se distingue par une population plus âgée et un IMC plus grand que les autres classes. Le groupe 3, quant à lui, est le moins obèse.

Concernant les comorbidités, le groupe 1 est plus marqué par les comorbidités cardiovasculaires ainsi que le diabète. De l'autre côté, le groupe 3 est plutôt caractérisé par des comorbidités liées à son faible IMC (i.e. Dénutrition, Problèmes digestifs, ostéoporose et ostéodensitométrie)

Pour les groupes les moins sévères, au sens des mesures spirométriques, nous retrouvons les groupes 2 et 4, ils ont des taux plus faibles de mortalité ainsi que d'évaluations symptomatiques. De plus, nous retrouvons des taux plus faibles de comorbidités relativement aux groupes plus sévères.

Tableau 17. Description des variables cliniques vis-à-vis des groupes obtenus selon PAM.

Variables cliniques	PAM				p-value
	1   N=719 24.22%	2   N=711 23.95%	3   N=759 25.57%	4   N=779 26.24%	
<b>Hommes, 1892 (63.74%)</b>	527 (73.3)	372 (52.3)	474 (62.5)	519 (66.6)	<0.001***
<b>Âge, 70.07 (10.89)</b>	77.3 (9)	65.7 (10.2)	68.0 (9.5)	69.4 (10.9)	<0.001***
<b>Décès, 335 (11.30%)</b>	114 (15.9)	61 (8.5)	120 (15.8)	40 (5.13)	<0.001***
<b>IMC (kg/m), 26.30 (5.84)</b>	27.1 (5.8)	26.1 (5.7)	25.0 (6)	26.9 (5.5)	<0.001***
<b>Tabagisme, N (%) :</b>					
<b>Jamais, 148 (5%)</b>	51 (7.09)	26 (3.6)	38 (5)	33 (4.24)	<0.001***
<b>Ancien Fumeur, 1671 (56.30%)</b>	505 (70.2)	332 (46.7)	402 (53)	432 (55.5)	
<b>Fumeur Actif, 1149 (38.70%)</b>	163 (22.7)	353 (49.6)	319 (42)	314 (40.3)	
<b>Toux, 1722 (58.02%)</b>	406 (56.5)	434 (61)	472 (62.2)	410 (52.6)	0.008**
<b>Expectoration, 1251 (42.15%)</b>	313 (43.5)	281 (39.5)	354 (46.6)	303 (38.9)	0.007**
<b>mMRC, N (%) :</b>					
<b>0-1   1295 (42.15%)</b>	225 (31.3)	419 (58.9)	171 (22.5)	480 (61.6)	<0.001***
<b>≥2   1673 (56.36%)</b>	494 (68.7)	292 (41.1)	588 (77.5)	299 (38.4)	
<b>Exacerbations, N (%) :</b>					
<b>0-1   2198 (74.05%)</b>	543 (75.5)	548 (77.1)	469 (61.8)	638 (81.9)	<0.001***
<b>≥2   770 (25.95%)</b>	176 (24.5)	163 (22.9)	290 (38.2)	141 (18.1)	
<b>VR/CPT, 137.17 (36.73)</b>	133 (13.9)	135 (11)	177 (43.9)	103 (15)	0.00***
<b>VEMS, 61.31 (20.08)</b>	50.9 (9)	71.0 (9.5)	40.8(13.3)	82.0 (13.4)	0.00***

Tableau 18. Description des comorbidités vis-à-vis des groupes obtenus selon PAM.

Comorbidités	PAM				p-value
	1   N=719 24.22%	2   N=711 23.95%	3   N=759 25.57%	4   N=779 26.24%	
<b>SAS, 490 (16.51%)</b>	120(16.7)	120(16.9)	68 (8.9)	182 (23.4)	<0.001***
<b>Asthme, 324 (11%)</b>	78 (10.8)	82 (11.5)	82 (10.8)	82 (10.5)	0.937
<b>DDB, 145 (4.88%)</b>	50 (6.9)	23 (3.23)	43 (5.6)	29 (3.7)	0.003**
<b>Cancer Poumon, 204 (6.87%)</b>	57 (7.9)	53 (7.45)	43 (5.6)	51 (6.5)	0.323
<b>Autre Cancer, 351 (11.82%)</b>	89 (12.4)	80 (11.3)	74 (9.7)	108 (13.9)	0.083
<b>Hypertension, 1090 (36.72%)</b>	330 (45.9)	230 (32.3)	262 (34.5)	268 (34.4)	<0.001***
<b>Cardiopathie Ischémique, 502 (17%)</b>	163 (22.7)	97 (13.6)	110 (14.5)	132 (16.9)	<0.001***
<b>Trouble du rythme, 346 (11.65%)</b>	135 (18.8)	54 (7.6)	65 (8.5)	92 (11.8)	<0.001***
<b>Insuffisance Cardiaque Gauche, 121 (4.07%)</b>	50 (6.9)	16 (2.2)	39 (5.1)	16 (2.05)	<0.001***
<b>Athérome, 170 (5.72%)</b>	55 (7.6)	34 (4.8)	43 (5.6)	38 (4.8)	0.067
<b>AVC, 99 (3.33%)</b>	23 (3.2)	19 (2.6)	26 (3.4)	31 (3.9)	0.566
<b>HTAP, 64 (2.15%)</b>	23 (3.2)	3 (0.4)	31 (4)	7 (0.9)	<0.001***
<b>Artériopathie, 190 (6.40%)</b>	68 (9.4)	38 (5.3)	54 (7.1)	30 (3.8)	<0.001***
<b>Diabète, 367 (12.36%)</b>	117 (16.3)	73 (10.3)	86 (11.3)	91 (11.7)	0.003**
<b>Dyslipidémie, 673 (22.67%)</b>	196 (27.3)	133 (18.7)	133 (17.5)	211 (27.1)	<0.001***
<b>Syndrome Métabolique, 105 (3.53%)</b>	32 (4.45)	19 (2.6)	26 (3.4)	28 (3.6)	0.341
<b>Dénutrition, 59 (2%)</b>	15 (2.09)	8 (1.1)	33 (4.3)	3 (0.4)	<0.001***
<b>Manifestations Sinusiennes, 160 (5.40%)</b>	34 (4.73)	32 (4.5)	35 (4.6)	59 (7.5)	0.019*
<b>Problèmes Digestifs, 75 (2.52%)</b>	15 (2.09)	14 (1.9)	30 (3.9)	16 (2)	0.038*
<b>Maladie Thromboembolique, 36 (1.21%)</b>	7 (0.97)	12 (1.7)	9 (1.2)	8 (1)	0.591
<b>Ostéoporose, 97 (3.26%)</b>	31 (4.31)	11 (1.5)	38 (5)	17 (2.2)	<0.001***
<b>Ostéodensitométrie, 97 (3.26%)</b>	31 (4.31)	11 (1.5)	38 (5)	17 (2.2)	<0.001***
<b>Anxiété, 430 (14.48%)</b>	104 (14.5)	102 (14.3)	152 (20)	72 (9.2)	<0.001***
<b>Dépression, 382 (12.87%)</b>	60 (8.3)	105 (14.8)	120(15.8)	97 (12.5)	0.027*

### V.5.3 Validation de la classification PAM (vs. HCPC et k-means)

Nous avons utilisé trois indice de validation interne pour la classification PAM (Vs. HCPC) et (Vs. K-means) : (**Voir Tableau 19**)

La connectivité correspond à la mesure dans laquelle les éléments sont placés dans le même cluster que leurs voisins les plus proches, dans l'espace de données. La connectivité a une valeur comprise entre 0 et l'infini et doit être minimisée.

L'indice de Dunn est le rapport entre la plus petite distance intergroupe et la plus grande distance intra-groupe. Il a une valeur comprise entre 0 et l'infini et doit être maximisé.

L'indice de silhouette mesure la qualité d'un partitionnement en s'appuyant sur la similitude d'une observation avec son propre cluster (cohésion), par rapport aux autres clusters (séparation). Sa valeur est comprise entre -1 et 1 et doit être maximisée.

**Tableau 19.** Score de la meilleure méthode de classification par rapport à l'indice de choix de validation interne.

<b>Indice</b>	<b>Score</b>	<b>Méthode</b>
<b>Connectivity</b>	617.72	PAM
<b>Dunn</b>	0.0197	PAM
<b>Silhouette</b>	0.143	k-Means

## **DISCUSSION GÉNÉRALE ET CONCLUSION**

À ce jour, la littérature portant sur les phénotypes et les trajectoires des patients BPCO se distingue par une grande hétérogénéité, aussi bien dans les méthodes utilisées (modalités utilisées pour identifier les phénotypes et les modèles considérées) que dans les résultats rapportés, à tel point qu'il est difficile d'en dégager de réelles certitudes.

Ce travail a consisté, dans un premier temps, à l'identification de profils de comorbidités par l'analyse en clusters, afin d'explorer l'histoire naturelle des comorbidités au cours de la BPCO, encore mal cernée. Dans un second temps, l'identification des phénotypes d'exacerbations par une classification non-supervisée a été étudiée. Ensuite, l'analyse des déterminants de la dyspnée par la régression pénalisée et l'analyse de sensibilité a été étudiée. Finalement, l'analyse de mortalité et les phénotypes cliniques des patients BPCO identifiés par l'analyse en cluster la plus optimale a été étudiée, afin d'éclaircir les conclusions discordantes actuellement retrouvées dans les travaux épidémiologiques sur le sujet.

## **Points forts et points faibles de la méthodologie**

La méthodologie utilisée afin de répondre aux objectifs de la thèse comporte des points forts, mais aussi quelques limites. Dans cette partie, seront discutées les forces et les faiblesses des différents aspects méthodologiques de ce travail, concernant les patients de la cohorte PALOMB, l'identification des profils des comorbidités, d'exacerbations, les déterminants de la dyspnée et enfin l'analyse de mortalité et les phénotypes cliniques.

## **PALOMB : Les caractéristiques sont conformes aux connaissances**

Les caractéristiques générales des patients de notre cohorte sont conformes aux connaissances de la maladie en ce qui concerne l'âge, le sexe, et l'exposition aux facteurs de risque tels que le tabac. En effet, notre cohorte est composée de 2/3 d'hommes et 1/3 de femmes, les premiers symptômes débutent principalement après 50 ans. Le tabagisme est un facteur de risque considérable puisqu'il est présent chez 95% de la population totale avec une consommation importante (plus de la moitié des fumeurs ont un tabagisme cumulé supérieur à 30 PA) et un tabagisme actif encore élevé (chez 36% de la population). Les principales classes de sévérité suivies en consultation de pneumologie sont les stades modérés (54%) et sévères (26%). Sur le plan clinique, les patients BPCO suivis en consultation par les pneumologues sont symptomatiques quel que soit le degré de sévérité de la maladie avec toutefois des symptômes plus fréquents et plus sévères pour les stades

avancés. Les comorbidités associées à la BPCO sont fréquentes, particulièrement en ce qui concerne les pathologies cardio-vasculaires et les pathologies anxio-dépressives.

## L'identification des profils des comorbidités

Cette étude avait pour objectif de d'étudier l'association entre les comorbidités et la sévérité de la BPCO dans une cohorte clinique de patients BPCO. Il s'agit principalement des patients BPCO suivis par des pneumologues. Seulement 28,4% des patients n'avaient pas de comorbidités associées à la BPCO. Quatorze comorbidités étaient significativement différentes entre les différents stades de sévérité de la BPCO. Dans cette grande population de patients BPCO, la médiane du nombre de comorbidités était de deux.

L'hypertension, le SAOS, la dyslipidémie, la cardiopathie ischémique et les troubles du rythme cardiaque étaient plus fréquents chez les patients des groupes B et D que dans les autres groupes, tout comme l'anxiété et la dépression. La dénutrition était la plus fréquente chez les patients GOLD D et l'ostéoporose était la plus fréquente chez les patients GOLD B et D. Le nombre de comorbidités était plus élevé chez les patients GOLD B et D. Même lorsque la sévérité des symptômes était similaire, la prise en charge de la BPCO semblait être différente selon que les patients présentaient ou non des comorbidités. Enfin, cinq profils de comorbidités ont été identifiés, le plus fréquent étant celui des maladies cardiovasculaires et du syndrome d'apnée obstructive.

Notre méthodologie est d'analyser la relation entre les comorbidités et la sévérité de la BPCO en utilisant trois approches différentes : l'impact du nombre de comorbidités, l'association univariée entre les comorbidités et la sévérité de la BPCO et l'analyse en clusters pour déterminer l'association entre les comorbidités, c'est-à-dire pour établir l'existence de différents phénotypes. Nos résultats sont conformes aux publications précédentes qui faisaient état d'une forte prévalence de comorbidité dans la BPCO, en particulier les maladies cardio-vasculaires. Une étude [105] a montré que, comparativement à la population des patients non atteinte de BPCO, les patients BPCO étaient plus susceptibles de recevoir un diagnostic de maladie cardiovasculaire (OR [95 % IC] : 2,4 [2,02-3] ;  $p < 0-0001$ ), y compris un risque deux à cinq fois plus élevé de cardiopathie ischémique, du trouble de rythme cardiaque, d'insuffisance cardiaque et de maladies artérielles. En outre, les patients atteints de BPCO ont déclaré plus souvent souffrir d'hypertension (OR [95 % IC] : 1,3 [1,1 – 1,5] ;  $p = 0-0007$ ), de diabète (OR [95 % IC] 1,3 [1,2 – 1,5] ;  $p < 0-0001$ ) et avoir déjà fumé (OR [95 % IC] : 4,2 [3,2 – 5,6] ;  $p < 0-0001$ ). Divo et al. [106] ont constaté dans leur cohorte que les maladies cardiovasculaires étaient fortement associées au risque de mortalité, mais que le risque de mortalité le plus élevé était associé à l'anxiété. Cependant, nous avons constaté une prévalence élevée du syndrome d'apnée obstructive du sommeil, probablement due au syndrome de chevauchement tel que rapporté par Soler et al [107]. En revanche, le syndrome d'apnée obstructive du sommeil est la comorbidité la plus fréquente chez les patients GOLD A et B, bien qu'il semble être associé à une BPCO modérée à sévère. Il est essentiel de diagnostiquer le syndrome de chevauchement chez les patients atteints de BPCO, car les patients atteints du syndrome de chevauchement qui ne sont pas traités par La ventilation à pression positive continue (CPAP) ont une mortalité plus élevée [108].

L'analyse en clusters a révélé cinq profils de comorbidités : Le groupe 1 (profil cardiaque) ; le groupe 2 (comprenait moins de comorbidités) ; le groupe 3 (le syndrome métabolique, SAS et l'anxiété-dépression) ; le groupe 4 (la dénutrition et l'ostéoporose) et le groupe 5 (la broncheectasie). Vanfleteren a trouvé 13 comorbidités dans un échantillon de 213 patients atteints de BPCO [109] avec cinq phénotypes de comorbidité : moins de comorbidité, cardiovasculaire, cachectique, métabolique et psychologique. Quatre de nos groupes sont concordants, à savoir les comorbidités cardiovasculaires, cachectiques, métaboliques et moins comorbidités. Néanmoins, tous les groupes ont été associés de manière plus significative à GOLD D et de manière moins significative à GOLD B. Cette découverte pourrait expliquer le risque plus élevé de mortalité chez les patients GOLD B et D, comme cela a déjà été signalé ailleurs [110]. De la même manière, Divo et ses collaborateurs [111] ont également identifié un certain nombre de clusters dans le réseau des comorbidités, dont un groupe cardiovasculaire et un autre groupe caractérisé par une limitation légère à modérée du débit d'air du poumon et un syndrome métabolique avec un IMC élevé, ces deux clusters concordant avec nos résultats.

Nos résultats montrent que même si l'étude de l'impact des comorbidités au cours de la BPCO est une question complexe, les comorbidités contribuent de manière importante à la sévérité clinique de nos patients, et que au même niveau d'obstruction bronchique, la prise en charge de la BPCO différerait selon qu'ils avaient ou non des comorbidités.

Notre étude comporte quelques limites. Premièrement, les comorbidités ont été enregistrées par des pneumologues ; des études antérieures ont montré que les comorbidités sont sous-diagnostiquées dans la vraie vie. Cela pourrait également être le cas dans notre étude pour la plupart des comorbidités, à l'exception du SAOS, car le SAOS a été diagnostiqué par une polysomnographie réalisée par le même pneumologue qui a diagnostiqué la BPCO. Cependant, nous pensons que ce biais est limité dans notre étude, car nous avons constaté une corrélation significative entre les comorbidités déclarées et les traitements administrés pour celles-ci. Deuxièmement, nous ne pouvons pas généraliser ces résultats aux patients BPCO dans la population générale, car notre population était gérée à la fois par un médecin généraliste et un pneumologue. Troisièmement, nous avons effectué une analyse exploratoire multivariée afin de mieux décrire les associations entre les différentes comorbidités. Ce type d'analyse utilise une méthode statistique qui traite une grande quantité d'informations provenant de variables hétérogènes dans des groupes homogènes. Il est bien connu que différents facteurs peuvent influencer l'analyse et donc les résultats : le choix des méthodes, hiérarchiques ou non, la détermination du nombre de clusters bien avant l'analyse, le choix des variables incluses dans l'analyse, la corrélation entre les variables sélectionnées et le jugement clinique des chercheurs. Pour limiter l'impact de la corrélation spécifique entre les variables, nous avons d'abord effectué une analyse en composantes principales. Ensuite, nous avons utilisé le graphe en éboulis de la valeur propre, le critère de Kaiser-Guttman et le pourcentage de variance expliqué pour déterminer le nombre de clusters.

## L'identification des profils d'exacerbations

Nous avons examiné la fréquence des exacerbations chez 932 patients BPCO légère, modérée, sévère et très sévère. Il s'agissait pour la plupart d'hommes qui étaient des fumeurs actuels atteints d'une BPCO symptomatique associée à des comorbidités fréquentes. 27,04% des 932 patients ont déclaré deux exacerbations ou plus au cours des 12 mois précédents. Plusieurs variables ont montré une association significative avec la fréquence des exacerbations : La toux chronique, l'expectoration chronique, les fumeurs actuels, la dyspnée et la distension thoracique. La méthode K<sub>m</sub>L a identifié quatre phénotypes de patients sur la base d'un suivi longitudinal de 3 ans de la fréquence des exacerbations : le phénotype A (exacerbateurs non fréquents), C (transitoires), B et D (exacerbateurs fréquents). Après ajustement sur les variables de confusion, l'analyse multivariée a montré que les patients qui avaient des exacerbations fréquentes l'année précédant leur inclusion (HR [IC95 %] : 3,72 [2,53–5,49]), une anxiété (HR [IC95 %] : 2,03 [1,24–3,31]) et l'absence de vaccination antigrippale annuelle (HR [IC95 %] : 1,97[1,20–3,24]) restent associées aux phénotypes d'exacerbateurs fréquents.

