WEB EXTRA MATERIAL

Supplementary methodology and results

Assessing the presence of the Hawthorne effect in an effectiveness randomised clinical trial comparing the usual care arm of the Salford Lung Study COPD with a non-trial UK population of COPD patients

Alexander Pate, MSci[1], Michael Barrowman, MSci[1], David Webb, MSc[2], Jeanne M Pimenta, PhD[2], Kourtney J Davis, PhD[3], Rachael Williams, MSc[4], Tjeerd Van Staa, MD, PhD[1,5], Matthew Sperrin, PhD[1]

[1]Farr Institute, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK; [2]Real World Evidence and Epidemiology, GlaxoSmithKline, Stockley Park, Uxbridge, Middlesex, UK; [3]Real World Evidence and Epidemiology, GlaxoSmithKline, Collegeville, PA, USA; [4]Clinical Practice Research Datalink, London, UK; [5]Utrecht University, Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht, The Netherlands

Contents

Section 1. Matching algorithm 2
Inclusion criteria 2
Exclusion criteria 2
Section 2. Outcome definitions and definition of events described in the matching algorithm 4
A. Acute exacerbation of COPD (AECOPD) 4
Defining exacerbation events 4
Defining exacerbation episodes 4
B. Strict definition of AECOPD 4
Defining exacerbation events 4
Defining exacerbation episodes 4
C. Chronic user of OCS 4
D. Pneumonia 4
E. Mortality 5
F. Numbers of days contact with primary care 5
G. Count of trial-related prescription items 5
H. Treatment switching 5
Section 3. Data source used to derive variables for the SLS COPD cohort 6
Section 4. Comparison of complete case and imputed datasets 7
Section 5. Additional clinical and demographic characteristics available not presented in Table 1 in main article 10
Section 8. Comparison of coding rates for patients in EMIS and VISION practices in the SLS COPD UC arm 13
Reference 16

Section 1. Matching algorithm

The Salford Lung Study in chronic obstructive pulmonary disease (SLS COPD) cohort comprised the entire usual care (SLS COPD UC) arm (N=1403). The process for generating the index date-matched cohorts, CPRD-GM (Clinical Practice Research Datalink in Greater Manchester; N=2 049) and CPRD-xGM (Clinical Practice Research Datalink outside of Greater Manchester; N=16 758), was as follows:

The first cohort was the UC arm of SLS COPD, which was used as is.

The second cohort was the Clinical Practice Research Datalink (CPRD) cohort. This comprised the linked primary care Office for National Statistics (ONS) mortality data, Hospital Episode Statistics (HES) and socioeconomic status (SES) data, and was generated in a multi-stage process. A large cohort of COPD patients was extracted from CPRD according to the following criteria:

- A diagnosis of COPD recorded within the clinical or referral files of CPRD
- At least 1 day of up-to-standard registration in CPRD
- Age ≥40 years
- Eligible for linkage to HES, SES, and ONS data.

A further step was taken to reduce the cohort to patients with codes that strongly indicate COPD (ie, those considered as either 1: COPD or 2: COPD-exacerbation-high-specificity). This cohort was classed as the CPRD cohort. The code list used was shown to identify COPD patients in the CPRD with an approximate positive predictive value (PPV) of between 80–90%.[1]

In addition, the following flags were available within the CPRD Practice File**:**

- Greater Manchester flag: 0 = Practice not in Greater Manchester, 1 = Practice in Greater Manchester
- Local Authority (LA) flag: pseudonymised number representing practice LA.

The CPRD cohort was then split into two groups according to the Greater Manchester flag: CPRD-GM (where flag = 1), and CPRD-xGM (rest of England, where flag = 0).

Restricted cohorts were then constructed in the CPRD population, based on the following inclusion/exclusion criteria (which matched the inclusion/exclusion criteria for SLS COPD) and the study period for SLS COPD.

Inclusion criteria

1. Patients with a documented general practitioner (GP) diagnosis of COPD (the code list for COPD diagnosis will be available for this study at https://www.gsk-clinicalstudyregister.com), and "currently receiving" maintenance (UC) therapy. UC was defined as all patients currently receiving either:
    a. inhaled corticosteroid (ICS) alone or in combination with a long-acting β_2-agonist (LABA) (this could be a fixed-dose combination or an ICS/LABA provided in two separate inhalers), or ICS and a long-acting muscarinic antagonist (LAMA) provided in two separate inhalers~~)~~, or
    b. long-acting bronchodilator therapy alone (eg, tiotropium or salmeterol), or the use of two bronchodilators (ie, LABA/LAMA), or
    c. triple therapy (ie, ICS/LABA plus a LAMA).

