


BMJ Open Study protocol: pneumonia and inhaled corticosteroid treatment patterns in chronic obstructive pulmonary disease – a cohort study using sequence analysis (PICCS)

Allan Klitgaard ^{1,2}, Rikke Ibsen,³ Ole Hilberg,^{1,2} Anders Løkke^{1,2}

To cite: Klitgaard A, Ibsen R, Hilberg O, *et al.* Study protocol: pneumonia and inhaled corticosteroid treatment patterns in chronic obstructive pulmonary disease – a cohort study using sequence analysis (PICCS). *BMJ Open* 2023;**13**:e072685. doi:10.1136/bmjopen-2023-072685

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2023-072685>).

Received 10 February 2023
Accepted 18 May 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Internal Medicine, Lillebaelt Hospital, Vejle, Denmark

²Department of Regional Health Research, University of Southern Denmark, Odense, Denmark

³12minds, Aarhus, Denmark

Correspondence to

Dr Allan Klitgaard;
alksorensen@health.sdu.dk

ABSTRACT

Introduction Treatment with inhaled corticosteroids (ICS) is a widely used treatment in chronic obstructive pulmonary disease. The main effects include a reduction in the number of exacerbations and, for some patients, an increase in expected mortality. Unfortunately, the treatment is also linked to an increased risk of pneumonia, and very little is known about which patients experience this increased risk. There is a need for identification of patient characteristics associated with increased risk of pneumonia and treatment with ICS.

Methods and analysis This is a register-based cohort study that uses the nationwide Danish registers. Data from several registers in the years 2008–2018 will be merged on an individual level using the personal identification numbers that are unique to every citizen in Denmark. Clusters based on pneumonia incidence and ICS treatment patterns will be explored with a sequence analysis in a 3-year follow-up period.

Ethics and dissemination This is a register-based study and research ethics approval is not required according to Danish Law and National Ethics Committee Guidelines. The results will be submitted to peer-reviewed journals and reported at appropriate national and international meetings.

INTRODUCTION

Inhaled corticosteroid (ICS) treatment is one of the medical treatments used in chronic obstructive pulmonary disease (COPD).^{1–2} The main effect of ICS is a reduction in acute exacerbations of COPD, and the treatment has shown statistically significant effects on quality of life and mortality.³ However, the treatment is associated with side effects such as increased risk of oral candidiasis, diabetes, osteoporosis, and cataract.^{4–9} Arguably, the most important side effect is the increased risk of pneumonia,^{7–10} as reflected in the Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease (GOLD) 2020 report.¹

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ A limitation of register-based studies is that relevant variables are derived from already predetermined data, and it is difficult to add other data for individual study purposes. The definitions of moderate pneumonia in our study, for instance, is constructed from redeemed prescriptions of medication typically used for these events.
- ⇒ This study is conducted on a large study population represented in the Danish registers of high quality. This ensures vast amounts of data, resulting in great statistical power and limited selection bias.
- ⇒ The registers have been used for many years, allowing us to study long-term trends in diseases.
- ⇒ The study design allows for exploration of associations between pneumonia and inhaled corticosteroid treatment over time.
- ⇒ The validity of the registers used in this study has been investigated and deemed high.

It is difficult to foresee which patients will benefit from ICS treatment and which patients will experience side effects. A Spanish study concluded that pulmonologists rarely agree on when to prescribe or withdraw from ICS.¹¹ However, during the past few years, a growing body of evidence has emerged to illuminate certain phenotypes within COPD that are more likely to benefit from ICS treatment: A history of frequent and/or severe exacerbations, a high level of blood eosinophils, and a history of asthma are all associated with an increased probability of beneficial ICS treatment effects.^{12–15} Unfortunately, not much is known about which patients with COPD that are more likely to experience an increased risk of pneumonia. In a pooled analysis of five clinical trials, a low body mass index was the only significant risk factor for ICS-induced pneumonia.¹⁶

COPD is recognised as a complex and heterogeneous disease, and patient phenotypes within COPD have been studied using various approaches.^{17 18} Currently accepted clinical phenotypes include the eosinophilic phenotype and the emphysema phenotype.¹⁹ Phenotypes within comorbidity have been described in cluster analysis studies.^{20 21} Phenotypes based on cluster analysis have been explored using cross-sectional data,^{20 22 23} but no studies have explored phenotypes within COPD using cluster-analysis on longitudinal data. This may be achieved using sequence analysis, and a study from Scotland did this to understand multimorbidity.²⁴ Sequence analysis is a non-parametric method commonly used in the social sciences to analyse trajectories and social processes.^{25 26} The method has the advantage of providing a holistic view of trajectories, describing how processes evolve over time and when transitions occur.