À la fin de notre étude, seulement 3,13% des patients BPCO avaient deux exacerbations ou plus, contre 2% dans l'étude SPIROMICS, 12% dans la cohorte ECLIPSE et 7% dans la cohorte française (EXACO) [112-114]. Nous avons considéré tous les stades de sévérité de la BPCO, contrairement à la forme sévère de la BPCO des patients des études mentionnées ci-dessus, ce qui explique les différences de résultats.

Entre les années 1 et 2, 85% des patients sont passés d'exacerbateurs fréquents à des exacerbateurs peu fréquents (exacerbation 0-1), tandis que 30% sont passés d'exacerbateurs peu fréquents à des exacerbateurs fréquents. Cette variabilité intra-sujet de la fréquence annuelle des exacerbations limite la capacité de ces critères à prédire le risque futur d'exacerbations, au niveau individuel. Il est possible qu'une période de suivi supérieure à trois ans réduise la variabilité du taux d'exacerbation dans le temps, ce qui permettrait au modèle d'avoir une meilleure efficacité prédictive.

Les vaccins antipneumococcique et antigrippaux annuels sont souvent utilisés pour traiter les exacerbations de la BPCO, la grippe étant une cause fréquente de ces exacerbations : Une étude récente a montré que les virus de la grippe sont détectés dans 2,50 % à 11,60 % des exacerbations de la BPCO [115], ce qui est conforme aux rapports publiés précédemment qui font état de la détection de la grippe chez 5 à 22 % des patients hospitalisés pour des exacerbations [116 - 119]. La vaccination contre la grippe est recommandée pour tous les patients atteints de BPCO par l'Organisation mondiale de la santé (OMS), les Centres « for Disease Control and Prevention » des États-Unis, le Centre européen de contrôle et de prévention des maladies (ECDC), de nombreuses agences nationales [120-122] et le Comité consultatif sur les pratiques de vaccination (ACIP) aux États-Unis [123].

L'originalité de notre étude était d'analyser l'impact de la vaccination annuelle contre la grippe, le statut de la vaccination antipneumococcique et de tous les traitements inhalés de la BPCO sur

l'analyse longitudinale de la fréquence des exacerbations. Nous avons pu montrer une forte association entre les patients non vaccinés contre la grippe et les phénotypes exacerbeurs fréquents. Malgré les recommandations de longue date visant à vacciner les patients vulnérables à haut risque contre la grippe saisonnière, les taux de vaccination restent sous-optimaux dans cette population.

Notre étude a identifié deux phénotypes exacerbeurs fréquents stables qui différaient en fonction de l'IMC. Le phénotype D (fréquent chez les patients obèses) était davantage associé à l'anxiété. Par conséquent, chez les patients atteints de BPCO présentant des exacerbations fréquentes, une gestion adéquate de l'anxiété doit être accordée à l'importance qu'elle mérite. L'éducation des patients sur la vaccination contre la grippe pourrait contribuer à réduire la fréquence des exacerbations.

En outre, des études ont démontré l'impact de la vaccination contre la coqueluche sur la réduction de la fréquence des exacerbations et pourraient constituer une considération intéressante pour de futures études.

Notre étude présente plusieurs points forts et certaines limites : 486 patients perdus de vue étaient des patients asymptomatiques. Notre étude comporte un biais de sélection mais elle reflète la pathologie compliquée de la BPCO. Ces résultats confirment que les patients atteints de BPCO pris en charge par un pneumologue sont plus sévère que ceux pris en charge en soins primaires. Comme les patients BPCO plus sévères étaient suivis par des pneumologues, nous ne pouvons pas exclure formellement un biais de sélection avec surestimation du phénotype d'exacerbation fréquent.

Nous avons trouvé quatre phénotypes avec deux phénotypes d'exacerbation fréquente avec une différence basée sur l'IMC. L'adhésion au traitement de la BPCO a été incluse dans notre analyse, ce qui n'avait pas été pris en compte auparavant.

Malgré ses limites, cette étude est l'une des rares à avoir étudié l'association entre les exacerbations fréquentes et les facteurs de risque chez les patients atteints de BPCO, en tenant compte de la prise en charge thérapeutique : une personne non vaccinée contre la grippe est un déterminant important des exacerbations fréquentes et la comorbidité de l'anxiété doit être traitée de manière adéquate.

## Les déterminants de la dyspnée (échelle mMRC)

Le résultat le plus intéressant de cette étude est l'identification de la relation spécifique entre la distension thoracique statique (définie par le rapport CI/CPT) et l'échelle mMRC de la dyspnée. En effet, au même degré de sévérité de la BPCO, la majorité des patients ayant une distension thoracique présentaient une dyspnée plus élevée que ceux dont l'hyperinflation était plus faible. La dyspnée est décrite comme une perception physique d'une respiration difficile avec différents degrés de sévérité [124], la dyspnée (échelle mMRC) est un paramètre largement utilisé pour l'évaluation de la sévérité de la BPCO [125]. GOLD a reconnu la nécessité d'une évaluation plus rigoureuse des symptômes, plutôt

qu'une mesure rapide et facile de l'essoufflement [126]. Cependant, de nombreuses études ont souligné l'importance de la dyspnée chez les patients atteints de BPCO, comme l'étude espagnole sur la BPCO [127]. Il existe un désaccord considérable sur la méthodologie appropriée pour mesurer la dyspnée. Nous avons opté pour l'échelle mMRC car il s'agit d'une mesure quantitative unidimensionnelle de la sévérité de la dyspnée, utilisée par les pneumologues du monde entier.

L'hyperinflation statique est associée à la dyspnée [128] et peut également être un prédicteur de mortalité lorsqu'elle est exprimée par le rapport CI/CPT [128]. La VR augmente en raison d'une limitation extrême du débit expiratoire au niveau du volume pulmonaire. Au fur et à mesure que la BPCO progresse, l'augmentation du VR est plus importante que l'augmentation de la CPT (rapport VR/CPT plus élevé), ce qui entraîne une diminution de la CV (capacité vitale) et une baisse conséquente du VEMS. Cependant, les valeurs seuils d'hyperinflation statique n'ont jamais été normalisées. Deux études [129, 130] ont suggéré que l'hyperinflation devrait être définie comme une augmentation de la CRF (capacité résiduelle fonctionnelle) plutôt qu'une augmentation du VR ou du rapport VR/CPT [131], mais ces études [129, 130] utilisent la VR pour analyser les différents traitements de l'emphysème. De nombreuses recherches [132, 133] se sont concentrées sur l'utilisation de la CI en pratique clinique pour évaluer l'hyperinflation statique. Malheureusement, nous n'avons pas pu analyser la CRF, la CI ou le rapport CI/CPT en raison de leurs limites dans la pratique clinique et de leurs valeurs manquantes dans nos données. C'est pourquoi nous avons opté pour un sous-groupe (n=421) de patients BPCO qui n'avaient pas de valeurs manquantes de CI et de CRF, pour étudier les corrélations entre l'échelle mMRC et le CRF, le rapport CRF/CPT ou CI/CPT, afin d'évaluer les effets d'autres indicateurs de l'hyperinflation par l'échelle mMRC dans le sous-groupe de 421 patients définis par les variables mentionnées précédemment. Le VEMS OR=3,47 [1,96-6,12] ; les exacerbations ( $\geq 2$ ) OR=2,31 [1,33- 4,17] et l'hyperinflation (IC/TLC) OR=0,57 [0,35-0,85] ont été associées à une dyspnée plus élevée (mMRC $\geq 2$ ) chez les patients du sous-groupe.

Notre étude présente plusieurs points forts et certaines limites. Tout d'abord, la classification GOLD serait différente selon l'échelle des symptômes, CAT (COPD Assessment Test) vs mMRC, mais dans notre cohorte, nous nous sommes concentrés sur l'échelle mMRC en développant une analyse de sensibilité. Il est bien fondé que de multiples considérations, telles que le choix des variables, l'association entre les facteurs choisis et le jugement scientifique des enquêteurs peuvent influencer l'analyse et donc les résultats. Pour limiter l'impact de la corrélation spécifique entre les variables, nous avons d'abord effectué une analyse croisée approfondie de différents facteurs ; les variables indépendantes incluses dans le modèle étaient également celles qui présentaient un  $P < 0,05$  en analyse bivariée. Enfin, la plupart des patients PALOMB sont couverts par l'assurance maladie, ce qui signifie qu'il existe une population non reconnue de patients non assurés qui ont probablement une prévalence plus élevée de la dyspnée.

Malgré ses limites, il s'agit d'une des rares études qui ont étudié l'association entre une dyspnée plus élevée et plusieurs facteurs de risque chez les patients atteints de BPCO.

## L'analyse de mortalité et l'identification des phénotypes cliniques

Les comorbidités, en particulier le cancer du poumon et les maladies cardiovasculaires, influencent considérablement le pronostic vital des patients atteints de BPCO. Plusieurs caractéristiques cliniques et fonctionnelles ont également été associées à la mortalité : sexe masculin, âge avancé, dyspnée, exacerbations fréquentes, BOD plus élevée et une fonction pulmonaire moins bonne. Nous avons ici une prise en charge de la morbidité cardiovasculaire plus faible que ce qui avait été signalé précédemment, peut-être en raison d'une représentation masculine plus faible [134, 135], même si la comorbidité de l'insuffisance cardiaque augmente le risque de mortalité toutes causes confondues des patients atteints de BPCO [134], en particulier chez les hommes.

Dans cette étude, nous avons également confirmé l'effet bénéfique de l'arrêt du tabac sur la mortalité à long terme [136]. Il s'agit donc d'une nouvelle preuve pour promouvoir un programme spécifique de sevrage tabagique afin d'augmenter le sevrage tabagique à long terme chez les patients atteints de BPCO.

La baisse du VEMS est généralement liée à une fréquence plus élevée d'exacerbations de la BPCO [137], qui est également associée à un pronostic plus défavorable. Cependant, dans la cohorte PALOMB, la différence en termes de pronostic n'est pas significative. Nous pouvons mettre en évidence plusieurs explications. Premièrement, la baisse moyenne du VEMS dans cette cohorte (-4,96mL/an) est inférieure aux attentes, car elle se situe généralement entre 30mL/an et 65mL/an dans les grandes études sur la BPCO [138, 139]. Deuxièmement, les sujets sont moins sévères puisque cette cohorte a été principalement recrutée par des médecins respiratoires de centres secondaires. Enfin, les données relatives à la baisse du VEMS n'étaient disponibles que pour 770 patients de la cohorte entière.

Nous avons également constaté que le fait d'être un homme était associé à un pronostic plus défavorable, que ce soit après ajustement en fonction de l'âge ou du score de BOD. Cependant, plusieurs études ont rapporté que le risque de décès était similaire chez les hommes et les femmes une fois que les analyses étaient ajustées pour tenir compte des différences dans les facteurs de confusion de base [134, 140, 141]. Une mortalité plus élevée chez les hommes pourrait être due à des maladies cardio-vasculaires mieux diagnostiquées (54,8 contre 36,2%,  $p < 0,0001$ ). En outre, les maladies cardiovasculaires étaient également plus fréquentes chez les hommes décédés que chez les femmes décédées.

Une des limites de cette étude est l'analyse de la survie à relativement court terme et l'absence de données disponibles concernant des causes spécifiques de mortalité, ce qui empêche d'analyser si certains taux de mortalité spécifiques diffèrent entre les hommes et les femmes.

Dans le cadre de cette étude, nous avons effectué différentes méthodes de classification (supervisée et non supervisée) sur les données de la cohorte PALOMB, afin d'identifier des phénotypes caractérisant les différents groupes. Ces méthodes ont été basées sur les variables

cliniques, les caractéristiques respiratoires et les comorbidités, notamment les maladies cardiovasculaires et le diabète.

La majorité des patients BPCO de la cohorte présentent au moins deux comorbidités (57.68%). Ces patients sont plus majoritaires dans les deux stades sévères de GOLD 2007 ("S", "TS" et les deux classes sévères de GOLD 2017 "B", "D"). En particulier, ils sont plus caractérisés par la présence des comorbidités cardiovasculaires, le diabète ainsi que les maladies métaboliques.

36.72% des patients présentent une HTA, 22.67% une dyslipidémie, 17% une cardiopathie ischémique et 12.36% un diabète. 16.51% des patients présentent un syndrome d'apnées du sommeil (dont 95% de type obstructif). Enfin, l'anxiété est présente chez 14.48% des patients et la dépression 12.87% des patients. L'HTA, le SAS, la dépression et l'anxiété sont significativement plus présents dans les stades modérés et très sévères. L'asthme est associé à la BPCO dans 11% des cas et sa répartition, selon la sévérité de la maladie, ne diffère pas selon le stade de sévérité de la maladie ( $p=0.623$ ). Le cancer du poumon est associé à la BPCO dès le stade léger et 6,87 % des patients de la population totale ont un cancer du poumon et 11.82% ont un autre type de cancer.

17% des patients présentent une dénutrition, 2.52% des problèmes digestifs, 1.21% une maladie thromboembolique, 3.26% une ostéoporose et une ostéodensitométrie. Ces quatre types de comorbidités ont montré des différences significatives vis à vis la classification GOLD 2007 et GOLD 2017.

Nous avons pu démontrer (selon les différentes méthodes employées) l'existence des clusters de patients français atteints de BPCO et nous avons évalué leur association avec les paramètres cliniques. Sur la base de notre hypothèse, nous avons pu déterminer quatre phénotypes selon la méthode de classification supervisée (HCPC) : le phénotype 1 (52.83%), 2 (10.85%), 3 (4.88%) et 4 (31.43%). En ce qui concerne la classification non supervisée, nous avons pu identifier quatre phénotypes selon la méthode de classification non supervisée (k-Means) : le phénotype 1 (27.20%), 2 (40.37%), 3 (22.17%) et 4 (9.9%). Et 4 autres phénotypes selon (PAM) : le phénotype 1 (24.22%), 2 (23.95%), 3 (25.57%) et 4 (26.24%). Les méthodes de classification (partitionnement) servent à délimiter des groupes de patients, ou typologies, à partir de leurs caractéristiques. D'une part, nous avons pu identifier un groupe de patients (Phénotype 1 - (52.83%), dit "groupe de référence"), au sens de HCPC, qui est caractérisé par : une similarité en terme de sexe (49,2% hommes et 50,8% femmes), des patients jeunes, un taux de mortalité moyen, un IMC faible, une proportion élevée de fumeurs actifs, moins de patients dyspnéiques, plus de distension thoracique (VR/CPT) et une sévérité modérée. A contrario, les phénotypes 3 (4.88%) et 4 (31.43%), au sens de HCPC, étaient plus sévères. Enfin le classe 2 (10.85%) était "transitoire". D'autre part, nous avons pu identifier, grâce au k-Means, un groupe de patients très sévère (classe 4 - 9.9%), au sens des mesures de spirométrie, qui est caractérisé par : plus de patients hommes (68.7%), un taux de mortalité très élevé, plus de distension thoracique (VR/CPT), plus d'exacerbations fréquentes et de patients dyspnéiques. Un autre groupe de patients sévère a été identifié (classe 1 - 27.20%). Il enregistre un grand taux de mortalité aussi et classé deuxième

en terme de sévérité spirométrique et symptomatique ainsi qu'en terme de comorbidités cardiovasculaires et de diabète.

En ce qui concerne le partitionnement autour des médoïdes (PAM), nous avons pu avoir une meilleure distinction des groupes par rapport aux évaluations de la sévérité et aux comorbidités. De plus, les groupes obtenus avaient des proportions plus uniformes que les méthodes précédentes. Cela a permis d'identifier un premier groupe très sévère (Groupe 3 : 25.57%), au sens des mesures spirométriques, caractérisé par un taux de mortalité élevé, un IMC plus faible ainsi que des proportions et des fréquences de symptômes plus élevées. En terme de comorbidités, ce groupe enregistre de grand taux de comorbidités métaboliques, ostéo-articulaires et neuropsychiatrique. Un deuxième groupe sévère, en terme de spirométrie, est le groupe 1 (24.22%). Nous y retrouvons un taux de mortalité semblable au groupe 3, mais se distingue par une population plus âgée, plus d'hommes et un IMC plus important. Contrairement au groupe 3, ce groupe est moins symptomatique et enregistre plutôt des taux élevés de comorbidités cardiovasculaires et de diabète.

La qualité des classifications a été examinée en utilisant des indices de validation interne/externe. D'après l'indice de "Connectivité" et de "Dunn", la méthode de classification PAM fut la meilleure en terme des distances intergroupes et intra-groupes. Par ailleurs, l'indice de "Silhouette" pointe vers la méthode "k-Means" comme la meilleure méthode de classification, au sens de la "cohésion" et la "séparation" des classes.

En fonction de l'approche choisie pour la modélisation (paramétrique / non paramétrique) et le type de données étudiées, on se place dans une famille particulière d'algorithmes de clustering dont voici une liste non-exhaustive. On peut tout d'abord compter les méthodes basées sur des centroïdes comme k-means [142] ou k-medoids [143], pour ne citer que les plus connus. Comme leur nom l'indique, elles ont la particularité d'être toutes basées sur une notion de centre de masse, placées aléatoirement à l'initialisation, qui vont idéalement se déplacer graduellement vers les centres de masse réellement présents dans le jeu de données. Ces méthodes, qui sont non-paramétriques, supposent que les groupes sont concentriques, c'est à dire semblables à un amas gaussien.

Les méthodes dites à regroupement hiérarchiques [144] produisent des hiérarchies qui mettent en évidence la force des liens entre les différentes observations ce qui permet ensuite d'en extraire des partitions adéquates. Elles peuvent être ascendantes ou descendantes. Les ascendantes partent de l'état initial où chaque observation définit un cluster puis le nombre de clusters est ensuite réduit de manière itérative en fusionnant les groupes les plus proches jusqu'à arriver à un seul groupe. Ensuite revient le choix de l'expert (ou de l'algorithme lui-même) de choisir quelles sont les meilleures partitions parmi l'ensemble des possibles fournies par la hiérarchie. La représentation de la hiérarchie se fait très souvent sous la forme d'un dendrogramme. Les descendantes suivent le même principe mais partent d'un seul cluster auquel toutes les observations appartiennent initialement pour le subdiviser progressivement. Les méthodes appartenant à cette catégorie diffèrent à la fois sur le choix de la métrique (ou mesure de dissimilarité) et sur la stratégie d'agglomération ou d'élague.