"Currently receiving" was defined as a prescription within 12 months before the index date. Code lists for ICS, LABA, and LAMA will be available for this study at https://www.gsk-clinicalstudyregister.com.

2. Male or female patients aged ≥40 years at index date.
3. Patients with a history of treatment with systemic/oral corticosteroids (OCS), antibiotics (in association with GP contact), and/or hospitalisation for at least one COPD exacerbation in the 3 years before index date. The algorithm for COPD exacerbation is described in Section 2 and the code list for COPD exacerbations will be available for this study at https://www.gsk-clinicalstudyregister.com.

Exclusion criteria

Patients meeting any of the following criteria were not included in the restricted cohorts:

1. Patients with unstable COPD: Patients with an exacerbation occurring within 2 weeks of index date. The index date had to be delayed until at least 2 weeks after the onset of an exacerbation and until the exacerbation had resolved.
2. Chronic user of OCS: Patients considered chronic users of OCS for respiratory or other indications. An algorithm for determining chronic use of OCS is given in Section 2.
3. Patients with fewer than 12 months of data available preceding the index date (ie, registered at a contributing GP practice less than 12 months ago).

It was proposed in the protocol that an additional exclusion criterion from SLS COPD "Patients with any life-threatening condition or uncontrolled/clinically significant disease—ie, diagnosis codes in any preceding period" should also be used. However, this definition could not be matched exactly to SLS COPD using CPRD, as this decision was down to the clinician's discretion. For this reason, this criterion was not applied.

To match index dates and follow-up times with the SLS COPD UC cohort, the following algorithm was applied to the CPRD population. This was done simultaneously for both the CPRD-GM and the CPRD-xGM cohorts, as we did not need to distinguish between them at this point.

1. A list of potentially eligible individuals in CPRD (patients who would be eligible at some point during the SLS COPD recruitment phase) was created. In other words, patients with the following characteristics were discarded:
    a. Died before first patient recruited to SLS COPD
    b. Did not attain the age of 40 years, or was not diagnosed with COPD, until after the final patient was recruited to SLS COPD.
2. Each remaining individual was matched with a patient from SLS COPD based on potential study entry date only—ie, with no consideration of covariates other than checking eligibility (inclusion and exclusion criteria). For everyone that remained, the following algorithm was applied:
    a. Randomly sampled (with replacement) a patient from the SLS COPD UC arm and assigned index date to CPRD patient
    b. Delayed index date if necessary per exclusion criteria 1 (patients with unstable COPD)
    c. If the CPRD patient was eligible at the index date (ie, after applying the inclusion and exclusion criteria), then they were included, with the SLS COPD patient's entry date taken as the CPRD patient's index date. The follow-up time for the CPRD patient was then up to a maximum of 12 months (which was the follow-up period for individuals in SLS COPD), but may have been truncated by death, loss to follow-up, etc.
    d. If the CPRD patient was not eligible at the entry date of the SLS COPD patient, they were discarded and the next CPRD patient was considered.

Once this process was complete, patients in the CPRD-xGM cohort who were not discarded comprised the comparison cohort for the SLS COPD UC. The corresponding patients within Greater Manchester (CPRD-GM cohort) comprised a secondary comparison cohort. They had an index date distribution that matched the SLS COPD UC cohort, and met the inclusion and exclusion criteria of the SLS COPD UC cohort at their matched entry date.

Section 2. Outcome definitions and definition of events described in the matching algorithm

A. Acute exacerbation of COPD (AECOPD)

Defining exacerbation events

In the CPRD cohorts, the definition of a COPD exacerbation was identified using a validated algorithm based on medical and treatment codes in primary care data that have been shown to result in a PPV of 86% and 63%.[1] This was referred to as the exacerbation algorithm. Exacerbation events were defined through any of the following:

1. Prescription of prespecified antibiotics (ABx) and OCS both on the same day, OR
2. Exacerbation symptom definition (exacerbation symptoms are codes suggesting increase in two or more of: breathlessness, cough, or sputum volume and/or purulence) and antibiotics, where medical code is on the same day as prescription, OR
3. Exacerbation symptom definition and OCS, where medical code is on the same day as prescription, OR
4. Lower respiratory tract infection code (not including pneumonia codes, but including acute bronchitis and other lower respiratory tract infection diagnosis codes), OR
5. Definite acute COPD exacerbation medical diagnosis code.