We hypothesise that, using sequence analysis, individual disease trajectories within COPD can be divided into groups forming typical trajectories based on how similar they are (clusters). The purpose of our present study is to explore if patients with COPD can be grouped in clusters according to ICS treatment patterns and pneumonia incidence.

METHODS AND ANALYSIS

Study design and research questions

This study is a nationwide cohort study from 2008 to 2018. The study will use data from the Danish national databases, and these data will be merged using a unique personal identification number (PIN) that is assigned to all citizens in Denmark from the Danish Civil Registration

System (CRS). This creates a dataset that will be used to answer the following research questions:

1. Can patients with COPD be grouped according to their ICS treatment pattern and pneumonia incidence?
2. Is there a difference in patient characteristics between these groups?

Data sources

This is an entirely register-based study that uses data from the vast Danish national registries (figure 1). Data are collected from the CRS, the Danish National Patient Register (DNPR), The Danish National Prescription Register (NPR), Statistics Denmark (DST) and the Danish Register for Chronic Obstructive Pulmonary Disease (DrKOL).

The Civil Registration System

The CRS uses a unique PIN that has been assigned to every citizen in Denmark since 1968.²⁷ This PIN allows for individual-level linkage across all registers, as it is used for registration in all health registers. It has been validated, and it is considered a vital tool in Danish epidemiological research.²⁸

The Danish National Patient Register

DNPR contains information on patients' contact with Danish hospitals such as admissions and outpatient clinics. It contains information on all hospitalised somatic patients from 1977 and all outpatients, emergency room patients, and psychiatric patients from 1995. In 2019, a new data model was established. Variables include treatment, diagnosis by International Classification of Diseases 10th Revision (ICD-10) codes, and length of stay.²⁹ The use of ICD-10 diagnosis codes as a method for identification

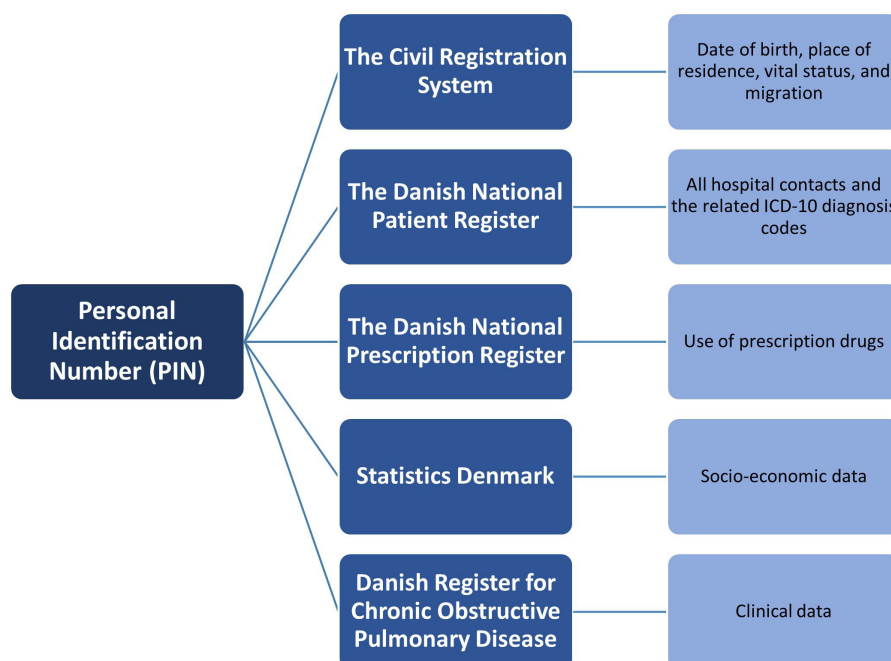


Figure 1 Overview of registers used in this study. ICD-10, International Classification of Diseases 10th Revision.