Avec ses 60 ans d'ancienneté, k-moyennes, mais plus souvent appelé k-means [142], est l'un des plus connus et des plus anciens algorithmes de classification automatique non-supervisée. L'initialisation des moyennes joue un rôle important sur les résultats en sortie. Si elles sont mal initialisées, on risque de faire converger l'algorithme sur un minimum local potentiellement

éloigné du minimum global. La manière la plus basique (méthode de Forgy) est d'initialiser les  $k$  moyennes initiales à  $k$  données d'entrée choisies aléatoirement. Des variantes de l'algorithme de base ont donc été proposées comme  $k$ means++ [145] dont la stratégie d'initialisation est également de choisir aléatoirement mais en abaissant les chances de choisir un point proche des centres déjà initialisés, par exemple. Les moyennes peuvent également être remplacées par des médianes [146] ou encore par la somme de fonctions variées comme l'algorithme Fuzzy C-means où l'appartenance n'est plus exclusive mais traduite par un degré d'appartenance. Les distances aux moyennes aussi peuvent être remplacées par une probabilité d'appartenance au groupe qui est supposé gaussien, ce qui donne lieu aux algorithmes de type Espérance-Maximisation EM. Cet algorithme et ses variantes sont très efficaces lorsque les données sont réparties en amas gaussiens mais leur efficacité s'effondre dès que les clusters sont de forme arbitraires, particulièrement pour des clusters non-concentriques. Aussi, le comportement de cet algorithme peut énormément varier selon le paramètre  $k$  qui est bien souvent difficile à déterminer compte tenu du cadre non-supervisé. Cependant la complexité de cet algorithme est exemplaire lorsqu'elle utilise les arbres  $k-d$  puisqu'elle se trouve ainsi égale à  $O(kTN)$  où  $T$  est le nombre d'itérations et  $N$  le nombre d'observations. Ainsi, bien implémentée, cette heuristique se trouve être une des plus rapides ce qui participe en partie à sa renommée et celle de ses variantes.

La méthode des  $k$ -médoides est une méthode de partitionnement qui se base sur des médoides pour créer des partitions (clusters). Un médoïde est l'élément d'un ensemble qui minimise la somme des distances entre lui et chacun des autres éléments de cet ensemble.  $K$  représente le nombre de médoides et de clusters que l'on souhaite. L'algorithme PAM (Partitioning Around Medoid) se décompose en deux parties. Dans un premier temps l'initialisation et dans un deuxième temps la recherche des meilleurs médoides. L'algorithme PAM est celui qui semble donner les meilleurs résultats, comparé à tous les autres algorithmes implémentant la méthode des  $k$ -médoides. C'est un algorithme qui a une complexité relativement grande, et une mauvaise capacité à travailler avec une grande quantité de données. Néanmoins cela reste un algorithme de référence de l'implémentation de la méthode des  $k$ -médoides. Si l'on veut tester la méthode des  $k$ -médoides, on se doit de tester cet algorithme. Les algorithmes CLARA et CLARANS sont des variantes de l'algorithme PAM, qui visent à augmenter la vitesse de l'algorithme en s'autorisant à être un peu moins précis. Par conséquent ces algorithmes donnent des résultats bien moins bons que l'algorithme PAM. Nous n'avons pas testé ces deux variantes car nous ne pensons pas qu'elles pourront nous permettre d'obtenir des résultats suffisamment satisfaisants.

La classification historique, fondée sur la sévérité de l'obstruction bronchique estimée sur le VEMS, ne rend que partiellement compte de la sévérité de la BPCO. Néanmoins, la BPCO est désormais considérée comme une maladie complexe à l'origine de multiples manifestations cliniques directement impliquées dans le pronostic des patients, d'où la notion de phénotypes de patients BPCO. Parmi ces phénotypes, on retrouve différents profils de patients tels que des patients dits « exacerbateurs fréquents », ou encore des patients avec une prédominance d'emphysème, ou plus une symptomatologie de bronchite chronique.

L'étude ECLIPSE a porté sur plus de 2000 patients BPCO suivis sur une période de 3 ans. Cette étude a montré que pour un même stade d'obstruction bronchique (c'est-à-dire un même niveau de VEMS), les patients avaient des symptômes très différents (intensité de la dyspnée, fréquence des exacerbations), une capacité à l'exercice différente et un retentissement

différent sur la qualité de vie. Ainsi, cette étude a mis en évidence un sous-groupe de patients dits « exacerbateurs fréquents », indépendant du stade de sévérité de la maladie, qui présentent plus de deux exacerbations par an tout au long du suivi malgré la prise en charge thérapeutique. Sur la première année de suivi, dans le stade modéré, les exacerbateurs fréquents représentaient 22 % des patients. À l'inverse du phénotype « exacerbateurs fréquents », il semble exister un phénotype « non exacerbateurs » de patients qui ne présentent aucune exacerbation tout au long du suivi. Chez les patients présentant une BPCO très sévère, 29 % des patients ne présentaient aucune exacerbation.

La cohorte « Initiatives BPCO », a évalué 322 patients à l'aide d'un outil de recueil standardisé permettant l'analyse de 8 variables différentes : l'âge, le tabagisme, le VEMS (%), l'index de masse corporel, le nombre d'exacerbations par an, la dyspnée selon le score mMRC et la qualité de vie selon le questionnaire de Saint Georges. L'analyse statistique en clusters des différents phénotypes cliniques a permis de classer les patients en 4 groupes distincts, qui ne recouvraient pas la classification GOLD.

Plusieurs travaux se sont également intéressés aux différences phénotypiques concernant la présentation scannographique de la BPCO. Mille deux patients de la cohorte COPD Gene study ont été analysés. Les auteurs décrivaient deux présentations radiologiques distinctes : l'une à prédominance « emphysémateuse », l'autre à prédominance « bronchique ». Ces deux types de présentations étaient corrélés à la sévérité d'un certain nombre de manifestations cliniques ou fonctionnelles, telles que le score de BODE, indépendamment de la sévérité de l'obstruction bronchique.

Les analyses en clusters pourraient permettre d'identifier, à partir de bases de données constituées sur des cohortes observationnelles de patients, des phénotypes non repérés par la simple observation clinique. Plusieurs équipes dans le monde ont adopté ces stratégies et les résultats de ces études, encore préliminaires, ont déjà apporté certaines informations nouvelles.

Les travaux récents indiquent qu'au moins 2 phénotypes très différents sont à haut risque de mortalité chez les patients BPCO. Le premier est composé de patients ayant une obstruction bronchique sévère (VEMS < 50% de la valeur théorique), une prédominance d'emphysème avec distension thoracique importante et une diffusion du monoxyde de carbone (DLCO) abaissée. Ces patients sont souvent dénutris et ont une forte prévalence d'ostéoporose et des exacerbations fréquentes, alors que les comorbidités cardiovasculaires et le diabète sont faibles. Ils ont une mortalité élevée, qui survient souvent avant l'âge de 70 ans. L'association entre emphysème, dénutrition et ostéoporose a été retrouvée dans une étude récente, dont les auteurs suggèrent la possibilité d'une physiopathologie commune à ces entités.

Le second phénotype, identifié par les analyses en clusters, concerne des patients en surpoids ou obèses, ayant une obstruction bronchique souvent moins sévère que le phénotype précédent, mais une prévalence élevée de facteurs de risque cardiovasculaire (hypertension artérielle, dyslipidémie), de diabète, de coronaropathie et d'insuffisance cardiaque. L'emphysème semble moins marqué, avec une distension thoracique plus faible et une prévalence d'ostéoporose moins élevée. Ces patients ont aussi un mauvais pronostic, mais la mortalité survient généralement plus tard que pour les patients du phénotype précédent. Les conséquences cliniques de ces phénotypes n'ont pas encore été évaluées de façon systématique, mais il est probable que ces 2 sous-groupes de patients BPCO nécessitent des traitements différents pour améliorer leur pronostic. Les patients en surpoids, ayant des comorbidités cardiovasculaires et

métaboliques, mais une maladie respiratoire moins sévère, pourraient grandement bénéficier d'une prise en charge thérapeutique intégrée de la poly pathologie, fondée sur une réduction des facteurs de risque, une activité physique régulière et des traitements médicamenteux connus pour leurs effets cardiovasculaires et métaboliques (antiagrégants plaquettaires, statines, bêtabloquants). Pour les patients ayant au contraire une maladie respiratoire sévère prédominante, avec peu de comorbidités cardiovasculaires et métaboliques, de nouvelles stratégies de diagnostic précoce et des thérapeutiques ciblant les remaniements pulmonaires (emphysème, fibrose péri bronchique, hypersécrétion de mucus) et l'inflammation broncho-pulmonaire devront être développées.

## Conclusion

### Apports principaux de cette thèse

#### Apports d'un point de vue clinique

L'apport de ce travail au sein de la cohorte PALOMB est d'abord clinique. Il contribue à une meilleure compréhension de l'histoire naturelle de la BPCO. Cette pathologie, dont l'expression est hétérogène (interaction avec les comorbidités, symptômes transitoires ou persistants, sévérité variable), est suspectée d'interagir avec les comorbidités et de partager des mécanismes communs. Mais les travaux épidémiologiques s'intéressant à la multimorbidité, en identifiant des profils basés sur la coexistence des comorbidités cardiovasculaire, métabolique, ostéo-articulaire, neuropsychiatrique et d'autres comorbidités chez les patients BPCO, sont encore rares et les connaissances en la matière demeurent lacunaires. Cette étude menée dans une grande cohorte clinique montre que la multimorbidité est courante chez les patients atteints de BPCO et que cinq profils de comorbidités peuvent être identifiés. Les patients GOLD B présentent plus de comorbidités que les patients GOLD C. La présence de comorbidités doit donc être prise en compte dans toute évaluation de la sévérité de la BPCO. Une analyse plus approfondie est nécessaire pour valider ces phénotypes dans une autre cohorte de validation.

Indépendamment des valeurs (seuils) fixées pour définir une dyspnée plus élevée, nos résultats suggèrent que la dyspnée est liée à la sévérité du VEMS, au sexe, aux exacerbations, aux comorbidités et à distension thoracique. Révélant que l'hyperinflation définie par le rapport CI/CPT était un meilleur déterminant de l'échelle mMRC de la dyspnée mMRC comparativement aux rapports VR/CPT et CRF/CPT. Ces résultats illustrent dans une cohorte de vraie vie l'aspect multifactoriel de la dyspnée chez les patients atteints de BPCO. L'élaboration d'un score prédictif en attribuant un poids à chaque catégorie de déterminants fournirait des informations intéressantes. Les facteurs de risque associés à une dyspnée plus élevée devraient être utilisés pour développer et valider extérieurement

un nouveau score de risque de dyspnée plus élevé dans d'autres cohortes. De futures études sont nécessaires pour valider cette règle de prédiction et sa pertinence dans la pratique clinique.

Notre étude montre que les déterminants les plus importants des exacerbations fréquentes sont : les antécédents d'exacerbations, l'anxiété et l'absence de vaccination contre la grippe. Ces analyses montrent l'existence et la pertinence clinique de deux phénotypes d'exacerbateurs fréquents chez les patients atteints de BPCO et le seuil actuellement utilisé pour définir ce phénotype.

Le suivi longitudinal a permis d'identifier des caractéristiques cliniques et des comorbidités associées à un risque de mortalité plus élevé chez les patients atteints de BPCO. En particulier, le cancer du poumon était celui qui pesait le plus lourdement sur le risque de mortalité. L'amélioration du diagnostic précoce du cancer du poumon pourrait modifier les trajectoires de ces patients BPCO à haut risque. Les différences de mortalité pourraient être encore plus prononcées à long terme, mais cela devra être confirmé par des analyses de survie supplémentaires de la cohorte.

Les analyses en clusters ont identifié un certain nombre de phénotypes cliniques chez les patients atteints de BPCO mais leurs résultats sont difficiles à utiliser en pratique clinique quotidienne. En plus, La complexité de la pathogénèse de la BPCO reflète la grande variabilité des phénotypes cliniques.

L'identification des phénotypes de patients atteints de BPCO laisse envisager une compréhension de l'hétérogénéité de cette maladie et une meilleure individualisation de la prise en charge thérapeutique.

Nous avons mis au défi, dans le cadre de cette étude, trois méthodes de classification (hiérarchique et non-hiérarchiques). Suite à une évaluation de la qualité des clusters obtenus, nous avons opté pour la méthode PAM, comme étant la plus adaptée pour notre jeu de données. Comme cité auparavant dans d'autres études, la méthode PAM délivre, en effet, de meilleurs résultats de classification par rapport au k-Means et HCPC quand nous manipulons des jeux de données de dimension relativement petit.

La BPCO est une pathologie hétérogène responsable d'une morbidité et d'une mortalité importante. Les traitements actuels de la maladie ont des effets symptomatiques certains (réduction de la dyspnée, amélioration de la capacité d'exercice, diminution de la fréquence des exacerbations) se traduisant par une amélioration de la qualité de vie des patients. L'amplitude des effets thérapeutiques observés reste insuffisante et les traitements actuels inefficaces pour diminuer la mortalité. Les travaux récents suggèrent l'intérêt d'identifier des sous-groupes de patients BPCO ayant des caractéristiques cliniques, fonctionnelles, respiratoires, scannographique et/ou biologiques différentes. Cette approche pourrait permettre de mieux cibler les traitements actuellement disponibles et de progresser dans la compréhension des mécanismes physiopathologiques, menant à un phénotype plutôt qu'à un autre, avec comme objectif le développement de nouvelles approches thérapeutiques.

## Apports d'un point de vue de modélisation statistique

Nous avons au cours de ce mémoire présenté en détail des méthodes de clustering, innovantes et récentes. Une attention particulière a été apportée à leurs défauts : grande complexité, dépendance à un paramètre difficile à choisir ( $k$ = nombre optimale de clusters), très impactant et non dynamique. De plus nous avons mis en lumière des défaillances de quelques méthodes comparus à d'autres, impactant lourdement la qualité du clustering et intervenant dans des circonstances pourtant communes. Nous avons proposé PAM, une méthode itérative peu complexe par rapport au nombre d'observation  $N$ , permettant d'obtenir les densités sous une forme arborescente.

## Perspectives

Nous savons que la prévalence de la BPCO est sous diagnostiquée et les symptômes peu spécifiques souvent sous évalués par les patients n'aident pas les praticiens dans l'orientation diagnostique. De nombreuses améliorations pourraient être effectuées sur le dépistage des patients à risque de BPCO : notre étude appuyée par les données de la littérature prouve que les patients BPCO présentent dans une très grande majorité de cas une pathologie chronique associée. Les recommandations de bonnes pratiques cliniques devraient inclure la notion de multimorbidité comme partie intégrante de la BPCO afin de permettre aux praticiens de mieux cibler les patients à fort risque de BPCO et de morbidité nécessitant une mesure de la fonction pulmonaire. Ceci favoriserait un dépistage ciblé plus précoce des patients BPCO ainsi qu'une prise en charge globale centrée sur le patient BPCO.

L'amplitude des effets thérapeutiques observés reste insuffisante et les traitements actuels inefficaces pour diminuer la mortalité. Nos travaux suggèrent l'intérêt d'identifier des sous-groupes de patients BPCO ayant des caractéristiques cliniques, fonctionnelles, respiratoires, scannographique et/ou biologiques différentes. Cette approche pourrait permettre de mieux cibler les traitements actuellement disponibles et de progresser dans la compréhension des mécanismes physiopathologiques, menant à un phénotype plutôt qu'à un autre, avec comme objectif le développement de nouvelles approches thérapeutiques.

## **BIBLIOGRAPHIE**

- [1] Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet*. 1997;349(9064):1498-1504. Doi: 10.1016/S0140-6736(96)07492-2
- [2] Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M, Stockley RA, Sin DD, Rodriguez-Roisin R. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013 Feb 15;187(4):347-65.
- [3] Sutherland ER, Cherniack RM. Management of chronic obstructive pulmonary disease. *N Engl J Med*. 2004;350(26):2689–2697.
- [4] Questionnaire Dépistage BPCO GOLD Web – [Internet]. 16 février 2016. [http://www.has-sante.fr/portail/upload/docs/application/pdf/201406/questionnaire\\_depistage\\_bpco\\_gold\\_web.pdf](http://www.has-sante.fr/portail/upload/docs/application/pdf/201406/questionnaire_depistage_bpco_gold_web.pdf)
- [5] Guide parcours de soins BPCO [Internet]. [Cité 15 juin 2015]. Disponible sur: [http://www.has-sante.fr/portail/plugins/ModuleXitiKLEE/types/FileDocument/doXiti.jsp?id=c\\_1242505](http://www.has-sante.fr/portail/plugins/ModuleXitiKLEE/types/FileDocument/doXiti.jsp?id=c_1242505)
- [6] Société de Pneumologie de Langue Française. Recommandation pour la Pratique Clinique. *Rev Mal Respir*. Mai 2010 ; 27(5) : 522-48.
- [7] Roche P, Cuvelier P, Marquette P, Chabot P. CHRONIQUE OBSTRUCTIVE (BPCO). [cité 16 févr 2016]; Disponible sur: <http://www.urps-mk-hn.fr/medias/files/cep-ecn-2010-bpc0.pdf>
- [8] Global strategy for the diagnosis, management and prevention of COPD, global initiative for chronic obstructive lung disease (GOLD) 2017. Accessible : <http://goldcopd.org>
- [9] Jochmann A, Neubauer F, Miedinger D, et al. General practitioners' adherence to the COPD GOLD guidelines: baseline data from the Swiss COPD Cohort Study. *Swiss Med Wkly* 2010; 142: w13567.
- [10] Amigo H, Erazo M, Oyarzún M, Bello S, Peruga A. Smoking and chronic obstructive pulmonary disease: attributable risk determination. *Rev Med Chil*. 2006 Oct; 134 (10):1275-82. Epub 2006 Dec 13.
- [11] Eisner MD, Anthonisen N, Coultas D, Kuenzli N, Perez-Padilla R, Postma D, Romieu I, Silverman EK, Balmes JR; Committee on Nonsmoking COPD, Environmental and Occupational Health Assembly. An official American Thoracic Society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2010 Sep 1;182(5):693-718.
- [12] Løkke A, Lange P, Scharling H, Fabricius P, Vestbo J. Developing COPD: a 25 year follow up study of the general population. *Thorax*. 2006 Nov;61(11):935-9.
- [13] Gan WQ, Man SF, Postma DS, Camp P, Sin DD. Female smokers beyond the perimenopausal period are at increased risk of chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Respir Res*. 2006 Mar 29; 7:52.