The algorithm we used differs slightly from the preferred algorithm identified in the paper. We were looking at primary care records only; therefore, hospitalisations could not be included in the algorithm. In addition, data on prescription length were not rich enough to identify consistently the length of prescriptions in the electronic health record (EHR). Therefore, the original algorithm of prescription of pre-specified ABx and OCS on the same day was relaxed from prescriptions of length 5–14 days, to include any prescription of ABx or OCS.

Defining exacerbation episodes

After determining all the events for COPD exacerbations (ie, any of the definitions bulleted above), they were combined to produce exacerbation episodes. The definition of an exacerbation episode was as follows:

1. Following an exacerbation event, project forwards 14 days and use as a potential end-of-episode date. Exacerbations within this period were assumed to be related to the initial exacerbation.
2. If there are no exacerbations in the 14 days after the potential end-of-episode date, halt the algorithm; the episode has been fully identified.
3. If there are exacerbations in the 14 days after the potential end-of-episode date, move the potential end-of-episode date 14 days forward of the start of the last exacerbation in these 14 days.
4. Repeat from step 2 until there are 14 days after the end-of-episode date clear of exacerbations.

All electronic health record analyses (CPRD or SLS EHR) involving COPD exacerbations were based on COPD exacerbation episodes, not COPD exacerbations**.** For analyses using the SLS COPD study database, exacerbations were used as recorded in the dataset.

B. Strict definition of AECOPD

Defining exacerbation events

In this definition, exacerbation events were defined solely through criterion 5 from the previous algorithm (A. defining moderate/severe COPD exacerbations). Criterion 5 is 'definite acute COPD exacerbation medical diagnosis code'. This means only specific AECOPD diagnosis codes were used to identify exacerbation events.

Defining exacerbation episodes

Exacerbation episodes were defined in the same way as full acute COPD exacerbations episodes.

C. Chronic user of OCS

To identify patients who were considered chronic users of OCS for respiratory or other indications, the respiratory team at GSK advised to use a definition of a set of at least four prescription records over a 12-month period, with the gap between any two consecutive prescriptions being 30 days at most.

D. Pneumonia

Pneumonia (defined by hospitalisation) was defined using ICD-10 hospital discharge codes (codes will be available for this study at https://www.gsk-clinicalstudyregister.com). The length of each pneumonia episode was defined through the hospitalisation start and end dates as recorded in the HES dataset. There were two

definitions of pneumonia: the first counted hospitalisations with a pneumonia code at any point during the hospitalisation; the second only counted hospitalisations with pneumonia as the primary admission diagnosis of the hospitalisation episode.

E. Mortality

All-cause death during the 12-month follow-up. For the CPRD cohorts, deaths were determined using ONS-linked mortality data.

F. Numbers of days contact with primary care

We counted the number of days the patient had contact with the primary care services. This included any contact with the GP—ie, including phone calls etc. This was defined by any recorded medical code. Any code recorded on the same day as a code indicating participation in a trial was not counted.

G. Count of trial-related prescription items

This is the number of distinct trial-related prescription items prescribed by the GP over the course of the 1-year follow-up. A trial-related prescription item was any item that caused the patient to meet inclusion criteria 1, maintenance therapy.

H. Treatment switching

We defined four 3-month periods consecutively across the year of follow-up. In each period, a patient could be in one of eight groups: No medication, LAMA, LABA, ICS, LAMA/LABA, LAMA/ICS, LABA/ICS, LAMA/LABA/ICS. The patient belonged to the group containing whatever prescriptions they had received in that period. Between each period, switching was defined as

- Step up (addition of one or more extra medications)
- Step down (removal of one or more extra medications)
- Switching (addition and removal of one or more medications)
- None (same medication group).

Section 3. Data source used to derive variables for the SLS COPD cohort

Table E1: Variables listed by data source from which they were derived

Variables	Data source
Pneumonia (outcome), sex, age, SES (using IMD 2010 scores), FEV_1%, ratio of FEV_1 to FVC (FEV_1/FVC), GOLD stage, smoking status, and BMI	SLS COPD study database
COPD (outcome), historical comorbidities (not available in the SLS COPD study database), Charlson comorbidity index (with COPD removed; not available in the SLS COPD study database), MRC dyspnoea score (not available in the SLS COPD study database), current asthma and current COPD medication group (for increased comparability with the CPRD), and immunisation history for influenza and pneumonia.	SLS COPD EHR database

COPD=chronic obstructive pulmonary disease. CPRD=Clinical Practice Research Datalink. EHR=electronic health record. FEV_1=forced expiratory volume in 1 s. FVC=forced vital capacity. GOLD=Global initiative for chronic Obstructive Lung Disease. IMD=index of multiple deprivation. MRC=Medical Research Council. SES=socioeconomic status. SLS COPD=Salford Lung Study in chronic obstructive pulmonary disease.