of pneumonia in hospitalised patients and construction of the Charlson Comorbidity Index (CCI) have both been validated.^{30 31}

The Danish National Prescription Register

NPR contains data since 1994 on all medication sold in Denmark in pharmacies or prescribed by a doctor or hospital. It does not contain medication prescribed during hospitalisation. It includes variables such as drug name, dose, price, and dates of redeemed prescriptions using anatomical therapeutic chemical (ATC) codes. The validity of the data is well established.³²

Statistics Denmark

DST contains various registers with longitudinal demographic information since 1970s and onwards.²⁷ These include immigration and emigration, vital status, partner status, and place of residence. DST contains income data since 1980, educational status since 1981, and occupational status since 1976.

The Danish register for COPD

DrKOL is a nationwide clinical database that contains clinical information on every patient with COPD that has visited a pulmonary outpatient clinic in Denmark since 2008.³³

Study population

All patients in Denmark with an ICD-10 diagnosis code of COPD (J44) and a first registration in DrKOL between 2008 and 2015 will be included (table 1). We consider clinical data from DrKOL vital to this study, and this restricts our study period to 2008 and onwards. Data in this study are available until the end of 2018. To ensure a 3-year follow-up period after first registered measurement, only patients with COPD with their first

measurement in the period 1 January 2008 until 31 December 2015 were included. Patients with missing or incomplete information of clinical measures are excluded.

Definition of baseline variables

All variables are derived from the registers. Baseline variables are allocated into two groups: health status variables and sociodemographic variables. An overview of baseline variables included in this study, including the register from which the data is derived, is shown in table 2.

Age, forced expiratory volume in first second (FEV₁), and body mass index (BMI) are included in the dataset as continuous variables. FEV₁ % of predicted will be divided into four groups (mild, moderate, severe, and very severe) according to GOLD guidelines.³⁴ BMI is categorised according to WHO groups. Age will be converted into a categorical variable with five groups by age in years (40–49, 50–59, 60–69, 70–79 and 80+). Medical Research Council Dyspnoea (MRC) scale is included in the dataset as an ordinal variable ranging from 1 to 5³⁵ and will be subsequently converted to the modified MRC (mMRC) scale and dichotomised into high/low dyspnoea according to GOLD guidelines.³⁴ Comorbidity is evaluated using ICD-10 codes from DNPR and will be collected as CCI within 3 years prior to inclusion date and as proportion of patients having comorbidity within each WHO category. CCI is divided into four groups (CCI=0, CCI=1, CCI=2 and CCI>2). A former asthma diagnosis is defined as an ICD-10 diagnosis code of asthma (J45) within 3 years prior to index. Medication at baseline is defined by redeemed prescription for inhaled medication in 4 months prior to index.³⁶ Medication is defined categorically as long-acting muscarinic antagonist (LAMA), long-acting beta-agonist (LABA), LABA+ICS, LABA+LAMA and LABA+LAMA+ICS. We include the number of moderate and severe pneumonias and non-pneumonic exacerbations 1 year prior to baseline. Pneumonias are defined in the same way as the outcome variable ‘pneumonia’ (see below). A moderate non-pneumonic exacerbation is defined as a redeemed prescription of oral corticosteroids, prednisolone (ATC H02AB06) and prednisone (ATC H02AB07), for short-term use. A severe non-pneumonic exacerbation is defined as a hospitalisation with ICD-10 primary diagnosis of J96 with a secondary diagnosis of J44. Incidents with ICD-10 diagnosis codes to define ‘pneumonia’ are excluded. Exacerbations and pneumonias must be separated by 4 weeks (28 days) to be considered separate events. If closer, it will be considered a pneumonia. Prescription of incident-relevant medication and hospitalisation must be separated by 4 weeks to be considered separate events. If closer, the severe event will be counted. Both pneumonias and non-pneumonic exacerbations are categorised by number of incidents (none, 1 moderate, ≥2 moderate, ≥1 severe). Cohabitation status will be defined as either living alone or cohabiting.