- [14] Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M, Stockley RA, Sin DD, Rodriguez-Roisin R. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013 Feb 15;187(4):347-65.
- [15] Lundbäck B, Lindberg A, Lindström M, Rönmark E, Jonsson AC, Jönsson E, Larsson LG, Andersson S, Sandström T, Larsson K; Obstructive Lung Disease in Northern Sweden Studies. Not 15 but 50% of smokers develop COPD? -Report from the Obstructive Lung Disease in Northern Sweden Studies. *Respir Med*. 2003 Feb;97(2):115-22.
- [16] Silverman EK, Vestbo J, Agusti A, Anderson W, Bakke PS, Barnes KC, Barr RG, Bleecker ER, Boezen HM, Burkart KM, Celli BR, Cho MH, Cookson WO, Croxton T, Daley D, DeMeo DL, Gan W, Garcia-Aymerich J, Hall IP, Hansel NN, Hersh CP, Kalsheker N, Kiley JP, Kim WJ, Lambrechts D, Lee SD, Litonjua AA, Lomas DA, London SJ, Nishimura M, Nørdestgaard BG, O'Donnell CJ, Postma DS, Puhon MA, Tesfaygi Y, Tobin MD, Vogelmeier C, Wilk JB, Wouters E, Young RP, Ziegler-Heitbrock L, MacNee W, Crapo JD. Opportunities and challenges in the genetics of COPD 2010: An International COPD Genetics Conference report. *COPD*. 2011 Apr;8(2):121-35.
- [17] Ekeowa UI, Gooptu B, Belorgey D, Häggglöf P, Karlsson-Li S, Miranda E, Pérez J, MacLeod I, Kroger H, Marciniak SJ, Crowther DC, Lomas DA. alpha1-Antitrypsin deficiency, chronic obstructive pulmonary disease and the serpinopathies. *Clin Sci (Lond)*. 2009 May 14; 116(12):837-50.
- [18] Dransfield MT, Davis JJ, Gerald LB, Bailey WC. Racial and gender differences in susceptibility to tobacco smoke among patients with chronic obstructive pulmonary disease. *Respir Med*. 2006 Jun;100(6):1110-6.
- [19] Charlson M, Charlson RE, Briggs W, Hollenberg J. Can disease management target patients most likely to generate high costs? The impact of comorbidity. *J Gen Intern Med* 2007; 22:464-9.
- [20] Mannino DM, Thorn D, Swensen A, Holguin F. Prevalence and outcomes of diabetes, hypertension and cardiovascular disease in COPD. *Eur Respir J*. 2008; 32:962-969.
- [21] Tillie-Leblond I, Marquette CH, Perez T, Scherpereel A, Zanetti C, Tonnel AB, et al. Pulmonary embolism in patients with unexplained exacerbation of chronic obstructive pulmonary disease: prevalence and risk factors. *Ann Intern Med*. 2006; 144:390-6.
- [22] Gunen H, Gulbas G, In E, Yetkin O, Hacıevliyagil SS. Venous thromboemboli and exacerbations of COPD. *Eur Respir J*. 2010; 35:1243—8.
- [23] Mascarenhas J, Lourenço P, Lopes R, Azevedo A, Bettencourt P. Chronic obstructive pulmonary disease in heart failure. Prevalence, therapeutic and prognostic implications. *Am Heart J*. 2008; 155:521—5.

- [24] Chaouat A, Naeije R, Weitzenblum E. Pulmonary hypertension in COPD. *Eur Respir J* 2008; 32:1371-85.
- [25] Chaouat A, Bugnet AS, Kadaoui N. Severe pulmonary hypertension and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2005; 172:189—94.
- [26] Donaldson GC, Hurst JR, Smith CJ, Hubbard RB, Wedzicha JA. Increased risk of myocardial infarction and stroke following exacerbation of COPD. *Chest*. 2010; 137:1091-7.
- [27] Rutten FH, Cramer MJ, Lammers JW, Grobbee DE, Hoes AW. Heart failure and chronic obstructive pulmonary disease: an ignored combination? *Eur J Heart Fail*. 2006; 8:707—11.
- [28] Wagner PD. Possible mechanisms underlying the development of cachexia in COPD. *EurRespirJ*. 2008; 31:492—501.
- [29] Cano N, Pichard C, Roth H, Court-Fortuné I, Cynober L, Gérard-Boncompain M, et al. Creactive protein and body mass index predict outcome in end stage respiratory failure. *Chest*. 2004; 126:540—6.
- [30] Vestbo J, Prescott E, Almdal T, Dahl M, Nordestgaard BG, Andersen T, et al. Body mass, fat-free body mass, and prognosis in patients with chronic obstructive pulmonary disease from a random population sample: findings from the Copenhagen city Heart Study. *Am J Respir Crit Care Med*. 2006; 173:79—83.
- [31] Landbo C, Prescott E, Lange P, Vestbo J, Almdal TP. Prognostic value of nutritional status in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 1999; 160:1856-61.
- [32] Chailleux E, Laaban JP, Veale D. Prognostic value of nutritional depletion in patients with COPD treated by long term oxygen therapy: data from the ANTADIR observatory. *Chest*. 2003; 123:1460-6.
- [33] Schols AM, Broekhuizen R, Weling-Scheepers CA, Wouters EF. Body composition and mortality in chronic obstructive pulmonary disease. *Am J Clin Nutr*. 2005; 82:53-9.
- [34] Soler-Cataluna JJ, Sanchez-Sanchez L, Martinez Garcia M, Sanchez PR, Salcedo E, Navarro M. Mid arm muscle area is a better predictor of mortality than body mass index in COPD. *Chest*. 2005; 128:2108—15.
- [35] Ferreira IM, Brooks D, Lacasse Y, Goldstein RS. Nutritional support for individuals with COPD: A meta-analysis. *Chest*. 2000; 117:672—8.
- [36] Pison CM, Cano NJ, Chérion C, Caron F, Court-Fortune I, Anto-nini MT, et al. Multimodal nutritional rehabilitation improves clinical outcomes of malnourished patients with chronic respiratory failure: a randomised controlled trial. *Thorax*. 2011; 66:953—60.
- [37] Jordan JG, Mann JR. Obesity and mortality in persons with obstructive lung disease using data from the NHANES III. *South Med J*. 2010; 103:232-330.

- [38] Dodge R, Cline MG, Quan SF. The natural history of insomnia and its relationship to respiratory symptoms. *Arch Intern Med.* 1995; 155:1797-800.
- [39] Budhiraja R, Parthasarathy S, Budhiraja P, Habib MP, Wendel C, Quan SF. Insomnia in patients with COPD. *Sleep* 2012; 35:369-75.
- [40] Mannino DM, Thorn D, Swensen A, Holguin F. Prevalence and outcomes of diabetes, hypertension and cardiovascular disease in COPD. *Eur Respir J.* 2008; 32:962-9.
- [41] Cazzola M, Bettoncelli G, Sessa E, Cricelli C, Biscione G. Prevalence of comorbidities in patients with COPD. *Respiration.* 2010; 80:112-9.
- [42] Alberti KG, Zimmet P, Shaw J. Metabolic syndrome: a new world wide definition: a consensus statement from the IDF. *Diabet Med.* 2006 ; 23 :469-80.
- [43] Ghanassia E, Jaussent A, Picot MC, Prefaut C. Prévalence de l'insulinorésistance et du syndrome métabolique chez le sujet BPCO. *Rev Mal Respir.* 2006 ;23 :393.
- [44] Curtis JR. Palliative and end-of-life care for patients with severe COPD. *Eur Respir J.* 2008; 32: 796-803.
- [45] Dinicola G, Julian L, Gregorich SE, Blanc PD, Katz PP. The role of social support in anxiety for persons with COPD. *J Psychosom Res.* 2013 ; 74 : 110-5.
- [46] Regvat J, Zmitek A, Vegnuti M, Kosnik M, Suskovic S. Anxiety and depression during hospital treatment of exacerbation of chronic obstructive pulmonary disease. *J Int Med Res.* 2011; 39: 1028-1038.
- [47] Schneider C, Jick SS, Bothner U. COPD and the risk of depression. *Chest.* 2010; 137: 341-7.
- [48] Tze-Pin NG, Niti M, Wan-Cheng T, Zhenying C, Kian-Chung O. Depressive symptoms and chronic obstructive pulmonary disease. *Arch Intern Med.* 2007; 167: 60-7.
- [49] Abrams TE, Vaughan-Sarrazin M, Vander Weg MW. Acute exacerbations of chronic obstructive pulmonary disease and the effect of existing psychiatric comorbidity on subsequent mortality. *Psychosomatics.* 2011; 52: 441-9.
- [50] Loganathan RS, Stover DE, Shi W, Venkatraman E. Prevalence of COPD in women compared to men around the time of diagnosis of primary lung cancer. *Chest.* 2006; 129 :1305-12.
- [51] Wasswa-Kintu S, Gan WQ, Man SF, Pare PD, Sin DD. Relationship between reduced forced expiratory volume in one second and the risk of lung cancer: a systematic review and meta-analysis. *Thorax.* 2005; 60:570-5.
- [52] Fabbri LM, Rabe KF. From COPD to chronic systemic inflammatory syndrome? *Lancet.* 2007; 370: 797-9.

- [53] Hurst JR, Vestbo J, Anzueto A, Locantore N, Müllerova H, Tal-Singer R, et al. Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) Investigators. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med*. 2010; 363: 1128-38.
- [54] Burgel PR, Paillasseur JL, Caillaud D, Tillie-Leblond I, Chanez P, Escamilla R, et al. Initiatives BPCO Scientific Committee. Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. *Eur Respir J*. 2010; 36: 531-9.33.
- [55] Han M, Washko G, Dransfield M, Liu L, Marchetti N, Casaburi R, et al. CT phenotype correlates with health status and BODE in COPD. *Eur Respir J*. 2010; 36: A5551.
- [56] Recommandations de la Société de Pneumologie de Langue Française sur la prise en charge de la BPCO (mise à jour 2009). Supplément 1, pp. S1-S76 (juin 2010).
- [57] Calverley PM, Anderson JA, Celli B, Ferguson GT, Jenkins C, Jones PW, et al. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *N Engl J Med*. 2007; 356: 775-89.
- [58] Tashkin DP, Celli B, Senn S, Burkhart D, Kesten S, Menjoge S, et al. A 4-year trial of tiotropium in chronic obstructive pulmonary disease. *N Engl J Med*. 2008; 359: 1543-54.
- [59] Vestbo J, Pauwels R, Anderson JA, Jones P, Calverley P. Early onset of effect of salmeterol and fluticasone propionate in chronic obstructive pulmonary disease. *Thorax*. 2005; 60: 301-4.
- [60] Szafranski W, Cukier A, Ramirez A, Menga G, Sansores R, Nahabedian S, et al. Efficacy and safety of budesonide/formoterol in the management of chronic obstructive pulmonary disease. *Eur Respir J*. 2003; 21: 74-81.
- [61] Hay JG, Stone P, Carter J, Church S, Eyre-Brook A, Pearson MG, et al. Bronchodilator reversibility, exercise performance and breathlessness in stable chronic obstructive pulmonary disease. *Eur Respir J*. 1992; 5: 659-64.
- [62] O'Donnell DE, Sciruba F, Celli B, Mahler DA, Webb KA, Kalberg J, et al. Effect of fluticasone propionate/salmeterol on lung hyperinflation and exercise endurance in COPD. *Chest*. 2006; 130: 647-56.
- [63] Gros PA, Hermogenes AW, Sacks HS, Lau J, Levandowski RA. The efficacy of influenza vaccine in elderly persons. *Ann Intern Med*. 1995; 123: 518-27.
- [64] Wongsurakiat P, Maranetra KN, Wasi C, et al. Acute respiratory illness in patients with COPD and the effectiveness of influenza vaccination: a randomized controlled study. *Chest*. 2004; 125: 2011-20.
- [65] Fletcher C, Peto R. The natural history of chronic airflow obstruction. *BMJ*. 1977; 1: 1645-8.

- [66] Anthonisen NR, Connett JE, Kiley JP, Altose MD, Bailey WC, Buist AS, Conway WA, Jr, Enright PL, Kanner RE, O'Hara P, et al. Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. *JAMA*. 1994; 272: 1497-505.
- [67] Conférence de consensus de la SPLF sur la réhabilitation respiratoire 2005 ; Recommandations HAS 2014 « Comment mettre en œuvre la réhabilitation respiratoire pour les patients ayant une bronchopneumopathie chronique obstructive ».
- [68] Ambrosino N, Simonds A. The clinical management in extremely severe COPD. *Respir Med*. 2007; 101: 1613-1624.
- [69] Bart F, Wallaert B. Indications de la réhabilitation respiratoire chez le BPCO. *Rev Mal Respir*. 2005 ; 22 : 7S17-7S18.
- [70] Troosters T, Casaburi R, Gosselink R, Decramer M. Pulmonary rehabilitation in chronic obstructive disease. *Am J Respir Crit Care Med*. 2005; 172: 19-38.
- [71] Tillie-Leblond I, Masure F, pour le groupe EXACO. « EXACO » : suivi sur 4 ans des exacerbations d'une cohorte de patients atteints de broncho pneumopathie chronique obstructive. *Rev Mal Respir* 2006 ; 23 :377-84.
- [72] Jebrack G, pour Initiatives BPCO. Recommandations et prise en charge de la BPCO en France : les recommandations sur la prise en charge de la BPCO ne sont pas suivies dans la vraie vie ! *Rev Mal Respir*. 2010 ; 27 : 11-18.
- [73] Kelkel E et al. COLIBRI : optimiser la pratique clinique et produire des données scientifiques pertinentes. *Rev Mal Respir*. 2015, article in press <http://dx.doi.org/10.1016/j.rmr.2015.02.090>
- [74] Silverman EK, Sandhaus RA. Clinical practice. Alpha1-antitrypsin deficiency. *N Engl J Med*. 2009; 360(26):2749-57.
- [75] Fishman A, Martinez F, Naunheim K et al.; National Emphysema Treatment Trial Research Group. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* 2003; 348(21):2059-73.
- [76] Hurst JR, Vestbo J, Anzueto A et al.; Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) Investigators. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med* 2010; 363(12):1128-38.
- [77] Roche N, Aguilaniu B, Burgel PR et al. ; groupe d'experts "Exacerbations de BPCO". Prévention des exacerbations de BPCO : un enjeu fondamental. *Rev Mal Respir*. 2012 ; 29(6) :756-74.
- [78] Albert RK, Connett J, Bailey WC et al.; COPD Clinical Research Network. Azithromycin for prevention of exacerbations of COPD. *N Engl J Med* 2011;365(8):689-98.

- [79] J. Mac Queen. "Some methods for classification and analysis of multivariate observations". In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. Berkeley, Calif.: University of California Press, 1967, p. 281–297.
- [80] Xin Jin et Jiawei Han. "K-Medoids Clustering". In: Encyclopedia of Machine Learning. Sous la dir. de Claude Sammut et Geoffrey I. Webb. Boston, MA: Springer US, 2010, p. 564–565. isbn: 978-0-387-30164-8. doi:10.1007/978-0-387-30164-8\_426. url: [https://doi.org/10.1007/978-0-387-30164-8\\_426](https://doi.org/10.1007/978-0-387-30164-8_426)
- [81] Robert Gray. "Vector quantization". In : IEEE Assp Magazine 1.2 (1984), p. 4–29.
- [82] Izong Cheng. "Mean shift, mode seeking, and clustering". In: IEEE Transactions on Pattern Analysis and Machine Intelligence 17.8 (1995), p. 790-799. Issn: 0162-8828. doi: <https://doi.org/10.1109/34.400568>
- [83] Stephen C Johnson. "Hierarchical clustering schemes". In : Psychometrika 32.3(1967), p. 241–254.
- [84] Ricardo JGB Campello, Davoud Moulavi et Joerg Sander. "Density-based clustering based on hierarchical density estimates". In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer. 2013, p. 160–172.
- [85] John H Wolfe. "Pattern clustering by multivariate mixture analysis". In : Multivariate Behavioral Research 5.3 (1970), p. 329–350.
- [86] Arthur P Dempster, Nan M Laird et Donald B Rubin. "Maximum likelihood from incomplete data via the EM algorithm". In: Journal of the royal statistical society. Series B (methodological) (1977), p. 1–38.
- [87] Geoffrey McLachlan et Thriyambakam Krishnan. The EM algorithm and extensions. T. 382. John Wiley & Sons, 2007.
- [88] Naftali Tishby, Fernando C Pereira et William Bialek. "The information bottleneck method". In: arXiv preprint physics/0004057(2000).
- [89] Thomas S Ferguson. "A Bayesian analysis of some nonparametric problems". In: The annals of statistics (1973), p. 209–230.
- [90] David Arthur et Sergei Vassilvitskii. "k-means++: The advantages of careful seeding". In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial et Applied Mathematics. 2007, p. 1027–1035.
- [91] Leonard Kaufman et Peter Rousseeuw. Clustering by means of medoids. North-Holland, 1987.
- [92] J. C. Dunn. "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters". In: Journal of Cybernetics 3.3 (1973), p. 32–57. doi: <http://dx.doi.org/10.1080/01969727308546046>

- [93] P. RAI & S. SINGH, A survey of clustering techniques, International Journal of Computer Applications (0975 - 8887) Volume 7- No.12, October 2010.
- [94] L. Kaufman and P. J. Rousseeuw. << Finding groups in Data: an introduction to Cluser Analysis m. John Wiley & Sons, 1990
- [95] R. Ng and J. Han, " Efficient and Effective Clustering Method for Spatial Data Mining", Proc. of 1994 Int'l Conf. on Very Large Data Bases (VLDB'94). Santiago, Chile, September 1994, pp. 144-155.
- [96] B. Elfron, T. Hastie et R. Tibshirani. Least angle regression. Annals of Statistics 32, 407-499, 2004.
- [97] J. Mairal. Simple Explanation for LARS. 2008.
- [98] Schwarz G. Estimating the Dimension of a Model. Ann Statist. mars 1978;6(2):461-4.
- [99] Goto T, Jo T, Matsui H, Fushimi K, Hayashi H, Yasunaga H. Machine Learning-Based Prediction Models for 30-Day Readmission after Hospitalization for Chronic Obstructive Pulmonary Disease. COPD. 2019 Dec; 16(5-6): 338-343. <https://doi.org/10.1080/15412555.2019.1688278>
- [100] Genolini C, Alacoque X, Sentenac M, Arnaud C. kml and kml3d: R packages to cluster longitudinal data. J Stat Softw. 2015; 65(4):1-34. DOI: <https://doi.org/10.18637/jss.v065.i04>
- [101] Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and Sp to analyze and compare ROC curves. BMC Bioinformatics. 2011; 12: 77. DOI: <https://doi.org/10.1186/1471-2105-12-77>
- [102] Mise en œuvre du décret n° 98–37 autorisant l'accès aux données relative au décès des personnes inscrites au RNIPP dans le cadre des recherché dans le domaine de la santé. 27 Mai 2015 <http://cesp.vjf.inserm.fr/svcd>
- [103] Celli BR, Jones P, Vestbo J, Anderson JA, Ferguson GT, Yates J, et al. The multidimensional BOD: association with mortality in the TORCH trial. Eur Respir J. 2008 ; 32 : Suppl. 52, 42s.
- [104] Burgel PR, Paillasseur JL, Janssens W, Piquet J, Riet GT, Aymerich JG, et al. A simple algorithm for the identification of clinical COPD phenotypes. Eur Respir J. 2017 ; 50 : 1701034.
- [105] Chen W, et al. Risk of cardiovascular comorbidity in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. Lancet Respir Med. 2015;3(8):631–9.
- [106] Divo M, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2012 ;186(2) :155–61.
- [107] Soler X, et al. High prevalence of obstructive sleep apnea in patients with moderate to severe chronic obstructive pulmonary disease. Ann Am Thorac Soc. 2015 ;12(8) :1219–25.
- [108] Marin JM, et al. Outcomes in patients with chronic obstructive pulmonary disease and obstructive sleep apnea: the overlap syndrome. Am J Respir Crit Care Med. 2010;182(3):325–31.