Section 4. Comparison of complete case and imputed datasets

***Table E2*: Comparison of complete case and imputed datasets for variables with missing data: CPRD-GM**

Variable	Category	CPRD-GM complete case	CPRD-GM imputed
N		1493	2049
SES IMD 2010 quintiles	5 (least deprived)	107 (7·17%)	127 (6·20%)
	4	180 (12·06%)	219 (10·69%)
	3	214 (14·33%)	283 (13·81%)
	2	332 (22·24%)	456 (22·25%)
	1 (most deprived)	660 (44·21%)	964 (47·05%)
FEV_1%	Mean (95% CI) Median (5–95% range)	56·80 (55·82–57·77) 55·80 (27·28–89·98)	57·01 (56·17–57·85) 55·87 (27·25–89·98)
FEV_1/FVC (%)	Mean (95% CI) Median (5–95% range)	60·73 (59·89–61·56) 60·00 (36·00–88·00)	61·14 (60·44–61·84) 60·91 (35·38–89·00)
GOLD stage	0 (FEV_1/FVC ≥70)	416 (27·86%)	601 (29·33%)
	1 (FEV_1/FVC <70, FEV_1% ≥80)	65 (4·35%)	86 (4·20%)
	2 (FEV_1/FVC <70, 50 ≤ FEV_1% <80)	506 (33·89%)	676 (32·99%)
	3 (FEV_1/FVC <70, 30 ≤ FEV_1% <50)	410 (27·46%)	547 (26·70%)
	4 (FEV_1/FVC <70, FEV_1% <30)	96 (6·43%)	139 (6·78%)
MRC dyspnoea score	1 (least breathlessness)	147 (9·85%)	203 (9·91%)
	2	525 (35·16%)	707 (34·50%)
	3	446 (29·87%)	592 (28·89%)
	4	314 (21·03%)	453 (22·11%)
	5 (most breathlessness)	61 (4·09%)	94 (4·59%)
Smoking	Never	97 (6·50%)	165 (8·05%)
	Ex	880 (58·94%)	1177 (57·44%)
	Current	516 (34·56%)	707 (34·50%)
BMI, kg/m^2	18·50–24·99	496 (33·22%)	669 (32·65%)
	<18·50	70 (4·69%)	96 (4·69%)
	25·00–29·99	466 (31·21%)	629 (30·70%)
	≥30·00	461 (30·88%)	655 (31·97%)

BMI=body mass index. CI=confidence interval. CPRD-GM=Clinical Practice Research Datalink in Greater Manchester. FEV_1=forced expiratory volume in 1 s. FVC=forced vital capacity. GOLD=Global initiative for chronic Obstructive Lung Disease. IMD=index of multiple deprivation. MRC=Medical Research Council. SES=socioeconomic status.