Table 1 Patient inclusion and exclusion

	N	Excluded (N)
Patients with COPD with a J44 diagnosis at the age of 40+ years old registered in the Danish COPD registry with a least one measurement in the period 2008–2018	n	
Exclude: Patients with index after 2015 (not 3 years follow-up)		N
Patients with COPD in the Danish COPD registry with first measurement 2008–2015	n	
Exclude: Patients with pneumonia in all follow-up months, errors in registered data or unrealistic data values		N
Final study population: patients with COPD in the Danish COPD registry with first measurement 2008–2015	n	
COPD, chronic obstructive pulmonary disease.		

**Table 2** Overview of baseline variables included in the study

Variable	Variable type	Values	Register used
Sociodemographic variables			
Age	Continuous		CRS
Sex	Categorical, binary	1=Female 2=Male	CRS
Cohabitation status	Categorical, binary	1=Living alone 2=Cohabiting	DST
Health status variables			
Comorbidity, CCI	Count		DNPR
BMI	Continuous		DrKOL
Smoking status	Categorical	0=Non-smoker 1=Former smoker 2=Current smoker	DrKOL
MRC dyspnoea scale	Categorical, ordinal	Range: 1–5	DrKOL
FEV ₁ % of expected	Continuous		DrKOL
Former asthma diagnosis	Categorical, binary	0=No 1=Yes	DNPR
Medication at baseline	Categorical	0=No medication 1=LAMA 2=LABA 3=LABA+ICS 4=LABA+LAMA 5=LABA+LAMA+ICS	
No of moderate pneumonias	Count		NPR
No of severe pneumonias	Count		DNPR
No of moderate non-pneumonic exacerbations	Count		NPR
No of severe non-pneumonic exacerbations	Count		DNPR

BMI, body mass index; CCI, Charlson Comorbidity Index; CRS, Civil Registration System; DNPR, Danish National Patient Register; DrKOL, Danish Register for Chronic Obstructive Pulmonary Disease; DST, Statistics Denmark; FEV₁, forced expiratory volume in 1 s; ICS, inhaled corticosteroid; LABA, long-acting beta-agonist; LAMA, long-acting muscarinic antagonist; MRC, Medical Research Council; MRC, Medical Research Council.

Definition of states

Patients are followed for 3 years starting from baseline. The follow-up period is divided into 36 monthly time units. For each month, patients are assigned to one of four states:

1. Dead.
2. Pneumonia.
3. ICS.
4. Nothing.

A patient can only have one of the four states in 1 month, and the states are ranked 1–4. Dead is an end state, so if a patient dies, this patient will have the state='dead' each month for the rest of the follow-up period. If a patient in the same month has both pneumonia and collects ICS, this patient will get the state 'pneumonia' in that month. The state 'nothing' is when a COPD patient in a month is neither dead, has pneumonia, nor uses ICS.

ICS treatment

Treatment with ICS is defined as redeeming a prescription for ICS medication (table 3). Months with ICS is

defined using DDD (defined daily dose) and expedition date.

Compliance is defined as:

Compliance=DDD for the expedition/duration between 2 expedition dates.

If compliance is $\geq 60\%$, there is no break between expeditions. If the compliance is $< 60\%$, the duration is the first expedition date+DDD, and there is a break

Table 3 Codes used for defining ICS treatment

Medication	ATC codes
ICS	R03BA (01–09)
ICS+LABA	R03AK (06–14)
ICS+LABA+LAMA	R03AL08, R03AL09, R03AL11, R03AL12

ATC, anatomical therapeutic chemical; ICS, inhaled corticosteroids; LABA, long-acting beta-agonist; LAMA, long-acting muscarinic antagonist.