- [109] Vanfleteren LE, et al. Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2013;187(7):728–35.
- [110] Lange P, et al. Prediction of the clinical course of chronic obstructive pulmonary disease, using the new GOLD classification: a study of the general population. *Am J Respir Crit Care Med*. 2012 ;186(10) :975–81.
- [111] Divo MJ, et al. COPD comorbidities network. *Eur Respir J*. 2015;46(3):640–50.
- [112] Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med*. 2010;363(12):1128-1138. DOI: <https://doi.org/10.1056/NEJMoa0909883>
- [113] Han MK, Quibrera PM, Carretta EE, et al. Frequency of exacerbations in patients with chronic obstructive pulmonary disease: an analysis of the SPIROMICS cohort. *Lancet Respir Med*. 2017; 5(8): 619-626. DOI: [https://doi.org/10.1016/S2213-2600\(17\)30207-2](https://doi.org/10.1016/S2213-2600(17)30207-2)
- [114] Le Rouzic Olivier, Nicolas Roche, Alexis B. Cortot, et al. Defining the “Frequent Exacerbator” Phenotype in COPD. *CHEST*, Volume 153, Issue 5, 1106 - 1115. DOI: <https://doi.org/10.1016/j.chest.2017.10.009>
- [115] K. Kostikas, A. Clemens, F. Patalano. Prediction and prevention of exacerbations and mortality in patients with COPD. *Expert Rev Respir Med*, 10 (7) (2016), pp. 739-753. DOI: <https://doi.org/10.1080/17476348.2016.1185371>
- [116] Beckham JD, Cadena A, Lin J, *et al*. Respiratory viral infections in patients with chronic, obstructive pulmonary disease. *J Infect* 2005; 50:322–330. DOI: <https://doi.org/10.1016/j.jinf.2004.07.011>
- [117] Ko FW, Ip M, Chan PK *et al*. Viral etiology of acute exacerbations of COPD in Hong Kong. *Chest* 2007 ; 132 :900–908. DOI: <https://doi.org/10.1378/chest.07-0530>
- [118] McManus TE, Marley AM, Baxter N *et al*. Respiratory viral infection in exacerbations of COPD. *Respir Med* 2008; 102:1575–1580. DOI: <https://doi.org/10.1016/j.rmed.2008.06.006>
- [119] Rohde G, Wiethage A, Borg I *et al*. Respiratory viruses in exacerbations of chronic obstructive pulmonary disease requiring hospitalisation: a case-control study. *Thorax* 2003; 58:37–42. DOI: <http://dx.doi.org/10.1136/thorax.58.1.37>
- [120] European Centre for Disease Prevention and Control (ECDC). Seasonal influenza vaccination in Europe. Overview of vaccination recommendations and coverage rates in the EU member states for the 2012–13 influenza season. 2015.
- [121] Vaccines against influenza WHO position paper - November 2012. *Wkly Epidemiol Rec*. 2012; 87: 461–476. PMID: 23210147.

- [122] Centers for Disease Control and Prevention (CDC). People at High Risk of Developing Flu-Related Complications 2015 [[http://www.cdc.gov/flu/about/disease/high\\_risk.htm](http://www.cdc.gov/flu/about/disease/high_risk.htm)] Accessed on 24 Feb 2016.
- [123] Grohskopf LA, Sokolow LZ, Broder KR, et al. Prevention and control of seasonal influenza with vaccines. *MMWR Recomm Rep* 2016; 65:1–54. DOI: <http://dx.doi.org/10.15585/mmwr.rr6505a1>
- [124] Mark B. Parshall, Richard M. Schwartzstein, Lewis Adams, Robert B. Banzett, Harold L. Manning, Jean Bourbeau, et al. An official American thoracic society statement: update on the mechanisms, assessment, and management of dyspnea. *Am J Respir Crit Care Med*. 2012 Feb 15. 185(4): 435-452. <https://doi.org/10.1164/rccm.201111-2042ST>
- [125] Louis Lavolette, Pierantonio Laveneziana. Dyspnoea : A multidimensional and multidisciplinary approach. *Eur Respir J*. 2014 Jun. 43 : 1750-1762. <https://doi.org/10.1183/09031936.00092613>
- [126] Jørgen Vestbo, Suzanne S. Hurd, Alvar G. Agustí, Paul W. Jones, Claus Vogelmeier, Antonio Anzueto, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013 Feb 15. 187(4): 347 - 365. <https://doi.org/10.1164/rccm.201204-0596PP>
- [127] Pere Almagro, Sergi Yun, Ana Sangil, Mónica Rodríguez-Carballeira, Meritxell Marine, Pedro Landete, et al. Palliative care and prognosis in COPD: a systematic review with a validation cohort; *Int J Chron Obstruct Pulmon Dis*. 2017; 12: 1721—1729. <https://dx.doi.org/10.2147/COPD.S135657>
- [128] Ciro Casanova, Claudia Cote, Juan P. de Torres, Armando Aguirre-Jaime, Jose M. Marin, Victor Pinto-Plata, et al. Inspiratory-to-total lung capacity ratio predicts mortality in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2005 Mar 15. 171(-):591–597. <https://doi.org/10.1164/rccm.200407-867OC>
- [129] Jorine E. Hartman, Nick H.T. ten Hacken, Karin Klooster, H. Marika Boezen, Mathieu H.G. de Greef and Dirk-Jan Slebos. The minimal important difference for residual volume in patients with severe emphysema. *Eur Respir J*. 2012 Nov; 40(5): 1137–1141. <https://doi.org/10.1183/09031936.00219111>
- [130] Alfonso Fiorelli, Antonio D’Andrilli, Michela Bezzi, Mohsen Ibrahim, Marco Anile, Daniele Diso, et al. Complications related to endoscopic lung volume reduction for emphysema with endobronchial valves: results of a multicenter study. *J Thorac Dis*. 2018 Jun 08. 10(27): S3315-S3325. <https://doi.org/10.21037/jtd.2018.06.69>
- [131] Andrea Rossi, Zaurbek R. Aisanov, Sergey Avdeev, Giuseppe Di Maria, Claudio F. Donner, José Luis Izquierdo, et al. Mechanisms, assessment and therapeutic implications of lung hyperinflation in COPD. *Respir Med*. 2015 Jul. 109(7), 785-802. <https://doi.org/10.1016/j.rmed.2015.03.010>

- [132] Roberto Duranti, Mario Filippelli, Roberto Bianchi, Isabella Romagnoli, Riccardo Pellegrino, Vito Brusasco, et al. Inspiratory Capacity and Decrease in Lung Hyperinflation with Albuterol in COPD. *Chest*. 2002 Dec. 122(6): 2009-2014. <https://doi.org/10.1378/chest.122.6.2009>
- [133] Maria A. Ramon, Jaume Ferrer, Elena Gimeno-Santos, David Donaire-Gonzalez, Esther Rodriguez, Eva Balcells, et al. Inspiratory capacity-to-total lung capacity ratio and dyspnoea predict exercise capacity decline in COPD. *Respirology*. 2016 Apr; 21(3): 476–482. <https://doi.org/10.1111/resp.12723>
- [134] Perez T, Deslée G, Burgel PR, Caillaud D, Le Rouzic O, Zysman M, et al. Predictors in routine practice of 6-min walking distance and oxygen desaturation in patients with COPD: impact of comorbidities. *Int J Chron Obstruct Pulmon Dis*. 2019 ; 14 : 1399-1410.
- [135] Axson EL, Ragutheeswaran K, Sundaram V, Bloom CI, Bottle A, Cowie MR, et al. Hospitalisation and mortality in patients with comorbid COPD and heart failure: a systematic review and meta-analysis. *Respir Res*. 2020; 21: 54.
- [136] Anthonisen NR, Skeans MA, Wise RA, Manfreda J, Kanner RE, Connett JE, et al. The Effects of a Smoking Cessation Intervention on 14.5-Year Mortality: A Randomized Clinical Trial. *Ann Intern Med*. 2005 ; 142 : 233–239.
- [137] Bhowmik A, Seemungal TAR, Sapsford RJ, Wedzicha JA. Relation of sputum inflammatory markers to symptoms and lung function changes in COPD exacerbations. *Thorax*. 2000 ; 55 : 114-120.
- [138] Petersen H, Sood A, Meek PM, Shen X, Cheng Y, Belinsky SA, et al. Rapid lung function decline in smokers is a risk factor for COPD and is attenuated by angiotensin-converting enzyme inhibitor use. *Chest*. 2014; 145(4): 695-703.
- [139] Welte T, Vogelmeier C, and Papi A. COPD: early diagnosis and treatment to slow disease progression. *Int J Clin Pract*. 2015 ; 69 : 336-349.
- [140] Celli B, Vestbo J, Jenkins CR, Jones PW, Ferguson GT, Calverley Peter MA, et al. Sex differences in mortality and clinical expressions of patients with chronic obstructive pulmonary disease. The TORCH experience. *Am J Respir Crit Care Med*. 2011 ; 183(3) : 317-322.
- [141] Prudente R, Franco EAT, Mesquita CB, Ferrari R, Godoy ID, Tanni SE. Predictors of mortality in patients with COPD after 9 years. *Int J Chron Obstruct Pulmon Dis*. 2018; 13: 3389-3398.
- [142] Arthur D, Vassilvitskii S. k-means++: the advantages of careful seeding, p 1027– 1035. In: SODA'07: proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA; 2007.
- [143] RDUSSEEUN LKPJ, KAUFMAN P. Clustering by means of medoids. 1987 ;
- [144] Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. 1973.

- [145] Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*. 1977 ;39(1) :1–38. Available from : <http://www.jstor.org/stable/2984875>.
- [146] Han J, Kamber M, Tung AKH. *Spatial clustering methods in data mining: A survey*; 2001.

## **ANNEXES**

## Liste des annexes

<b>Annexe 1.</b> PALOMB English questionnaire 1/3 .....	145
<b>Annexe 2.</b> PALOMB English questionnaire 2/3 .....	146
<b>Annexe 3.</b> PALOMB English questionnaire 3/3 .....	147
<b>Annexe 4.</b> Les déterminants de la dyspnée selon l'échelle mMRC (1/3).....	148
<b>Annexe 5.</b> Les déterminants de la dyspnée selon l'échelle mMRC (2/3) .....	149
<b>Annexe 6.</b> Les déterminants de la dyspnée selon l'échelle mMRC (3/3) .....	150
<b>Annexe 7.</b> Les phénotypes et les déterminants d'exacerbations (1/6) .....	151
<b>Annexe 8.</b> Les phénotypes et les déterminants d'exacerbations (2/6) .....	152
<b>Annexe 9.</b> Les phénotypes et les déterminants d'exacerbations (3/6) .....	153
<b>Annexe 10.</b> Les phénotypes et les déterminants d'exacerbations (4/6) .....	154
<b>Annexe 11.</b> Les phénotypes et les déterminants d'exacerbations (5/6) .....	155
<b>Annexe 12.</b> Les phénotypes et les déterminants d'exacerbations (6/6) .....	156
<b>Annexe 13.</b> Les déterminants de la mortalité (1/13).....	157
<b>Annexe 14.</b> Les déterminants de la mortalité (2/13).....	158
<b>Annexe 15.</b> Les déterminants de la mortalité (3/13).....	159
<b>Annexe 16.</b> Les déterminants de la mortalité (4/13).....	160
<b>Annexe 17.</b> Les déterminants de la mortalité (5/13).....	161
<b>Annexe 18.</b> Les déterminants de la mortalité (6/13).....	162
<b>Annexe 19.</b> Les déterminants de la mortalité (7/13).....	163
<b>Annexe 20.</b> Les déterminants de la mortalité (8/13).....	164
<b>Annexe 21.</b> Les déterminants de la mortalité (9/13).....	165
<b>Annexe 22.</b> Les déterminants de la mortalité (10/13) .....	166
<b>Annexe 23.</b> Les déterminants de la mortalité (11/13) .....	167
<b>Annexe 24.</b> Les déterminants de la mortalité (12/13) .....	168
<b>Annexe 25.</b> Les déterminants de la mortalité (13/13) .....	169



## Annexe 1. PALOMB English questionnaire 1/3

Questions	Items
Date of visit	
ID number	
gender	
Date of birth	
Address (postal code)	Outside France, Country
Occupation	1/Farmers 2/craftsmen 3/managers 4/intermediate occupation 5/employees 6/workers 7/retired 8/no professional activity
Habits	home / institution
Location of birth	City/department/country
Family status	couple/unmarried/divorced/ widow
Social Insurance	
Adressed by	
<b>CLINIC</b>	
Height/ Weight	(cm)/(Kg)
Blood pressure	(max/min mmHg)
Pulse	(nb pulsations/min)
<b>Risk factor</b>	
Smoking <i>Number pack years</i>	yes/no if yes, past smoker or active smoker
Minimal advice	Yes/no
e-cigarette	Yes/no
<b>Environment</b>	
Job exposure	
Genetic predisposition	alpha 1 antitrypsin deficiency
<b>Clinical examination</b>	
<b>Purpose of the visit*</b>	
Year of first symptoms	
Cough*	Yes/no
Phlegm*	Yes/no
<b>Dyspnea/MRC</b>	
Chest pain *	Yes/no
Exacerbations*	(nb/year)
Daily time of walking	(min)
<b>Pulmonary examination</b>	
<b>Spirometry</b>	
FEV1 (post bronchodilator)*	(L)
LVC or FVC*	(L)
TPC	(% predicted)

## Annexe 2. PALOMB English questionnaire 2/3

RV	(% predicted)
TLCO/VA	(% predicted)
IC	(% predicted)
<b>Blood gases</b>	
SaO2	(%)
PaO2	(mmHg)
PaCO2	(mmHg)
<b><u>Exercise</u></b>	
6 minutes walking test	(meters)
<b>COMORBIDITIES</b>	
OAS	Yes/no
Asthma	Yes/no
bronchiectasis	Yes/no
Cancer	Yes/no
<b><u>Cardio-vascular</u></b>	
Hypertension	Yes/no
Ischemic cardiopathy	Yes/no
Heart rhythm disorder	Yes/no
Left cardiac insufficiency	Yes/no
Atheroma	Yes/no
Vascular Stroke	Yes/no
Pulmonary hypertension	Yes/no
Arteriopathy	Yes/no
<b><u>METABOLIC</u></b>	
Diabetes	Yes/no
Dyslipidemia	Yes/no
Metabolic syndrome	Yes/no
Denutrition	Yes/no
Osteoporosis	Yes/no
Depression	Yes/no
Anxiety	Yes/no
Rhinitis/rhinosinusitis	Yes/no
<b>Treatment</b>	
<b><u>COPD medication</u></b>	<b>Link with VIDAL database</b>
Compliance	Good/moderate/low
Oxygen therapy	Long term /Nocturnal /during waking/short term
Non invasive ventilation	Yes/no

### Annexe 3. PALOMB English questionnaire 3/3

---

Continue Positive Pressure	Yes/no
Smoking cessation management	Yes/no
NRT use	Yes/no
Physiotherapy breathing	Yes/no
Therapeutic education	Yes/no
Respiratory rehabilitation	Yes/no
<b>PREVENTION</b>	
Influenza and pneumococcal Vaccinations	Yes/no
Psychological support	Yes/no
Exacerbations : action plan	Yes/no ; written/oral
Physical activity	Yes/no

---

Next visit	15 days/1month/3months/6months/1year
------------	--------------------------------------

---

Annexe 4. Les déterminants de la dyspnée selon l'échelle mMRC (1/3)

**SUPPLEMENTARY INFORMATION**

**The determinants of dyspnoea evaluated by the mMRC scale: The French Palomb cohort.**

El Hassane Ouaalaya,<sup>1</sup> Laurent Falque,<sup>3</sup> Jean Michel Dupis,<sup>4</sup> Marielle Sabatini,<sup>5</sup> Alain Bernady,<sup>6</sup> Laurent Nguyen,<sup>7</sup> Annaig Ozier,<sup>7</sup> Cécilia Nocent-Ejnaini,<sup>5</sup> Frédéric Le Guillou,<sup>8</sup> Mathieu Molimard,<sup>9</sup> Chantal Raherison-Semjen<sup>1, 2</sup>

<sup>1</sup>Bordeaux University, INSERM, Bordeaux Population Health Research Center, team: EPICENE, UMR1219, Bordeaux, France. <sup>2</sup>CHU de Bordeaux, Pole cardiothoracique, Respiratory Diseases Department, Bordeaux, France. <sup>3</sup>Cabinet Medical, Avenue Thiers, Bordeaux, France. <sup>4</sup>Cabinet Medical Pessac, Pessac, France. <sup>5</sup>CHG Côte Basque, Bayonne, France. <sup>6</sup>Medical Center TOKI EDER, Cambo les bains, France. <sup>7</sup>Polyclinique Saint-Augustin, Bordeaux, France. <sup>8</sup>Cabinet Medical, la Rochelle, France. <sup>9</sup>Univ. Bordeaux, INSERM, Bordeaux Population Health Research Center, team Pharmaco-epidemiology, UMR 1219, Bordeaux, France.

**The list of studied comorbidities**

Hypertension, Dyslipidemia, ischemic cardiopathy, Apnea syndrome, anxiety, cancer, depression, diabetes, arrhythmia, asthma, left heart failure, bronchiectasis, atheroma, stroke, pulmonary arterial hypertension, lower limb arteriopathy, embolic thromboembolism, metabolic syndrome, undernutrition, osteoporosis, and sinus manifestation.