***Table E3*: Comparison of complete case and imputed datasets for variables with missing data: CPRD-xGM**

Variable	Category	CPRD-xGM complete case	CPRD-xGM imputed
N		12 077	16 758
SES IMD 2010 quintiles	5 (least deprived)	1765 (14·61%)	2500 (14·92%)
	4	2490 (20·62%)	3431 (20·47%)
	3	2361 (19·55%)	3348 (19·98%)
	2	2858 (23·66%)	3899 (23·27%)
	1 (most deprived)	2603 (21·55%)	3580 (21·36%)
FEV_1%	mean (95% CI) median (5–95% range)	55·84 (55·50–56·17) 55·11 (26·80–87·62)	56·01 (55·72–56·30) 55·28 (26·11–88·46)
FEV_1/FVC (%)	mean (95% CI) median (5–95% range)	60·23 (59·93–60·53) 60·00 (34·00–87·00)	60·57 (60·32–60·81) 60·57 (34·00–88·32)
GOLD stage	0 (FEV_1/FVC ≥70)	3205 (26·54%)	4743 (28·30%)
	1 (FEV_1/FVC <70, FEV_1% ≥80)	479 (3·97%)	646 (3·85%)
	2 (FEV_1/FVC <70, 50 ≤ FEV_1% <80)	4096 (33·92%)	5512 (32·89%)
	3 (FEV_1/FVC <70, 30 ≤ FEV_1% <50)	3377 (27·96%)	4503 (26·87%)
	4 (FEV_1/FVC <70, FEV_1% <30)	920 (7·62%)	1354 (8·08%)
MRC dyspnoea score	1 (least breathlessness)	1193 (9·88%)	1778 (10·61%)
	2	4096 (33·92%)	5545 (33·09%)
	3	3708 (30·70%)	4957 (29·58%)
	4	2564 (21·23%)	3612 (21·55%)
	5 (most breathlessness)	516 (4·27%)	866 (5·17%)
Smoking	Never	773 (6·40%)	1349 (8·05%)
	Ex	7344 (60·81%)	10 033 (59·87%)
	Current	3960 (32·79%)	5376 (32·08%)
BMI, kg/m^2	18·50–24·99	3835 (31·75%)	5334 (31·83%)
	<18·50	564 (4·67%)	838 (5·00%)
	25·00–29·99	3950 (32·71%)	5423 (32·36%)
	≥30·00	3728 (30·87%)	5163 (30·81%)

BMI=body mass index. CI=confidence interval. CPRD-xGM=Clinical Practice Research Datalink outside of Greater Manchester. FEV_1=forced expiratory volume in 1 s. FVC=forced vital capacity. GOLD=Global initiative for chronic Obstructive Lung Disease. IMD=index of multiple deprivation. MRC=Medical Research Council. SES=socioeconomic status.

***Table E4*: Comparison of complete case and imputed datasets for variables with missing data: SLS COPD UC**

Variable	Category	SLS COPD UC complete case	SLS COPD UC imputed
N		863	1403
SES IMD 2010 quintiles	5 (least deprived)	54 (6·26%)	72 (5·13%)
	4	61 (7·07%)	106 (7·56%)
	3	128 (14·83%)	203 (14·47%)
	2	191 (22·13%)	295 (21·03%)
	1 (most deprived)	429 (49·71%)	727 (51·82%)
FEV_1%	mean (95% CI) median (5–95% range)	60·68 (59·40–61·96) 61·00 (29·70–93·00)	60·29 (59·30–61·29) 60·90 (28·40–91·94)
FEV_1/FVC (%)	mean (95% CI) median (5–95% range)	54·37 (53·45–55·30) 54·55 (31·73–76·15)	54·90 (54·18–55·62) 55·70 (31·73–76·53)
GOLD stage	0 (FEV_1/FVC ≥70)	107 (12·40%)	175 (12·47%)
	1 (FEV_1/FVC <70, FEV_1% ≥80)	66 (7·65%)	132 (9·41%)
	2 (FEV_1/FVC <70, 50 ≤ FEV_1% <80)	371 (42·99%)	692 (49·32%)
	3 (FEV_1/FVC <70, 30 ≤ FEV_1% <50)	242 (28·04%)	320 (22·81%)
	4 (FEV_1/FVC <70, FEV_1% <30)	77 (8·92%)	84 (5·99%)
MRC dyspnoea score	1 (least breathlessness)	112 (12·98%)	157 (11·19%)
	2	328 (38·01%)	515 (36·71%)
	3	288 (33·37%)	473 (33·71%)
	4	120 (13·90%)	234 (16·68%)
	5 (most breathlessness)	15 (1·74%)	24 (1·71%)
Smoking	Never	35 (4·06%)	59 (4·21%)
	Ex	417 (48·32%)	678 (48·33%)
	Current	411 (47·62%)	666 (47·47%)
BMI, kg/m^2	18·50–24·99	279 (32·33%)	442 (31·50%)
	<18·50	37 (4·29%)	76 (5·42%)
	25·00–29·99	273 (31·63%)	417 (29·72%)
	≥30·00	274 (31·75%)	468 (33·36%)

BMI=body mass index. CI=confidence interval. COPD=chronic obstructive pulmonary disease. FEV_1=forced expiratory volume in 1 s. FVC=forced vital capacity. GOLD=Global initiative for chronic Obstructive Lung Disease. IMD=index of multiple deprivation. MRC=Medical Research Council. SES=socioeconomic status. SLS COPD UC=Salford Lung Study in chronic obstructive pulmonary disease—usual care arm.