Table 4 Codes used for defining pneumonia

Moderate pneumonia (ATC codes)	Severe pneumonia (ICD-10 codes)
▶ J01CA04: amoxicillin	▶ J12: viral pneumonias
▶ J01CA02: ampicillin	▶ J13–18: bacterial pneumonias
▶ J01CE02: phenoxymethylpenicillin	▶ A481: Legionnaires' disease
▶ J01CR01: Aampicillin and beta-lactamase inhibitor	▶ A709: ornithosis
▶ J01CR02: amoxicillin and beta-lactamaseinhibitor	▶ B012: varicella pneumonia
▶ J01FA01: erythromycin	
▶ J01FA06: roxithromycin	
▶ J01FA09: clarithromycin	
▶ J01FA10: azithromycin	

ATC, anatomical therapeutic chemical; ICD-10, International Classification of Diseases 10th Revision.

until the next expedition date. Patients can collect more DDD's than they should use in one period and can then use the DDD's over an extended period. We assume that the collected DDD's correspond to the usage of ICS over time, and if a patient collects more DDD's than covers the period between 2 expedition dates, the compliance is accumulated. From the last collection of ICS, the end date for ICS usage is the last collection date+DDD.

Pneumonia

Pneumonia includes both severe and moderate pneumonia. A patient gets the monthly state 'pneumonia' if the patient in the given month either collects medication for pneumonia (ATC-codes) or if the patient is admitted or have an ER contact with A or B pneumonia ICD-10 diagnosis (table 4).

Statistical analysis plan

Sequence analysis will be used to explore and identify relevant clusters.^{25 26} We use single channel sequence analysis with one sequence per person. Each patient has their own distinct sequence in the 36-month follow-up period based on the four possible monthly states (dead, pneumonia, ICS, and nothing). Patients who have the state 'nothing' and patients who have the state 'ICS' the whole follow-up period will be manually gathered in two clusters 'Only nothing' and 'Only ICS'. Furthermore, we assume that patients who die in the follow-up period are to be considered terminal, and all these patients are gathered in the same cluster named 'Terminal'. This ensures that they do not 'disturb' the clusters with patients who survive all 36 months.

The sequence analysis is a two-step analysis, where we include all patients who are not manually distributed to a cluster. First step is to group sequences together based on an assessment of how similar the sequence trajectories are. Here, we use optimal matching (OM) to find the dissimilarity. OM is the method most often used to assess

the dissimilarity between all pairs of sequence.²⁵ The OM stage allows us to produce a dissimilarity matrix, which can be used to distinguish typical trajectories. At the OM stage, choices must be made on costs for three possible operations (substitution, insertion, and deletion) that allow two sequences to match. These costs are set by a substitution matrix (SM) for substitution operations and an indel value for insertion and deletion operations. In this OM analysis, the observed transitions matrix is used to construct the substitution costs. This is a data-driven approach to define SM based on transition rates.³⁷ We use this approach because we do not assume that 'all states are equally different' (cost=1). Insertion-deletion (indel) cost is set to constant 1, and the distances are not normalised, since all sequences have the same length. Different indel values will be tested in sensitivity analyses. In step 2, we use a partitioning around medoids (PAM) algorithm to find the clusters in the dissimilarity matrix. PAM is used to group patients with similar sequence trajectories and allocate patients to the nearest cluster.³⁸ The optimal number of clusters will be found based on an internal validation test.³⁹ R will be used.^{40 41}

Descriptive baseline variables will be presented for the total population and by clusters. Continuous variables will be presented as both categorical variables as described above and continuous variables with mean and SD.

Differences in baseline characteristics between the clusters will be analysed with a multivariable logistic regression model with clusters as the outcome variable and the 'Only nothing' cluster as reference. We include all baseline variables except medication at baseline and mMRC scale. mMRC is highly correlated with FEV₁ % of predicted, and we consider the latter more important. We do not consider medication at baseline relevant for the model. Categorical variables are included as described above. Age is included as a continuous variable, but BMI and FEV₁ % of predicted are included as categorical variables because of an expected non-linear relationship. The model will be tested for multicollinearity and influential outliers. Because of the large period of patient inclusion (2008–2015), we will also examine the model for time-dependency by including patient index year as a covariate. A p<0.05 will be considered significant.

Furthermore, we will explore the difference between clusters regarding average number of the different states 'nothing', 'ICS' and 'pneumonia' in each cluster in the 36 months follow-up period.

No subgroup analyses have been planned. We will have no missing data, because we include only patients with complete registration in DrKOL at baseline, and all other variables are derived from complete registers.