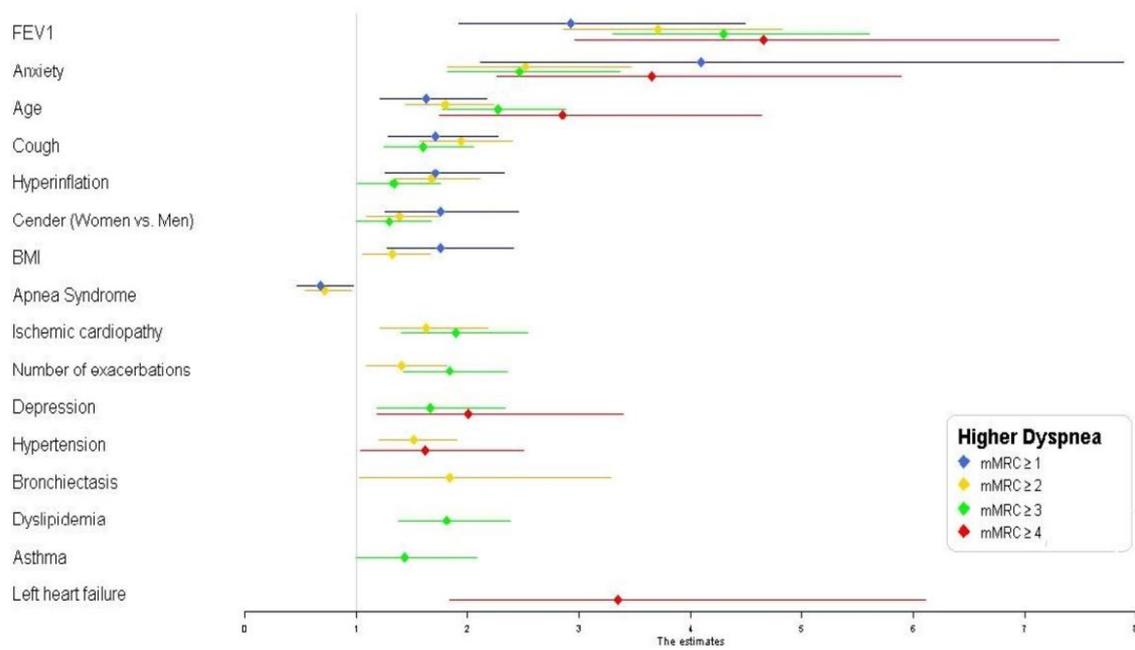
**Annexe 5. Les déterminants de la dyspnée selon l'échelle mMRC (2/3)**

49 **Table S1** Static hyperinflation in COPD patients by dyspnoea grade (mMRC)  
50

Cut-offs for static hyperinflation, n (%)	Dyspnoea mMRC « 0 » n=221	Dyspnoea mMRC « 1 » n=571	Dyspnoea mMRC « 2 » n=627	Dyspnoea mMRC « 3 » n=425	Dyspnoea mMRC « 4 » n=106	P Value
Without hyperinflation, RV/TLC <130%, 806 (40.85)	142(64.25)	294(51.48)	225(35.88)	112(26.35)	26(24.52)	<b>&lt;0.0001</b>
Mild hyperinflation, 130% ≤ RV/TLC <220%, 1150 (58.29)	77(34.84)	274(47.98)	398(63.47)	308(72.47)	78(73.58)	
Severe hyperinflation, 220% ≤ RV/TLC, 17 (0.86)	2(0.92)	3(0.52)	4(0.63)	5(1.17)	2(1.88)	

51  
52  
53  
54  
55

**Fig. S1** Forest plot of odds ratios and 95% confidence intervals.



56  
57  
58  
59  
60  
61

## Annexe 6. Les déterminants de la dyspnée selon l'échelle mMRC (3/3)

62  
63  
64  
65

**Table S2** Multivariate analysis (adjusted for age, gender and FEV<sub>1</sub>) indicating variables associated with higher dyspnoea, considering mMRC more than 1, 2, 3 and 4.

	OR [95% CI]			
	Higher dyspnoea (mMRC grade ≥1)	Higher dyspnoea (mMRC grade ≥2)	Higher dyspnoea (mMRC grade ≥3)	Higher dyspnoea (mMRC grade ≥4)
Anxiety	4.07 [2.11 – 7.84]	2.58 [1.86 – 3.56]	2.55 [1.87 – 3.47]	3.67 [2.27 – 5.93]
Cough	1.86 [1.39 – 2.49]	1.98 [1.60 – 2.45]	1.63 [1.28 – 2.09]	-----
Hyperinflation	1.76 [1.29 – 2.39]	1.69 [1.34 – 2.12]	1.33 [1.01 – 1.76]	-----
BMI	1.53 [1.14 – 2.06]	1.34 [1.07 – 1.68]	-----	-----
Apnea Syndrome	-----	-----	-----	-----
Ischemic cardiopathy	-----	1.62 [1.21 – 2.17]	1.75 [1.30 – 2.36]	-----
Number of exacerbations during the past year	-----	1.42 [1.10 – 1.82]	1.81 [1.41 – 2.33]	-----
Depression	-----	-----	1.67 [1.19 – 2.34]	2.08 [1.21 – 3.58]
Hypertension	-----	1.51 [1.20 – 1.89]	-----	1.60 [1.04 – 2.48]
Bronchiectasis	-----	1.87 [1.04 – 3.34]	-----	-----
Dyslipidemia	-----	-----	1.75 [1.33 – 2.31]	-----
Asthma	-----	-----	-----	-----
Left heart failure	-----	-----	1.65 [1.01 – 2.68]	3.27 [1.79 – 5.95]

66  
67  
68  
69

**Table S3** Summary of the differences in between the subsample (n=421) and the rest of the sample (n=1,552).

Subjects characteristics, n (%)		Subsample n=421 (21.33)	Rest of the sample n=1,552 (78.66)	P Value
Variable names	Data in all patients			
Males, n (%)	1,289(65.33)	250 (59.38)	1,039 (66.94)	0.4457
Age, yrs. mean(SD)	66.1 (10.98)	66.77 (10.54)	65.96 (11.08)	0.1818
BMI, (kg/m <sup>2</sup> ), mean(SD)	26.69 (10.81)	27.05 (6.08)	26.61 (11.63)	0.3010
<21 (underweight), n (%)	331(16.78)	56 (13.30)	275 (17.71)	0.0990
[21-26](normal), n (%)	683(34.62)	123 (29.21)	560 (36.08)	
[26-29](overweight), n (%)	396(20.07)	92 (21.85)	304 (19.58)	
≥29 (obese), n (%)	563(28.54)	102 (24.22)	461 (29.70)	
FEV <sub>1</sub> , % pred, Mean (SD)	58.11 (19.52)	57.17 (19.49)	58.32 (19.51)	0.3045
Chronic cough, n (%)	1,124(56.97)	227 (53.96)	897 (57.88)	0.0921
Chronic sputum, n (%)	795(40.33)	167 (39.66)	628 (40.46)	0.0502
Current smokers, n (%)	793(40.24)	129 (30.64)	664 (42.78)	0.0142
0-1 exacerbation, n (%)	1,427(72.33)	263 (62.47)	1,164 (75)	0.3837
≥ 2 exacerbations, n (%)	546(27.67)	158 (37.52)	388 (25)	
0 – 1 Dyspnoea mMRC	792 (40.14)	125 (29.77)	667 (42.97)	0.0037
≥2 Dyspnoea mMRC	1,181 (59.86)	296 (70.38)	885 (57.02)	
Pulmonary rehabilitation, n (%)	157(8)	42 (9.97)	115 (7.49)	0.0089
Smoking cessation, n (%)	155(7.85)	69 (16.38)	86 (5.54)	<0.0001
Hyperinflation (RV/TLC≥130%), n (%)	1167(59.14)	218 (51.78)	949 (61.14)	0.7589

70

## Supplementary Data (for publication)

### Susceptibility to frequent exacerbation in COPD patients: impact of the exacerbations history, vaccinations and comorbidities?

El Hassane Ouaalaya,<sup>1</sup> Laurent Falque,<sup>3</sup> Jean Michel Dupis,<sup>4</sup> Marielle Sabatini,<sup>5</sup> Alain Bernady,<sup>6</sup> Laurent Nguyen,<sup>7</sup> Annaig Ozier,<sup>7</sup> Cécilia Nocent-Ejnaini,<sup>5</sup> Frédéric Le Guillou,<sup>8</sup> Mathieu Molimard,<sup>9</sup> Maéva Zysman,<sup>10</sup> Chantal Raherison-Semjen<sup>1,2</sup>

<sup>1</sup>Bordeaux University, INSERM, Bordeaux Population Health Research Center, team: EPICENE, UMR1219, Bordeaux, France. <sup>2</sup>CHU de Bordeaux, Pole cardiothoracique, Respiratory Diseases Departement, Bordeaux, France. <sup>3</sup>Polyclinique Bordeaux Rive Droite, Lormont, France. <sup>4</sup>Cabinet Medical Pessac, Pessac, France. <sup>5</sup>CHG Côte Basque, Bayonne, France. <sup>6</sup>Medical Center TOKI EDER, Cambo les bains, France. <sup>7</sup>Polyclinique Saint-Augustin, Bordeaux. France. <sup>8</sup>Cabinet Medical, la Rochelle, France. <sup>9</sup> Bordeaux University, INSERM, Bordeaux Population Health Research Center, team Pharmaco-epidemiology, UMR 1219, Bordeaux, France. <sup>10</sup> Bordeaux University, INSERM, Cardiothoracic research center of Bordeaux, team: 2, UMR1045, Bordeaux, France.

**Table A.1.** The determinants of frequent exacerbators ( $\geq 2$ /year) at inclusion, the first and the second year of follow-up in the 446 COPD patients.

Variables	Odds ratio (95% CI)		
	At inclusion	The first year of follow-up	The second year of follow-up
Anxiety	-----	2.13[1.02-4.46]	2.53[1.23-5.19]
Chronic cough	2.72 [1.67-4.42]	-----	-----
Chronic sputum	-----	-----	-----
Diabetes	-----	-----	-----
Atheroma	-----	-----	-----
FEV <sub>1</sub> (% pred.)	1.92[1.20-3.08]	-----	-----
Residual volume (RV $\leq$ 140% Vs RV $>$ 140%)	-----	-----	-----
Dyspnea (mMRC 0-1 Vs mMRC $\geq$ 2)	-----	-----	-----
Gender (Women vs. Men)	1.73[1.07-2.79]	-----	-----
Depression	2.25[1.22-4.15]	-----	-----
Exacerbations at inclusion	-----	3.53[2.02-6.20]	1.78[1.00-3.16]
Exacerbations during the first year	-----	-----	4.08[2.17-7.65]

## Annexe 8. Les phénotypes et les déterminants d'exacerbations (2/6)

**Table A.2.** Comparisons of comorbidities in COPD patients who completed the 3-year follow-up separated into Cluster A, B, C and D.

Frequency of Major Comorbidities	All COPD patients followed for 3 years, n=446	A n =261 (58.50%) Infrequent	B n =85 (19.05%) Frequent	C n=82 (18.38 %) Transient	D n=18 (4.03%) Frequent	P value
Hypertension, n (%)	179 (40.13)	110(42.15)	33(38.82)	28(34.15)	8(44.44)	0.6013
Dyslipidemia, n (%)	113 (25.34)	66(25.29)	18(21.18)	25(30.49)	4(22.22)	0.5682
Cancer, All causes, n (%)	95 (21.30)	50(19.16)	26(30.59)	17(20.73)	2(11.11)	0.1044
Ischemic cardiopathy, n (%)	74 (16.59)	52(19.92)	8(9.41)	11(13.41)	3(16.67)	0.1187
Apnea Syndrome, n (%)	139(31.17)	98(37.55)	15(17.65)	23(28.05)	3(16.67)	0.0025
Anxiety, n (%)	47(10.54)	13(4.98)	11(12.94)	17(20.73)	6(33.33)	<0.0001
Depression, n (%)	63(14.13)	26(9.96)	16(18.82)	13(15.85)	8(44.44)	0.0003
Diabetes, n (%)	44(9.87)	25(9.58)	9(10.59)	6(7.32)	4(22.22)	0.2881
Arrhythmia, n (%)	58 (13.00)	36(13.79)	11(12.94)	9(10.98)	2(11.11)	0.9191
Asthma, n (%)	62(13.90)	30(11.49)	17(20.00)	13(15.85)	2(11.11)	0.2325
Arteriopathy of the lower limbs, n (%)	27 (6.05)	16(6.13)	6(7.05)	5(6.09)	0(0.00)	0.8056
Manifestation of Sinus, n (%)	38 (8.52)	24(9.19)	8(9.41)	5(6.09)	1(5.56)	0.9045
Atheroma, n (%)	27(6.05)	13(4.98)	10(11.76)	4(4.88)	0(0.00)	0.0799
Bronchiectasis, n (%)	31 (6.95)	20(7.66)	7(8.24)	4(4.88)	0(0.00)	0.5105
Left heart failure, n (%)	24(5.38)	13(4.98)	4(4.71)	6(7.32)	1(5.56)	0.8583
Stroke, n (%)	20(4.48)	14(5.36)	3(3.53)	1(1.22)	2(11.11)	0.2089
Metabolic syndrome, n (%)	18 (4.04)	8(3.07)	5(5.88)	4(4.88)	1(5.56)	0.6502
Osteoporosis, n (%)	5(1.12)	2(0.77)	2(2.35)	1(1.22)	0(0.00)	0.6433
Digestive disorders, n (%)	8 (1.79)	5(1.91)	2(2.35)	0(0.00)	1(5.56)	0.1989
Pulmonary arterial hypertension, n (%)	7(1.57)	5(1.92)	1(1.18)	1(1.22)	0(0.00)	0.8873
Thromboembolism, n (%)	1(0.22)	0(0.00)	1(1.18)	0(0.00)	0(0.00)	0.2351
Undernutrition, n (%)	7(1.57)	1(0.38)	3(3.53)	2(2.44)	1(5.56)	0.0805
Densitometry, n (%)	2 (0.45)	1(0.38)	0(0.00)	1(1.22)	0(0.00)	0.6637
Cardiovascular comorbidities, n (%)	230 (51.57)	143(54.79)	37(43.53)	39(47.56)	11(61.11)	0.2152

## Annexe 9. Les phénotypes et les déterminants d'exacerbations (3/6)

**Table A.3.** Comparisons of medications used by COPD patients who completed the 3-year follow-up separated into Cluster A, B, C and D.

COPD Treatment	All COPD patients followed for 3 years, n=446	A n =261 (58.50%) Infrequent	B n =85 (19.05%) Frequent	C n=82 (18.38 %) Transient	D n=18 (4.03%) Frequent	P value
SABA, n (%)	203 (45.52)	106(40.61)	46(54.12)	41(50.00)	10(55.56)	0.0912
LABA, n (%)	150 (33.63)	79(30.27)	22(25.88)	39(47.56)	10(55.56)	0.0022
LAMA, n (%)	245 (54.93)	130(49.81)	53(62.35)	46(56.1)	16(88.89)	0.0045
LABA or LAMA, n (%)	311 (69.73)	164(62.84)	63(74.12)	66(80.49)	18(100.00)	0.0003
LABA and LAMA, n (%)	84 (18.83)	45(17.24)	12(14.12)	19(23.17)	8(44.44)	0.0154
Inhaled corticosteroids, n (%)	37 (8.30)	14(5.36)	10(11.76)	8(9.76)	5(27.78)	0.0037
Fixed combination (ICS + LABA), n (%)	128 (28.70)	65(24.90)	33(38.82)	26(31.71)	4(22.22)	0.0776
LABA and LAMA and ICS, n (%)	24 (5.38)	6(2.30)	8(9.41)	6(7.32)	4(22.22)	0.0004
Annual influenza vaccination, n (%)	147 (32.96)	60(22.99)	36(42.35)	40(48.78)	11(61.11)	<0.0001
Pneumococcal vaccination, n (%)	206 (46.19)	103(39.46)	48(56.47)	44(53.66)	11(61.11)	0.0080
Oxygen therapy, n (%)	49 (10.99)	16(6.13)	14(16.47)	15(18.29)	4(22.22)	0.0013
<b>Monotherapy</b>						
SABA, n (%)	28 (6.28)	23(8.81)	3(3.53)	2(2.44)	0(0.00)	0.0658
LABA, n (%)	33 (7.40)	22(8.43)	3(3.53)	8(9.76)	0(0.00)	0.2247
LAMA, n (%)	26 (5.83)	14(5.36)	5(5.88)	5(6.10)	2(11.11)	0.7943
Inhaled corticosteroids, n (%)	2 (0.45)	2(0.77)	0(0.00)	0(0.00)	0(0.00)	0.6999
<b>Dual Therapy</b>						
SABA + LABA, n (%)	26 (5.83)	7(2.68)	6(7.06)	12(14.63)	1(5.56)	0.0009
SABA + LAMA, n (%)	87 (19.51)	46(17.62)	25(29.41)	12(14.63)	4(22.22)	0.0651
SABA + ICS, n (%)	4 (0.90)	2(0.77)	1(1.18)	1(1.22)	0 (0.00)	0.9436
LABA + LAMA, n (%)	43 (9.64)	30(11.49)	3(3.53)	8(9.76)	2(11.11)	0.1935
LABA + ICS, n (%)	3 (0.67)	2(0.77)	1(1.18)	0 (0.00)	0 (0.00)	0.7929
<b>Triple Therapy</b>						
SABA + LABA + LAMA, n (%)	17 (3.81)	9(3.45)	1(1.18)	5(6.10)	2(11.11)	0.1393
SABA + LABA + ICS, n (%)	2 (0.45)	1(0.38)	0(0.00)	0(0.00)	1(5.56)	0.0102
SABA + LAMA + ICS, n (%)	2 (0.45)	1(0.38)	0(0.00)	1(1.22)	0(0.00)	0.6637
LABA + LAMA + ICS, n (%)	10 (2.24)	2(0.77)	3(3.53)	3(3.66)	2(11.11)	0.0151

## Annexe 10. Les phénotypes et les déterminants d'exacerbations (4/6)

**Table A.4.** Factors associated with the frequent exacerbator cluster (D) in patients with full 3-year follow-up.

	Odds ratio (95% CI)
Depression	5.09[1.90-13.64]
Annual influenza vaccination	3.15[1.17-8.43]

**Table A.5.** Comparisons of mortality in COPD patients who completed the 3-year follow-up separated into Cluster A, B, C and D.

Mortality	All COPD patients followed for 3 years, n=446	A n =261 (58.50%) Infrequent	B n =85 (19.05%) Frequent	C n=82 (18.38%) Transient	D n=18 (4.03%) Frequent	P value
Death globally, n (%)	55 (12.33)	27(10.34)	12(14.12)	15(18.29)	1(5.56)	0.1982
Mortality after 1 years, n (%)	3 (0.67)	1(0.38)	2(2.35)	0(0.00)	0(0.00)	0.3324
Mortality after 2 years, n (%)	12 (2.69)	5(1.92)	3(3.53)	4(4.88)	0(0.00)	
Mortality after 3 years, n (%)	40 (8.97)	21(8.05)	7(8.24)	11(13.41)	1(5.56)	

**Table A.6.** Comparisons of exacerbations and mMRC dyspnea Scale in COPD patients who completed the 3-year follow-up separated into Cluster A, B, C and D.