The complete case analyses for the primary outcome variable, count of moderate/severe COPD exacerbations, provide very similar results to the original analysis (Table 3) which produced percentiles of 98·37 and 96·71 for the Poisson (count of exacerbations) and Cox (time until first exacerbation) models, respectively.

Section 5. Additional clinical and demographic characteristics available not presented in Table 1 in main article

Table E5: Additional clinical and demographic characteristics available but not included manuscript Table 1

Variable	Category	CPRD-GM	CPRD-xGM	SLS COPD UC
N		2049	16 758	1403
Index date	5th percentile	03/06/12	31/05/12	31/05/12
	25th percentile	09/10/12	03/10/12	03/10/12
	50th percentile	09/07/13	17/07/13	23/07/13
	75th percentile	24/04/14	24/04/14	12/05/14
	95th percentile	23/09/14	22/09/14	23/09/14
Charlson comorbidity index	0	341 (16·64%)	2569 (15·33%)	271 (19·32%)
	1–2	1020 (49·78%)	8226 (49·09%)	703 (50·11%)
	3–4	462 (22·55%)	4094 (24·43%)	297 (21·17%)
	5+	226 (11·03%)	1869 (11·15%)	132 (9·41%)
MRC dyspnoea score	Missing	199 (9·71%)	1719 (10·26%)	12 (0·86%)
	1 (least breathlessness)	179 (8·74%)	1530 (9·13%)	154 (10·98%)
	2	634 (30·94%)	4912 (29·31%)	510 (36·35%)
	3	532 (25·96%)	4509 (26·91%)	470 (33·50%)
	4	417 (20·35%)	3299 (19·69%)	233 (16·61%)
	5 (most breathlessness)	88 (4·29%)	789 (4·71%)	24 (1·71%)
BMI, kg/m^2	Missing	68 (3·32%)	940 (5·61%)	281 (20·03%)
	18·50–24·99	651 (31·77%)	5005 (29·87%)	351 (25·02%)
	<18·50	94 (4·59%)	793 (4·73%)	52 (3·71%)
	25·00–29·99	608 (29·67%)	5121 (30·56%)	349 (24·88%)
	≥30·00	628 (30·65%)	4899 (29·23%)	370 (26·37%)
Vaccinations	Influenza	1856 (90·58%)	15 105 (90·14%)	1267 (90·31%)
	Pneumococcal	320 (15·62%)	2259 (13·48%)	241 (17·18%)

BMI=body mass index. CPRD-GM=Clinical Practice Research Datalink in Greater Manchester. CPRD-xGM=Clinical Practice Research Datalink outside of Greater Manchester. MRC=Medical Research Council. SLS COPD UC=Salford Lung Study in chronic obstructive pulmonary disease—usual care arm.

Section 6. Outcome comparisons using complete case datasets

***Table E6*: Primary analysis carried out using complete case cohorts. Random intercept for SLS COPD placed in distribution of random intercepts of LAs from CPRD in fully adjusted multilevel model**

Variable	CPRD-xGM 2·5th percentile	CPRD-xGM median value	CPRD-xGM 97·5th percentile	SLS COPD UC value	SLS COPD UC percentile	Unusual flag
Count of moderate/severe COPD exacerbations[*]	0·91	1·00	1·13	1·11	94·06	No
Time to moderate/severe COPD exacerbation[†]	0·92	1·00	1·13	1·10	94·01	No

[*]Poisson model; [†]Cox model.

COPD=chronic obstructive pulmonary disease. CPRD-xGM=Clinical Practice Research Datalink outside of Greater Manchester. LA=Local Authority. SLS COPD UC =Salford Lung Study in chronic obstructive pulmonary disease—usual care arm.

Section 7. Pneumonia analyses

Pneumonia analyses were conducted to fulfil a safety objective and not primary objectives of representativeness or Hawthorne effect. However, for transparency, results of the analysis are provided here. The analyses were performed using the SLS COPD database and CPRD linked with secondary care data (HES), as hospital data were required to determine pneumonia hospitalisations. The crude rate of hospitalised pneumonia appears lower in the SLS COPD UC cohort (76·01 per 1000 person-years [62·1 to 92·1]) in comparison with both CPRD-xGM (106·85 [101·59 to 112·30]) and CPRD-GM (101·99 [87·90 to 117·70]). This is similar to hospitalisations with pneumonia as primary admission diagnosis.