Patient and public involvement

Patients were not involved in the development of the research question or the design of this study.



ETHICS AND DISSEMINATION

This study will be conducted in accordance with the Declaration of Helsinki. The study is entirely register based, and research ethics approval is not required according to Danish Law and National Ethics Committee Guidelines.

Storage and management of data

Data are obtained from DST. Data are also stored at DST, and all analyses will be carried out using a secure remote access to Forskermaskinerne (virtual desktop) at DST. Data management, data control, and data analysis will be carried out by experienced statisticians from i2minds.

Dissemination of results

The results of this study will be published in international peer-review journals with an emphasis on open access. The results will also be shared with relevant public stakeholders such as patient associations and medical and academic communities within pulmonary medicine.

DISCUSSION

This study has strengths and limitations. The largest strength of the study is its use of data from a large nationwide real-world population of patients with COPD, in a cohort that allows for complete follow-up. On the other hand, we are only able to use data that have already been entered into the registers. This is the case in our definition of moderate pneumonia and exacerbation. We define pneumonia by redeemed prescriptions, because we do not have access to clinical diagnoses in the primary care setting. We do, however, believe this to be a feasible strategy. The distinction is not clear between pneumonia and acute exacerbation of COPD (AECOPD) caused by a lower respiratory tract infection (LRTI),^{42–44} and they may be seen as a continuum of severity instead of distinctly different entities.⁴⁵ This is complicated by the fact that pneumonia is often clinically defined as the presence of typical symptoms and a new pulmonary infiltration on chest radiograph, but primary care practitioners do often not obtain these.⁴⁶ Furthermore, chest radiography may have limited sensitivity and positive predictive value for identifying pneumonia,⁴⁷ and patients can have pulmonary infiltrates on chest radiographs in AECOPD without clinical symptoms of pneumonia.⁴⁸ Even randomised controlled trials of inhaled medication for COPD have widely different definitions of pneumonia, and many have not used radiological confirmation of the diagnosis.⁴⁹ In summary, distinguishing between AECOPD and pneumonia is difficult both clinically and in research. Our proposed method identifies both pneumonia and AECOPD suspected to be caused by bacterial LRTI, while it likely excludes AECOPD of non-bacterial aetiology. Importantly, it does so in a real-world clinical outpatient setting, where the distinction between AECOPD and pneumonia is often not clear.

Contributors AK: conceptualisation, methodology, writing—original draft, writing—review and editing, funding acquisition. RI: conceptualisation,

methodology, writing—original draft, writing—review and editing. OH: conceptualisation, methodology, funding acquisition. AL: conceptualisation, methodology, writing—review and editing, funding acquisition.

Funding This work is supported by the Region of Southern Denmark, the University of Southern Denmark, and an unrestricted grant from Boehringer Ingelheim, grant number AGR-2018-731-5845.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Allan Klitgaard <http://orcid.org/0000-0002-6805-5695>