Variables	All COPD patients followed for 3 years, n=446	A n =261 (58.50%) Infrequent	B n =85 (19.05%) Frequent	C n=82 (18.38%) Transient	D n=18 (4.03%) Frequent	P value
<b>Exacerbations</b>						
0	204 (45.74)	182(69.73)	0(0.00)	22(26.83)	0(0.00)	<0.0001
1	121 (27.13)	79(30.27)	0(0.00)	41(50.00)	1(5.56)	
≥2	121 (27.13)	0(0.00)	85(100.00)	19(23.17)	17(94.44)	
<b>mMRC Dyspnea Scale</b>						
0	74 (16.59)	54(20.69)	8(9.41)	12(14.63)	0(0.00)	0.0163
1	110 (24.66)	72(27.59)	14(16.47)	20(24.39)	4(22.22)	
2	98 (21.97)	55(21.07)	24(28.24)	16(19.51)	3(16.67)	
3	118 (26.46)	54(20.69)	30(35.29)	27(32.93)	7(38.89)	
4	46 (10.31)	26(9.96)	9(10.59)	7(8.54)	4(22.22)	

## Annexe 11. Les phénotypes et les déterminants d'exacerbations (5/6)

**Table A.7.** Summary of the differences in between the subsample (n=446 patients who completed the 3-year follow-up) and the rest of the sample (n=486 patients lost to follow-up).

Subjects characteristics, n (%)		The Subsample n=446 (47.85)	The rest of the sample n=486 (52.15)	P Value
Variable names	Data in all patients N=932			
Males, n (%)	605 (64.91)	302 (67.71)	303 (62.35)	0.0863
Age, yrs. mean(SD)	66.22 (10.25)	66.95 (10.04)	65.56 (10.40)	0.4427
BMI, (kg/m <sup>2</sup> ), mean(SD)	27.14 (14.52)	27.11 (6.29)	27.17 (19.20)	0.9439
FEV1, % pred, Mean (SD)	57.53 (20.73)	58.84 (19.98)	56.33 (21.34)	0.0645
Chronic cough, n (%)	521 (55.90)	253 (56.73)	268 (55.14)	0.6269
Chronic sputum, n (%)	398 (42.70)	201 (45.07)	197 (40.53)	0.1623
Current smokers, n (%)	328 (35.19)	153 (34.30)	175 (36.01)	0.5865
≥ 2 exacerbations, n (%)	252 (27.04)	121 (27.13)	131 (26.95)	0.9520
≥2 Dyspnea mMRC	546 (58.58)	257 (57.62)	289 (59.47)	0.5685
Pulmonary rehabilitation, n (%)	62 (6.65)	32 (7.17)	30 (6.17)	0.5397
Smoking cessation, n (%)	90 (9.66)	48 (10.76)	42 (8.64)	0.2736
Hyperinflation (RV/TLC≥120%), n (%)	659 (70.71)	296 (66.37)	363 (74.69)	0.0053

## Annexe 12. Les phénotypes et les déterminants d'exacerbations (6/6)

**Table A.8.** Comorbidities associated with COPD stratified by infrequent (<2/yr) or frequent exacerbators (≥2 /yr).

Frequency of Major Comorbidities	All COPD, n=932	Infrequent exacerbators (<2), n=680 (72.96)	Frequent exacerbators (≥2), n= 252 (27.04)	P Value
Hypertension, n (%)	381 (40.88)	277 (40.74)	104 (41.27)	0.8828
Dyslipidemia, n (%)	239 (25.64)	173 (25.44)	66 (26.19)	0.8160
Cancer, All causes, n (%)	186 (19.96)	144 (21.18)	42 (16.67)	0.1260
Ischemic cardiopathy, n (%)	148 (19.96)	113 (16.62)	35 (13.89)	0.3114
Apnea Syndrome, n (%)	209 (22.42)	178 (26.18)	31 (12.30)	<0.0001
Anxiety, n (%)	109 (11.70)	49 (7.21)	60 (23.81)	<0.0001
Depression, n (%)	139 (14.91)	76 (11.18)	63 (25.00)	<0.0001
Diabetes, n (%)	110 (11.80)	71 (10.44)	39 (15.48)	0.0343
Arrhythmia, n (%)	118 (12.66)	85 (12.50)	33 (13.10)	0.8082
Asthma, n (%)	111 (11.91)	75 (11.03)	36 (14.29)	0.1728
Arteriopathy of the lower limbs, n (%)	58 (6.22)	47 (6.91)	11 (4.36)	0.3890
Manifestation of Sinus, n (%)	57 (6.11)	42 (6.17)	15 (5.95)	0.6061
Atheroma, n (%)	47 (5.04)	31 (4.56)	16 (6.35)	0.2673
Bronchiectasis, n (%)	53 (5.69)	37 (5.44)	16 (6.35)	0.5950
Left heart failure, n (%)	41 (4.40)	25 (3.68)	16 (6.35)	0.0772
Stroke (Cerebrovascular Accident), n (%)	39 (4.18)	28 (4.12)	11 (4.37)	0.8669
Metabolic syndrome, n (%)	39 (4.18)	26 (3.82)	13 (5.16)	0.3659
Osteoporosis, n (%)	21 (2.25)	11 (1.62)	10 (3.97)	0.0317
Digestive disorders, n (%)	16 (1.71)	13 (1.91)	3 (1.19)	0.6470
Pulmonary arterial hypertension, n (%)	13 (1.39)	9 (1.32)	4 (1.59)	0.7604
Thromboembolism, n (%)	7 (0.75)	3 (0.44)	4 (1.59)	0.0719
Undernutrition, n (%)	19 (2.04)	6 (0.88)	13 (5.16)	<0.0001
Densitometry, n (%)	13 (1.39)	7 (1.03)	6 (2.38)	0.1181
Cardiovascular comorbidities, n (%)	477 (51.18)	349 (51.32)	128 (50.79)	0.8857

**Annexe 13.** Les déterminants de la mortalité (1/13)

1 **Main determinants associated with mortality in COPD patients**

2 El Hassane Ouaalaya, Maéva Zysman, Laurent Falque, Jean Michel Dupis, Marielle Sabatini,

3 Alain Bernady, Laurent Nguyen, Annaig Ozier, Cécilia Nocent-Ejnaini, Frédéric Le Guillou,

4 Mathieu Molimard, Chantal Raherison-Semjen.

5

6 **Online Data Supplement**

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

## Annexe 14. Les déterminants de la mortalité (2/13)

### 26 **The list of studied comorbidities**

27 Hypertension, dyslipidemia, ischemic cardiopathy, apnea syndrome, anxiety, cancer,  
28 depression, diabetes, arrhythmia, asthma, left heart failure, bronchiectasis, atheroma, stroke,  
29 pulmonary arterial hypertension, lower limb arteriopathy, embolic thromboembolism,  
30 metabolic syndrome, undernutrition, osteoporosis, and sinus manifestation.

31 Patients with presence of cardiovascular comorbidities are patients how had at least one of  
32 the following comorbidities: Hypertension, ischemic cardiopathy, arrhythmia, left heart  
33 failure, atheroma, stroke, pulmonary arterial hypertension or lower extremity artery disease.

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

## Annexe 15. Les déterminants de la mortalité (3/13)

### 52 **Online data supplement Figure Legends**

53 **Supplementary Fig. 1: Kaplan–Meier analyses for assessing all-cause mortality at 5 years in**  
54 **the PALOMB COPD cohort according to the static hyperinflation.**

55

56 **Supplementary Fig. 2: Change in FEV1 (ml/year) between Visit 1 and Visit 2 by disease stage**  
57 **(GOLD 2007 and GOLD 2017). Each bar represents a change in FEV1 of 200 ml per year.**

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

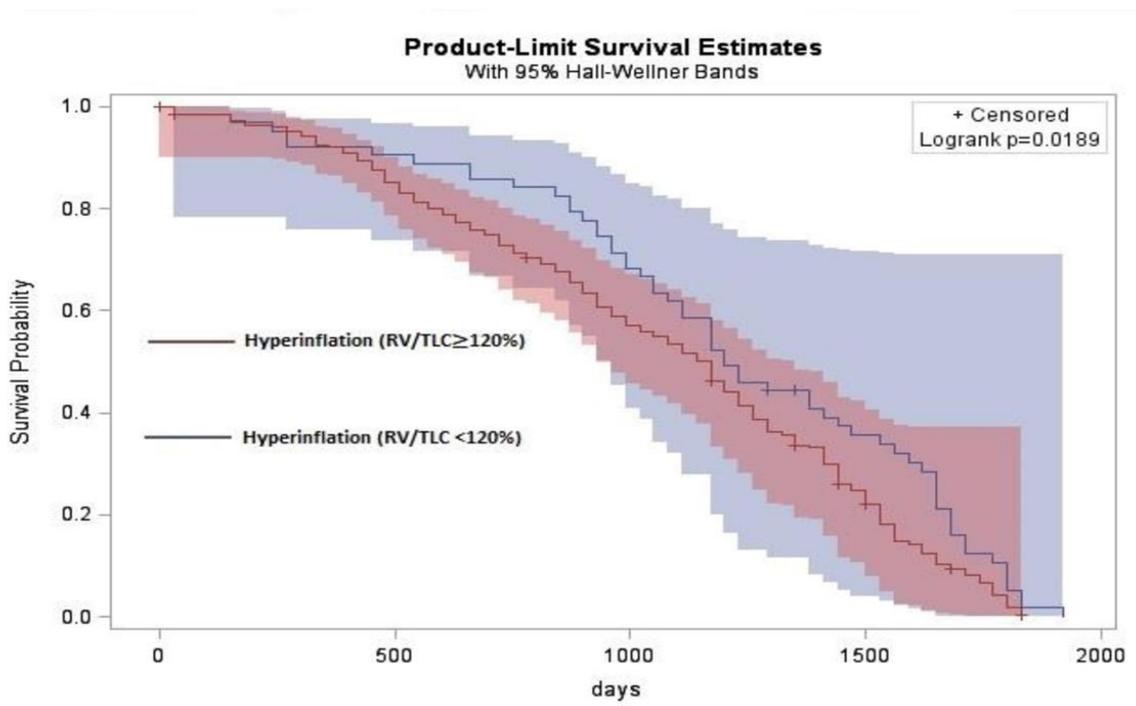
76

77

Annexe 16. Les déterminants de la mortalité (4/13)

78 **Supplementary Fig. 1**

79



80

81

82

83

84

85

86

87

88

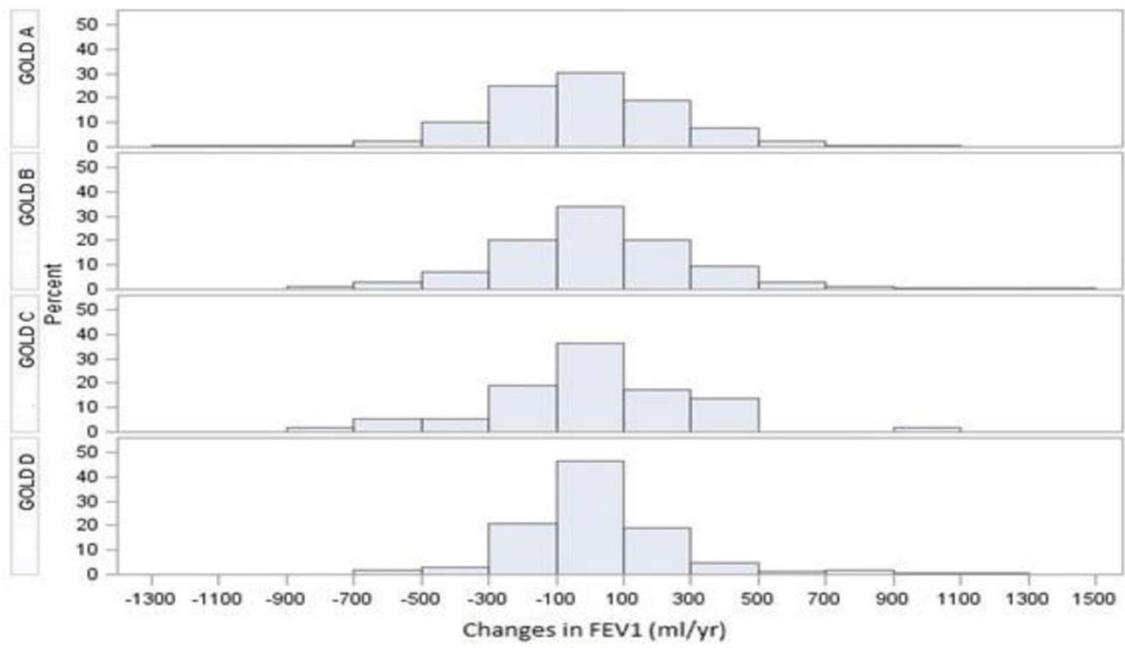
89

90

Annexe 17. Les déterminants de la mortalité (5/13)

91 **Supplementary Fig. 2**

92



93

94

95

96

97

98

99

100

101

102

103

**Annexe 18. Les déterminants de la mortalité (6/13)**

104 **Supplementary Table 1:** Association between patient follow-up time in 2,653 subjects with  
 105 COPD and death.

106

	All COPD, n=2,653	Deceased patients, n=339 (12.78)	live patients, n=2,314 (87.22)	P Value
Patients who had at least one follow-up visit, n(%)	932 (35.13)	108 (31.86)	824 (35.61)	0.1767
Patient follow-up time (months), mean(SD)	29.13 (18.13)	21.94 (13.21)	30.19 (18.51)	<0.0001
1-12 months, n(%)	1119 (42.18)	183 (53.98)	936 (40.45)	<0.0001
13-24 months, n(%)	437 (16.47)	77 (22.71)	360 (15.56)	
25-36 months, n(%)	364 (13.72)	43 (12.68)	321 (13.87)	
37-48 months, n(%)	308 (11.61)	26 (7.67)	282 (12.19)	
49-64 months, n(%)	425 (16.02)	10 (2.95)	415 (17.93)	

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

**Annexe 19. Les déterminants de la mortalité (7/13)**

123 **Supplementary Table 2:** gender differences and mortality rate in 2,653 COPD patients by the  
 124 number of comorbidities and the cardiovascular comorbidities.

125

	Deceased patients, n=339 (12.78)		Live patients, n=2,314 (87.22)		P Value
	Males, 254 (74.93)	Females, 85 (25.07)	Males, 1,451 (62.71)	Females, 863 (37.29)	
<b>Number of comorbidities 0 – 1, n (%), 1,136 (42.82)</b>	85 (25.07)	33 (9.73)	573 (24.67)	445 (19.23)	<b>0.0010</b>
<b>Number of comorbidities ≥2, n (%), 1,517 (57.12)</b>	169 (49.85)	52 (15.33)	878 (37.94)	418 (18.06)	<b>0.0095</b>
<b>Cardiovascular comorbidities, n (%), 1,277 (48.1)</b>	158 (46.61)	45 (13.27)	776 (33.53)	298 (12.87)	0.1000

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

**Annexe 20. Les déterminants de la mortalité (8/13)**

142 **Supplementary Table 3:** Determinants of Mortality adjusted by gender and Age in 2,653 COPD  
 143 patients.

	Multivariate Analysis adjusted for :	
	Gender	Gender and Age
<b>Lung cancer</b>	4.81 [3.36 - 6.89]	4.81 [3.36 - 6.89]
<b>Metabolic syndrome</b>	2.47 [1.45 - 4.22]	2.47 [1.45 - 4.21]
<b>Smoking cessation</b>	2.42 [1.32 - 4.40]	2.41 [1.32 - 4.40]
<b>Age</b>	1.94 [1.50 - 2.51]	-----
<b>FEV<sub>1</sub></b>	1.85 [1.41 - 2.42]	1.85 [1.41 - 2.42]
<b>Depression</b>	1.81 [1.26 - 2.58]	1.80 [1.26 - 2.58]
<b>Dyspnea</b>	1.73 [1.30 - 2.30]	1.73 [1.30 - 2.30]
<b>Atheroma</b>	1.60 [1.03 - 2.48]	1.60 [1.03 - 2.48]
<b>Hyperinflation</b>	1.55 [1.11 - 2.16]	1.55 [1.11 - 2.16]
<b>Ischemic cardiopathy</b>	1.37 [1.01 - 1.86]	1.37 [1.01 - 1.86]

144

145

146

147

148

149

150

151

152

153

154

155

156

**Annexe 21. Les déterminants de la mortalité (9/13)**

157 **Supplementary Table 4:** Decline in FEV1 (litre and % pred.) between the inclusion and the  
 158 first visit of patient follow-up by mortality rate in 770 patients (Patients who had at least one  
 159 follow-up visit with complete data on FEV1).

160

	All COPD, n=770	Deceased patients, n=78 (10.13)	live patients, n=692 (89.87)	P Value
FEV1 at inclusion, litre, Mean (SD)	1.74 (0.71)	1.42 (0.60)	1.77 (0.71)	<0.0001
FEV1 at the first visit of follow-up, litre, Mean (SD)	1.73 (0.72)	1.37 (0.57)	1.77 (0.72)	<0.0001
FEV1 at inclusion, % pred, Mean (SD)	58.39 (21.05)	48.43 (16.78)	59.51 (21.20)	<0.0001
FEV1 at the first visit of follow-up, % pred, Mean (SD)	58.17 (21.99)	46.41 (15.77)	59.50 (22.21)	<0.0001
FEV1 decline, ml/yr, Mean (SD)	-4.96 (295.24)	-55.89 (280.80)	0.78 (296.47)	0.0966
FEV1 % decline of baseline values, Mean (SD)	1.01 (0.20)	0.97 (0.18)	1.01 (0.20)	0.1264

161

162

163

164

165

166

167

168

169

170

171

172

173

174

## Annexe 22. Les déterminants de la mortalité (10/13)

175 **Supplementary Table 5:** Summary of the differences in between the subsample (n=770) and  
 176 the rest of the sample (n=1,883).