***Table E7*: Crude counts and rates of pneumonia hospitalisations**

Variable	Category	CPRD-GM	CPRD-xGM	SLS COPD UC
N		2046	16 745	1403
Count of pneumonia hospitalisations	Number of events Rate per 1000 person-years (95% CI)	187 101·99 (87·90–117·70)	1551 106·85 (101·59–112·30)	104 76·01 (62·10–92·10)
Count of hospitalisations with pneumonia as primary admission diagnosis	Number of events Rate per 1000 person-years (95% CI)	120 66·03 (54·74–78·95)	951 66·09 (61·96–70·43)	77 56·53 (44·61–70·65)

CI=confidence interval. CPRD-GM=Clinical Practice Research Datalink in Greater Manchester. CPRD-xGM=Clinical Practice Research Datalink outside of Greater Manchester. SLS COPD=Salford Lung Study in chronic obstructive pulmonary disease. SLS COPD UC=Salford Lung Study in chronic obstructive pulmonary disease—usual care arm.

In fully adjusted models, participation in SLS COPD UC was not associated with an unusual rate of hospitalised pneumonia episodes (percentile = 9·76), or an unusual hazard ratio when considering time to first hospitalisation for pneumonia (percentile = 9·11). The percentiles for the equivalent statistics when considering hospitalisations with pneumonia as primary admission diagnosis were 79·81 and 66·96, respectively. Although these are much higher, they are still well within the usual range.

***Table E8*: Random intercept for SLS COPD placed in distribution of random intercepts of LAs from CPRD in fully adjusted multilevel model**

Variable	CPRD-xGM 2·5th percentile	CPRD-xGM median value	CPRD-xGM 97·5th percentile	SLS COPD UC value	SLS COPD UC percentile	Unusual flag
Count of pneumonia episodes[*]	0·91	0·99	1·13	0·94	9·76	No
Time until first pneumonia episode[†]	0·90	1·00	1·16	0·92	9·11	No
Count of pneumonia episodes with pneumonia as primary admission diagnosis[*]	0·96	1·00	1·04	1·02	79·81	No
Time until first pneumonia episode with pneumonia as primary admission diagnosis[†]	0·92	1·00	1·07	1·01	66·96	No

[*]Poisson model; [†]Cox model.

CPRD= Clinical Practice Research Datalink. CPRD-xGM=Clinical Practice Research Datalink outside of Greater Manchester. LA=local authority. SLS COPD UC =Salford Lung Study in chronic obstructive pulmonary disease—usual care arm.

Section 8. Comparison of coding rates for patients in EMIS and VISION practices in the SLS COPD UC arm

When modelling the count of medical codes, and after adjusting for all covariates, we found patients registered with EMIS practices had on average 1·64 more codes recorded over the course of the year.

Table E9: Comparison of coding rates for patients in EMIS and VISION practices in the SLS COPD UC arm

Outcome = any medical code.

Parameter	Category	Relative rate	Relative rate lower 95% CI	Relative rate upper 95% CI	p value
Data source = EMIS		1·64	1·50	1·79	<0·0001
Sex	Female	0·98	0·91	1·06	0·62
Age		1·10	1·06	1·15	<0·0001
Age squared		0·99	0·96	1·02	0·45
SES IMD 2010 quintiles	4	1·25	1·00	1·56	0·05
	3	1·21	0·98	1·49	0·07
	2	1·15	0·94	1·40	0·18
	1 (most deprived)	1·44	1·19	1·74	0·0002
Current medication (prescriptions in last 3 months)	LABA only	0·86	0·63	1·17	0·34
	LAMA only	0·69	0·58	0·81	<0·0001
	ICS only	0·63	0·50	0·79	<0·0001
	LABA/LAMA	0·89	0·64	1·24	0·49
	LAMA/ICS	0·57	0·45	0·72	<0·0001
	LABA/ICS	0·64	0·56	0·74	<0·0001
	LABA/LAMA/ICS	0·69	0·61	0·78	<0·0001
History of depression		1·16	1·06	1·27	0·0009
History of anxiety		1·11	1·02	1·22	0·02
History of asthma		1·08	1·00	1·17	0·05
History of pneumonia		1·10	0·98	1·25	0·10
History of GORD/peptic ulcer		1·13	1·04	1·23	0·0047
Exacerbation history in prior 12 months		2·05	0·95	4·45	0·07
Exacerbation history squared		0·62	0·17	2·27	0·47