REFERENCES

- Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Global Initiative for Chronic Obstructive Lung Disease; 2020.
- Terry PD, Dhand R. Inhalation therapy for stable COPD: 20 years of gold reports. *Adv Ther* 2020;37:1812–28.
- Lipson DA, Crim C, Criner GJ, *et al*. Reduction in all-cause mortality with fluticasone Furoate/Umeclidinium/Vilanterol in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2020;201:1508–16.
- Gayle A, Dickinson S, Poole C, *et al*. Incidence of type II diabetes in chronic obstructive pulmonary disease: a nested case–control study. *npj Prim. Care Respir. Med.* 2019;29:28.
- Janson C, Lisspers K, Ställberg B, *et al*. Osteoporosis and fracture risk associated with inhaled corticosteroid use among Swedish COPD patients: the Arctic study. *Eur Respir J* 2021;57:2000515.
- Ställberg B, Janson C, Lisspers K, *et al*. Inhaled corticosteroids and the risk of type 2 diabetes among Swedish COPD patients. *NPJ Prim Care Respir Med* 2020;30:47.
- Miravittles M, Auladell-Rispau A, Monteagudo M, *et al*. Systematic review on long-term adverse effects of Inhaled corticosteroids in the treatment of COPD. *Eur Respir Rev* 2021;30:210075.
- Saeed MI, Eklöf J, Achir I, *et al*. Use of Inhaled corticosteroids and the risk of developing type 2 diabetes in patients with chronic obstructive pulmonary disease. *Diabetes Obes Metab* 2020;22:1348–56.
- Wang JJ, Rochtchina E, Tan AG, *et al*. Use of inhaled and oral corticosteroids and the long-term risk of cataract. *Ophthalmology* 2009;116:652–7.
- Chen H, Sun J, Huang Q, *et al*. Inhaled corticosteroids and the pneumonia risk in patients with chronic obstructive pulmonary disease: a meta-analysis of randomized controlled trials. *Front Pharmacol* 2021;12:691621.
- Miravittles M, González-Torralba F, Represas-Represas C, *et al*. Pulmonologists' opinion on the use of inhaled corticosteroids in chronic obstructive pulmonary disease patients in Spain: a cross-sectional survey. *Int J Chron Obstruct Pulmon Dis* 2022;17:1577–87.
- Dalín DA, Løkke A, Kristiansen P, *et al*. A systematic review of blood Eosinophils and continued treatment with Inhaled corticosteroids in patients with COPD. *Respir Med* 2022;198:S0954-6111(22)00145-7:106880..
- Lipson DA, Barnhart F, Brealey N, *et al*. Once-Daily single-Inhaler triple versus dual therapy in patients with COPD. *N Engl J Med* 2018;378:1671–80.
- Wedzicha JA, Banerji D, Chapman KR, *et al*. Indacaterol–Glycopyrronium versus Salmeterol–Fluticasone for COPD. *N Engl J Med* 2016;374:2222–34.
- Papi A, Vestbo J, Fabbri L, *et al*. Extrafine inhaled triple therapy versus dual bronchodilator therapy in chronic obstructive pulmonary disease (tribute): a double-blind, parallel group, randomised controlled trial. *The Lancet* 2018;391:1076–84.