177

Subjects characteristics, n (%)		Subsample n=770 (29.02)	Sample n=1,883 (70.98)	P Value
Variable names	Data in all patients N=2,653			
Males, n (%)	1,705 (64.27)	508 (65.97)	1,197 (63.57)	0.2406
Age, yrs. Mean (SD)	66.12 (10.87)	66.33 (10.24)	66.04 (11.12)	0.6166
BMI, (kg/m <sup>2</sup> ), mean (SD)	26.61 (11.11)	27.50 (15.70)	26.24 (8.54)	0.0006
<21 (underweight), n (%)	457 (17.23)	124 (16.10)	333 (17.68)	0.0058
[21-26](normal), n (%)	934 (35.21)	247 (32.08)	687 (36.48)	
[26-29](overweight), n (%)	509 (19.19)	144 (18.70)	365 (19.38)	
≥29 (obese), n (%)	753 (28.38)	255 (33.12)	498 (26.45)	
FEV1, % pred, Mean (SD)	58.32 (40.26)	58.39 (21.05)	58.30 (45.86)	0.3956
GOLD stage I, >80%, n (%)	382 (14.4)	116 (15.06)	266 (14.13)	0.0237
GOLD stage II, 50-80%, n (%)	1,383 (52.13)	369 (47.92)	1,014 (53.85)	
GOLD stage III, 30-50%, n (%)	706 (26.61)	233 (30.26)	473 (25.12)	
GOLD stage IV, <30%, n (%)	182 (6.86)	52 (6.75)	130 (6.90)	
Chronic cough, n (%)	1,537 (57.93)	430 (55.84)	1,107 (58.79)	0.1631
Chronic sputum, n (%)	1,105 (41.65)	345 (44.81)	760 (40.36)	0.0351
Current smokers, n (%)	1,030 (38.82)	279 (36.23)	751 (39.88)	0.0800
Dyspnea, 0-1mMRC, n (%)	1,120 (42.22)	331 (42.99)	789 (41.90)	0.6073
Dyspnea, ≥ 2 mMRC, n (%)	1,533 (57.78)	439 (57.01)	1,094 (58.10)	
Infrequent exacerbations (< 2), n (%)	1,971 (74.29)	558 (72.47)	1,413 (75.04)	0.1688
Frequent exacerbations (≥2), n (%)	682 (25.71)	212 (27.53)	470 (24.96)	
Pulmonary rehabilitation, n (%)	186 (7.01)	49 (6.36)	137 (7.28)	0.4037
Smoking cessation, n (%)	235 (8.86)	85 (11.04)	150 (7.97)	0.0115
Hyperinflation (RV/TLC≥130%), n (%)	1,902 (71.69)	529 (68.70)	1,373 (72.92)	0.0287

178

179

## Annexe 23. Les déterminants de la mortalité (11/13)

180 **Supplementary Table 6:** Characteristics and 5-year mortality rates in COPD patients recruited  
 181 in the PALOMB cohort according to the five classes identified using the CART-based algorithm.

182

Subjects characteristics,		Class 1 n =516 (19.45%)	Class 2 n =1,135 (42.78%)	Class 3 n =454 (17.11%)	Class 4 n =96 (3.62%)	Class 5 n =452 (17.04%)	P value
Variable names	All COPD patients, n=2,653						
Males, n (%)	1,705 (64.27)	365 (70.74)	663 (58.41)	361 (79.52)	61 (63.54)	255 (56.42)	<0.0001
Age, yrs. Mean(SD)	66.12 (10.87)	73.08 (9.80)	63.77 (9.44)	70.6 (9.80)	64.08 (9.20)	60.02 (10.90)	<0.0001
BMI, (kg/m <sup>2</sup> ), Mean (SD)	26.61 (11.11)	27.7 (6.40)	25.4 (15.32)	30.66 (5.87)	23.21 (5.48)	25.04 (4.61)	<0.0001
<21 (underweight), n (%)	457 (17.23)	59 (11.34)	257 (22.64)	16 (3.52)	42 (43.75)	83 (18.36)	<0.0001
[21-26](normal), n (%)	934 (35.21)	176 (34.11)	444 (39.12)	91 (20.04)	29 (30.21)	194 (42.92)	
[26-29](overweight), n (%)	509 (19.19)	99 (19.19)	233 (20.53)	66 (14.54)	12 (12.50)	99 (21.90)	
≥29 (obese), n (%)	753 (28.38)	182 (35.27)	201 (17.71)	281 (61.89)	13 (13.54)	76 (16.81)	
FEV1, % pred, Mean (SD)	58.32 (40.26)	47.76 (18.17)	55.54 (56.68)	64.93 (13.65)	27.51 (5.06)	77.28 (11.48)	<0.0001
GOLD stage I, >80%, n (%)	382 (14.40)	31 (6.01)	126 (11.10)	58 (12.78)	0 (0)	167 (36.95)	<0.0001
GOLD stage II, 50-80%, n (%)	1,383 (52.13)	161 (31.20)	582 (51.28)	355 (78.19)	0 (0)	285 (63.05)	
GOLD stage III, 30-50%, n (%)	706 (26.61)	243 (47.09)	388 (34.19)	39 (8.59)	36 (37.50)	0 (0)	
GOLD stage IV, <30%, n (%)	182 (6.86)	81 (15.70)	39 (3.44)	2 (0.44)	60 (62.50)	0 (0)	
Dyspnea, ≥ 2 mMRC, n (%)	1,533 (57.78)	478 (92.64)	737 (64.93)	222 (48.90)	96 (100)	0(0)	<0.0001
≥ 2 exacerbations, n (%)	682 (25.71)	177 (34.30)	294 (25.90)	85 (18.72)	44 (45.83)	82 (18.14)	<0.0001
Any cardiovascular comorbidity or diabetes, n (%)	1,303 (49.11)	516 (100)	333 (29.34)	454 (100)	0(0)	0(0)	<0.0001
5-year mortality, n (%)	339 (12.78)	114 (22.09)	128 (11.28)	50 (11.01)	28 (29.17)	19 (4.20)	<0.0001
Age at death years, Mean(SD)	69.95 (10.22)	74.39 (8.31)	66.12 (10.22)	72.56 (9.24)	68.58 (7.83)	63.25 (13.01)	<0.0001

183

184

185

186

187

188

189

## Annexe 24. Les déterminants de la mortalité (12/13)

190 **Supplementary Table 7:** Comorbidities associated with COPD patients recruited in the  
 191 PALOMB cohort according to the five classes identified using the CART-based algorithm.

192

Frequency of Major Comorbidities	All COPD, n=2,653	Class 1 n =516 (19.45%)	Class 2 n =1,135 (42.78%)	Class 3 n =454 (17.11%)	Class 4 n =96 (3.62%)	Class 5 n =452 (17.04%)	P value
Hypertension, n (%)	969 (36.52)	373 (72.29)	240 (21.15)	356 (78.41)	0 (0)	0 (0)	<0.0001
Dyslipidemia, n (%)	596 (22.47)	200 (38.76)	175 (15.42)	184 (40.53)	4 (4.17)	33 (7.30)	<0.0001
Lung cancer, n(%)	186 (7.01)	35 (6.78)	101 (8.90)	23 (5.07)	3 (3.13)	24 (5.31)	0.0108
Cancer, other causes, n (%)	320 (12.06)	76 (14.73)	132 (11.63)	58 (12.78)	5 (5.21)	49 (10.84)	0.0673
Ischemic cardiopathy, n (%)	454 (17.11)	224 (43.41)	92 (8.11)	138 (30.40)	0 (0)	0 (0)	<0.0001
Apnea Syndrome, n (%)	440 (16.58)	106 (20.54)	122 (10.75)	147 (32.38)	6 (6.25)	59 (13.05)	<0.0001
Anxiety,n (%)	375 (14.13)	134 (25.97)	145 (12.78)	37 (8.15)	23 (23.96)	36 (7.96)	<0.0001
Depression, n (%)	366 (13.80)	98 (18.99)	146 (12.86)	43 (9.47)	26 (27.08)	53 (11.73)	<0.0001
Diabetes, n (%)	330 (12.44)	136 (26.36)	71 (6.26)	123 (27.09)	0 (0)	0 (0)	<0.0001
Arrhythmia, n (%)	310 (11.68)	121 (23.45)	85 (7.49)	81 (17.84)	4 (4.17)	19 (4.20)	<0.0001
Asthma, n (%)	298 (11.23)	57 (11.05)	145 (12.78)	34 (7.49)	13 (13.54)	49 (10.84)	0.0460
Arteriopathy of the lower limbs, n (%)	155 (8.21)	59 (11.43)	55 (4.85)	31 (6.83)	2 (2.08)	8 (1.77)	<0.0001
Manifestation of Sinus, n (%)	121 (6.41)	24 (4.65)	59 (5.20)	16 (3.52)	5 (5.21)	17 (3.76)	0.5668
Atheroma, n (%)	150 (5.65)	60 (11.63)	46 (4.05)	35 (7.71)	3 (3.13)	6 (1.33)	<0.0001
Bronchiectasis, n (%)	130 (4.90)	29 (5.62)	59 (5.20)	22 (4.85)	8 (8.33)	12 (2.65)	0.0876
Left heart failure, n (%)	119 (4.49)	85 (16.47)	9 (0.79)	25 (5.51)	0(0)	0 (0)	<0.0001
Stroke, n (%)	96(3.62)	30 (5.81)	28 (2.47)	21 (4.63)	3 (3.13)	14 (3.10)	0.0104
Metabolic syndrome, n (%)	92 (3.47)	36 (6.98)	12 (1.06)	41 (9.03)	0 (0)	3 (0.66)	<0.0001
Osteoporosis, n (%)	83 (3.13)	26 (5.04)	39 (3.44)	7 (1.54)	7 (7.29)	4 (0.88)	0.0001
Digestive disorders, n (%)	46 (2.44)	8 (1.55)	22 (1.94)	6 (1.32)	2 (2.08)	8 (1.77)	0.9238
Pulmonary arterial hypertension, n (%)	62 (2.34)	32 (6.20)	13 (1.15)	8 (1.76)	6 (6.25)	3 (0.66)	<0.0001
Thromboembolism, n (%)	25 (1.32)	10 (2.91)	10 (0.88)	4 (0.88)	0 (0)	1 (0.22)	0.0617
Undernutrition, n (%)	49 (1.85)	15 (2.91)	23 (2.03)	0 (0)	10 (10.42)	1 (0.22)	<0.0001
Densitometry, n (%)	41 (1.55)	14 (2.71)	20 (1.76)	3 (0.66)	2 (2.08)	2 (0.44)	0.0254

193

194

195

**Annexe 25. Les déterminants de la mortalité (13/13)**

196 **Supplementary Table 8:** The BOD indices according to the five classes identified using the  
 197 CART-based algorithm.

198

The BOD indices	All COPD, n=2,653	Class 1 n =516 (19.45%)	Class 2 n =1,135 (42.78%)	Class 3 n =454 (17.11%)	Class 4 n =96 (3.62%)	Class 5 n =452 (17.04%)	P value
<b>BOD. Median (interquartile range)</b>	2 (1 – 3)	4 (3 – 5)	2 (1 – 3)	1 (0 – 2)	6 (5 – 6)	0 (0 – 1)	<0.0001
<b>BOD (0 points), n (%)</b>	560 (21.11)	0 (0)	111 (9.78)	130 (28.63)	0 (0)	319 (70.58)	<0.0001
<b>BOD (1 point), n (%)</b>	558 (21.03)	6 (1.16)	259 (22.82)	165 (36.34)	0 (0)	128 (28.32)	
<b>BOD (2 points), n (%)</b>	514 (19.37)	84 (16.28)	302 (26.61)	123 (27.09)	0 (0)	5 (1.11)	
<b>BOD (3 points), n (%)</b>	447 (16.85)	162 (31.4)	254 (22.38)	31 (6.83)	0 (0)	0 (0)	
<b>BOD (4 points), n (%)</b>	271 (10.21)	109 (21.12)	157 (13.83)	5 (1.10)	0 (0)	0 (0)	
<b>BOD (5 points), n (%)</b>	199 (7.50)	105 (20.35)	50 (4.41)	0 (0)	44 (45.83)	0 (0)	
<b>BOD (6 points), n (%)</b>	82 (3.09)	40 (7.75)	2 (0.18)	0 (0)	40 (41.67)	0 (0)	
<b>BOD (7 points), n (%)</b>	22 (0.83)	10 (1.94)	0 (0)	0 (0)	12 (12.50)	0 (0)	

199

**Contexte :** La BPCO est une pathologie hétérogène responsable d'une morbidité et d'une mortalité importante. La classification GOLD de la sévérité de la BPCO introduit des incertitudes, son applicabilité dans la pratique est plus que douteuse. Les exacerbations et les comorbidités contribuent à la sévérité globale de l'affection, au complexe des symptômes et à l'évolution de la maladie. L'évaluation de la dyspnée en pratique clinique n'est pas chose aisée, et il est difficile d'apprécier la part attribuable respiratoire de la dyspnée chez des patients ayant des comorbidités. Les données actuelles restent insuffisantes pour affirmer ou infirmer l'importance de la majorité des candidats phénotypes.

**Objectifs :** Principalement, ce travail a pour buts de caractériser l'évolution des trajectoires des patients BPCO et d'identifier des candidats phénotypes chez des patients BPCO. Secondairement, ce travail a pour buts : d'identifier les clusters stables de comorbidités, de caractériser les déterminants de la dyspnée, de déterminer les facteurs associés aux phénotypes des Exacerbateurs fréquents et d'évaluer l'impact des symptômes, des fonctions pulmonaires et des comorbidités sur la mortalité.

**Méthodes :** Les patients BPCO (VEMS/CVF<70% post-BD) ont été inclus depuis Janvier 2014 par leur pneumologue en utilisant le questionnaire de la cohorte PALOMB. Ce questionnaire comprenait les domaines suivants : les critères démographiques, les symptômes cliniques, les tests fonctionnels, les comorbidités, les vaccinations et les traitements inhalés. Après cinq ans de suivi, les données sur leur statut vital ont été obtenues en utilisant le RNIPP. L'identifier des clusters stables de comorbidités a été réalisé à partir des modèles de classification supervisée sur les composantes principales. L'analyse de sensibilité et l'estimateur des moindres carrés pénalisé (modèle LASSO) ont ensuite été utilisés afin d'évaluer les déterminants de la dyspnée. Le suivi longitudinal de la fréquence des exacerbations a permis d'identifier trois phénotypes à partir d'une classification non-supervisée. De plus, Le modèle de Cox a été utilisé afin d'identifier les facteurs associés à la mortalité toute cause. Les arbres de classification et de régression (CART) ont été utilisés pour répartir les patients dans des sous-groupes et la pertinence clinique a été déterminée en comparant la mortalité à 5 ans. L'analyse en cluster la plus optimale (supervisé vs non-supervisé) a été utilisée pour identifier de façon robuste les phénotypes cliniques.

**Résultats :** L'analyse en cluster a permis de différencier 5 clusters de comorbidités : cardiovasculaire ; syndrome métabolique et SAOS ; dénutrition ; patients avec bronchectasie et un cluster de patients ayant moins de comorbidités. Nos résultats suggèrent que la dyspnée est liée à la sévérité de la BPCO, au sexe, aux exacerbations, aux comorbidités et à l'hyperinflation statique. Révélant que l'hyperinflation statique définie par le rapport IC/TLC était un meilleur déterminant de l'échelle de dyspnée mMRC comparativement aux rapports RV/TLC et FRC/TLC. Sur la base de notre hypothèse, nous avons pu déterminer quatre phénotypes : le phénotype A (Exacerbateurs non fréquents), B (exacerbations fréquentes chez les patients en sous poids), C (transitoires) et D (exacerbations fréquentes chez les patients obèses). Les déterminants les plus importants du phénotype exacerbateur fréquent sont les antécédents d'exacerbations, l'absence de vaccination antigrippale annuelle et l'anxiété. Le cancer et les comorbidités cardiovasculaires contribuent à l'augmentation de la mortalité, en plus des critères connus en rapport avec la sévérité de la maladie.

**Conclusion :** Nos résultats montrent que la multimorbidité est fréquente, hétérogène et varie en fonction du stade de sévérité de la BPCO. Ils illustrent l'aspect plurifactoriel de la dyspnée chez les patients BPCO, confirment l'existence et la pertinence clinique d'un phénotype d'Exacerbateurs fréquents de patients BPCO.

**Background:** COPD is a heterogeneous and multisystemic disease with progressive increasing morbidity and mortality. The GOLD classification introduces uncertainties, its validity in predicting mortality is still uncertain, and its applicability in practice is more than questionable. Our understanding of the longitudinal determinants of COPD exacerbations remains unclear. COPD patients frequently suffer from comorbidities, with a significant impact on mortality. Evaluating dyspnea in clinical practice is not easy, and it is difficult to assess the respiratory attributable part of dyspnea in patients with comorbidities. Current data remains insufficient to confirm the importance of the majority of candidate phenotypes.

**Research aims:** Mainly, this work aims to characterize the evolution of the trajectories of COPD patients and to identify candidate phenotypes in COPD patients. Secondly, this work aims to: identify stable clusters of comorbidities, characterize the determinants of dyspnoea, determine the factors associated with the phenotypes of frequent exacerbators and assess the impact of symptoms, pulmonary functions and comorbidities on mortality.

**Methods:** Diagnosis of COPD was made using spirometry with post-BD FEV<sub>1</sub>/FVC <70%. Since January 2014, pulmonologists have been using the PALOMB web site questionnaire to include COPD patients. This questionnaire included the following domains: demographic criteria, clinical symptoms, lung function, comorbidities and therapeutic management. After 5 years of follow-up, vital status was recorded through the RNIPP. Stable clusters of comorbidities were identified using the supervised classification models on the principal components. The sensitivity analysis and the least absolute shrinkage and selection operator (LASSO) were then used to assess the determinants of dyspnea. The unsupervised classification was used to identify three phenotypes by modelling the longitudinal changes in the frequency of COPD exacerbations. In addition, the Cox model was used to identify factors associated with all-cause mortality. Classification and regression trees (CART) were used for allocating patients to the subgroups and clinical relevance was determined by comparing 5-year mortality. The most optimal cluster analysis (supervised vs unsupervised) was used to robustly identify clinical phenotypes.

**Results:** The cluster analysis showed five phenotypes of comorbidities: cluster 1 included cardiac profile; cluster 2 included less comorbidities; cluster 3 included metabolic syndrome, apnea and anxiety-depression; cluster 4 included malnutrition and osteoporosis and cluster 5 included bronchiectasis. Regardless of the cut-off values set to identify higher dyspnoea, our results suggested that dyspnoea is related to the severity of airflow limitation, gender, exacerbations, comorbidities and hyperinflation. Revealing that the hyperinflation defined by IC/TLC ratio was a better determinant of the mMRC dyspnoea scale in comparison to both RV/TLC and FRC/TLC ratio. Based on our hypothesis, we were able to determine four phenotypes: A (infrequent), B (frequent in underweight patients), C (transient), and D (frequent in obese patients). The most important determinants of the frequent exacerbating phenotype are: the presence of anxiety, chronic sputum, and unvaccinated against influenza. In addition to known criteria related to COPD severity, cancer and cardiovascular comorbidities contribute to increased mortality in COPD patients.

**Conclusion:** Our results show that within the Palomb cohort (a large clinic-based cohort) multimorbidity is frequent, heterogeneous and varies according to the stage of COPD severity. The presence of comorbidities should therefore be included in any assessment of COPD severity. These cohort results illustrate in real-life the multifactorial aspect of dyspnoea in COPD patients, confirm the existence and clinical relevance of two frequent exacerbators' phenotypes and the currently used threshold to define this phenotype.