Parameter	Category	Relative rate	Relative rate lower 95% CI	Relative rate upper 95% CI	p value
FEV_1%		0·98	0·92	1·04	0·44
FEV_1% squared		1·01	0·98	1·04	0·58
FEV_1/FVC ratio		1·07	1·01	1·13	0·03
FEV_1/FVC ratio squared		1·00	0·97	1·03	0·90
MRC dyspnoea score	2	1·21	1·04	1·39	0·01
	3	1·29	1·11	1·49	0·0008
	4	1·38	1·17	1·62	0·0001
	5 (most breathlessness)	1·39	1·01	1·90	0·04
Smoking	Ex	1·05	0·86	1·28	0·62
	Current	1·10	0·90	1·34	0·37
Influenza vaccination		1·32	1·13	1·55	0·0004

When COPD-related medical codes were modelled, the relative rate dropped to 1·20.

CI=confidence interval. COPD=chronic obstructive pulmonary disease. EMIS=Egton Medical Information Systems. FEV_1=forced expiratory volume in 1 s. FVC=forced vital capacity. GORD=gastro-oesophageal reflux disease. ICS=inhaled corticosteroid. IMD=index of multiple deprivation. LABA=long-acting β_2-agonist. LAMA=long-acting muscarinic receptor antagonist. MRC=Medical Research Council. SES=socioeconomic status. SLS COPD UC=Salford Lung Study in chronic obstructive pulmonary disease—usual care arm.

***Table E10*: Comparison of COPD coding rates for patients in EMIS and VISION practices in the SLS COPD UC arm**

Outcome = any COPD-related medical code.

Parameter	Category	Relative rate	Relative rate lower 95% CI	Relative rate upper 95% CI	p value
Data source = EMIS		1·20	1·09	1·32	0·0002
Sex	Female	1·14	1·05	1·24	0·0021
Age		1·00	0·95	1·05	0·91
Age squared		0·93	0·90	0·96	<0·0001
SES IMD 2010 quintiles	4	0·95	0·76	1·20	0·68
	3	1·00	0·81	1·22	0·97
	2	1·01	0·83	1·23	0·95
	1 (most deprived)	0·92	0·76	1·11	0·37
Current medication (prescriptions in last 3 months)	LABA only	0·71	0·47	1·07	0·10
	LAMA only	0·79	0·66	0·96	0·02
	ICS only	0·91	0·72	1·16	0·45
	LABA/LAMA	0·64	0·40	1·03	0·06
	LAMA/ICS	0·72	0·55	0·93	0·01
	LABA/ICS	0·89	0·77	1·04	0·14
	LABA/LAMA/ICS	0·93	0·81	1·07	0·30
History of depression		1·02	0·92	1·12	0·73
History of anxiety		1·09	0·99	1·21	0·09
History of asthma		1·00	0·92	1·09	0·94
History of pneumonia		1·15	1·02	1·31	0·02
History of GORD/peptic ulcer		0·96	0·87	1·05	0·39
Exacerbation history in prior 12 months		17·01	7·62	37·98	<0·0001
Exacerbation history squared		0·04	0·01	0·15	<0·0001
FEV_1%		1·00	0·94	1·07	0·94
FEV_1% squared		1·00	0·96	1·03	0·94
FEV_1/FVC ratio		0·94	0·88	0·99	0·03
FEV_1/FVC ratio squared		1·03	1·00	1·07	0·08
MRC dyspnoea score	2	0·99	0·86	1·15	0·93
	3	1·05	0·90	1·22	0·56
	4	1·08	0·92	1·28	0·35
	5 (most breathlessness)	1·13	0·83	1·55	0·43
Smoking	Ex	0·98	0·80	1·21	0·86
	Current	0·92	0·74	1·14	0·45
Influenza vaccination		1·10	0·94	1·28	0·25

CI=confidence interval. EMIS=Egton Medical Information Systems. FEV_1=forced expiratory volume in 1 s. FVC=forced vital capacity. GORD=gastro-oesophageal reflux disease. ICS=inhaled corticosteroid. IMD=index of multiple deprivation. LABA=long-acting β_2-agonist. LAMA=long-acting muscarinic receptor antagonist. MRC=Medical Research Council. SES=socioeconomic status. SLS COPD UC=Salford Lung Study in chronic obstructive pulmonary disease—usual care arm.

Reference

1 Rothnie KJ, Müllerová H, Hurst JR, et al. Validation of the recording of acute exacerbations of COPD in UK primary care electronic healthcare records. *PLoS One* 2016; **11:** e0151357.