- 16 Hartley BF, Barnes NC, Lettis S, *et al.* Risk factors for exacerbations and pneumonia in patients with chronic obstructive pulmonary disease: a pooled analysis. *Respir Res* 2020;21:5.
- 17 Marques A, Souto-Miranda S, Machado A, *et al.* COPD profiles and treatable traits using minimal resources: identification, decision tree and stability over time. *Respir Res* 2022;23:30.
- 18 Lange P, Halpin DM, O'Donnell DE, *et al.* Diagnosis, assessment, and Phenotyping of COPD: beyond FEV₁. *INT J Chron obstruct Pulmon DIS 11 SPEC Iss(Spec ISS)*. 2016:3–12.
- 19 Yousuf A, McAuley H, Elneima O, *et al.* The different phenotypes of COPD. *Br Med Bull* 2021;137:82–97.
- 20 Chubachi S, Sato M, Kameyama N, *et al.* Identification of five clusters of Comorbidities in a longitudinal Japanese chronic obstructive pulmonary disease cohort. *Respir Med* 2016;117:S0954-6111(16)30149-4:272–9..
- 21 Vanfleteren LEGW, Spruit MA, Groenen M, *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;187:728–35.
- 22 Burgel P-R, Paillasser J-L, Caillaud D, *et al.* Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. *Eur Respir J* 2010;36:531–9.
- 23 Augustin IML, Spruit MA, Houben-Wilke S, *et al.* The respiratory Physiome: clustering based on a comprehensive lung function assessment in patients with COPD. *PLoS One* 2018;13:e0201593.
- 24 Cezard G, Sullivan F, Keenan K. Understanding Multimorbidity Trajectories in Scotland using sequence analysis. *Sci Rep* 2022;12:16485.
- 25 Abbott A, Tsay A. Sequence analysis and optimal matching methods in sociology. *Sociol Methods Res* 2000;29:3–33.
- 26 Liao TF, Bolano D, Brzinsky-Fay C, *et al.* Sequence analysis: its past, present, and future. *Soc Sci Res* 2022;107:S0049-089X(22)00078-3:102772..
- 27 Mainz J, Hess MH, Johnsen SP. The Danish unique personal Identifier and the Danish civil registration system as a tool for research and quality improvement. *Int J Qual Health Care* 2019;31:717–20.
- 28 Schmidt M, Pedersen L, Sørensen HT. The Danish civil registration system as a tool in epidemiology. *Eur J Epidemiol* 2014;29:541–9.
- 29 Schmidt M, Schmidt SAJ, Sandegaard JL, *et al.* The Danish national patient registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015;7:449.
- 30 Thygesen SK, Christiansen CF, Christensen S, *et al.* The predictive value of ICD-10 diagnostic coding used to assess Charlson Comorbidity index conditions in the population-based Danish national Registry of patients. *BMC Med Res Methodol* 2011;11:83.
- 31 Skull SA, Andrews RM, Byrnes GB, *et al.* ICD-10 codes are a valid tool for identification of pneumonia in hospitalized patients aged > or = 65 years. *Epidemiol Infect* 2008;136:232–40.
- 32 Rasmussen L, Valentin J, Gesser KM, *et al.* Validity of the prescriber information in the Danish national prescription registry. *Basic Clin Pharmacol Toxicol* 2016;119:376–80. Available <http://doi.wiley.com/10.1111/bcpt.2016.119.issue-4>
- 33 Lange P, Tøttenborg SS, Sorknæs AD, *et al.* Danish register of chronic obstructive pulmonary disease. *Clin Epidemiol* 2016;8:673–8.
- 34 Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease (2023 report). 2023.
- 35 Bestall JC, Paul EA, Garrod R, *et al.* Usefulness of the medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax* 1999;54:581–6.
- 36 Løkke A, Hilberg O, Lange P, *et al.* Disease Trajectories and impact of one moderate exacerbation in gold B COPD patients. *Int J Chron Obstruct Pulmon Dis* 2022;17:569–78.
- 37 Studer M, Ritschard G. What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society* 2016;179:481–511.
- 38 Leonard PJR. *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley. 1990.
- 39 Studer M. Weightedcluster library manual: a practical guide to creating Typologies of Trajectories in the social sciences with R. *LIVES Working Papers* 2013;24.
- 40 Gabadinho A, Ritschard G, Müller NS, *et al.* Analyzing and Visualizing state sequences in R with **Traminer**. *J Stat Softw* 2011;40:1–37.
- 41 Seqdist R documentation: distances (Dissimilarities) between sequences. n.d. Available: <http://traminer.unige.ch/doc/seqdist.html>
- 42 Williams N, Staples KJ, Ostridge K. Distinguishing features of pneumonia and exacerbations in COPD. Annual Congress 2015 2015.
- 43 Huerta A, Crisafulli E, Menéndez R, *et al.* Pneumonic and Nonpneumonic exacerbations of COPD: inflammatory response and clinical characteristics. *Chest* 2013;144:S0012-3692(13)60656-X:1134–42..
- 44 Lieberman D, Lieberman D, Gelfer Y, *et al.* Pneumonic vs Nonpneumonic acute exacerbations of COPD. *Chest* 2002;122:1264–70.
- 45 Williams NP, Ostridge K, Devaster J-M, *et al.* On behalf of the, impact of Radiologically stratified exacerbations: insights into pneumonia Aetiology in COPD. *Respir Res* 2018;19:143.
- 46 Metlay JP, Waterer GW, Long AC, *et al.* Diagnosis and treatment of adults with community-acquired pneumonia an official clinical practice guideline of the American Thoracic society and infectious diseases society of America. *Am J Respir Crit Care Med* 2019;200:e45–67.
- 47 Self WH, Courtney DM, McNaughton CD, *et al.* High discordance of chest X-ray and computed tomography for detection of pulmonary opacities in ED patients: implications for diagnosing pneumonia. *Am J Emerg Med* 2013;31:401–5.
- 48 Saleh A, López-Campos JL, Hartl S, *et al.* The effect of incidental consolidation on management and outcomes in COPD exacerbations: data from the European COPD audit. *PLoS One* 2015;10:e0134004.
- 49 Wise RA, Bafadhel M, Crim C, *et al.* Discordant diagnostic criteria for pneumonia in COPD trials: a review. *Eur Respir Rev* 2021;30:210124.