Incidence of COPD and quality of subsequent treatment among people with a history of using illicit opioids: a matched cohort study in England (PROTOCOL)

Date: 12 March 2021

Authors (alphabetical order)

Thomas D. Brothers, UCL
Spiros Denaxas (UCL)
Arturo Gonzalez-Izquierdo, UCL
Dan Lewer, UCL (main/corresponding author: d.lewer@ucl.ac.uk)
Prianka Padmanathan, University of Bristol
Jennifer K. Quint, Imperial College London
Muhammad Qummer Ul Arfeen, UCL
Paola Zaninotto (UCL)

1 Summary

Chronic obstructive pulmonary disease (COPD) is common among people who use illicit opioids. This study will estimate the incidence of diagnosed COPD and the rate of death due to COPD among patients in primary care in England with previous records of illicit opioid use, and compare this to patients without records of illicit opioid use. Among patients with a new COPD diagnosis, we estimate the association between illicit opioid use and the probability of preventative healthcare such as flu and pneumococcal vaccines or support with smoking cessation, and the association between illicit opioid use and adverse outcomes such as acute exacerbations and death. Data will be drawn from the Clinical Practice Research Datalink (CPRD), using a validated method to identify patients with a history of illicit opioid use. Patients without a history of illicit opioid use will be selected using a process called 'exposure density sampling' to create a cohort matched on age, sex, GP practice, and date of cohort entry.

2 Background

- **High prevalence of chronic obstructive pulmonary disease (COPD) among people who use illicit opioids.** Cross-sectional spirometry studies in community drug and alcohol services have found COPD prevalence (defined as FEV1 < 70% of predicted) of: 91/184 (49%) among people who smoke heroin in Liverpool,[1] 260/753 (35%) in a larger sample of people who smoke heroin in Liverpool,[2] 36/129 (28%) among patients at opioid agonist treatment clinics in Switzerland,[3] and a pooled value of 18% from an international systematic review of COPD prevalence among people who smoke opiates.[4] These values are particularly high given the young age of participants (often 30s-40s).
- **High mortality rates due to COPD**. In a cohort of 198,247 people who use opiates identified from criminal justice and drug treatment services, 130/3974 (3.3%) of deaths had a primary cause of COPD,

with an SMR of 12.6.[5] In a cohort of 6683 people entering treatment for heroin dependence, 48/732 (6.6%) of deaths had a primary cause of COPD, with an SMR of 19.0.[6]

- **Various mechanisms that explain the high frequency of COPD in this population**, including (1) tobacco smoking; (2) for those who smoke drugs, direct thermal injury from smoking cocaine/crack[7] and irritation of airways by particles.[8]
- **Qualitative research shows barriers to healthcare**, including: stigma and perceptions that patients are 'drug seeking'; diagnostic overshadowing (in which symptoms such as a persistent cough are not fully investigated because they are attributed to drug use); priorities such as finding accommodation or money that compete with healthcare; lack of knowledge about opioid dependence among clinicians; and bureaucratic barriers such as the need to give a home address when making an appointment.[9–13]
- **Need for research to inform healthcare models.** Although the unmet need for healthcare related to COPD is well-documented, there are few evidence-based approaches to improve access. Spirometry studies show that patients are willing to participate and receive a diagnosis, but treatment is mainly in primary care and may not be accessible. This study will use primary care data to identify the gaps in secondary prevention measures (such as immunisation and support with smoking cessation), to inform development of more accessible models of care. For example, if patients have poor access to flu and pneumococcal vaccines, these vaccines could be offered when methadone and buprenorphine are dispensed.

3 Research questions

Among people with a history of using illicit opioids:

1. What is the incidence of diagnosed COPD, and how does this compare to patients without a history of illicit opioid use?
2. What is the rate of death due to COPD, and how does this compare to patients without a history of illicit opioid use?
3. What is the likelihood of receiving evidence-based secondary prevention after a new diagnosis of COPD, and how does this compare to people without a history of illicit opioid use?
4. What is the likelihood of COPD exacerbations and death after a new diagnosis of COPD, and how does this compare to people without a history of illicit opioid use?
5. Are those who die due to COPD less likely to receive a diagnosis in primary care prior to death?

4 Population

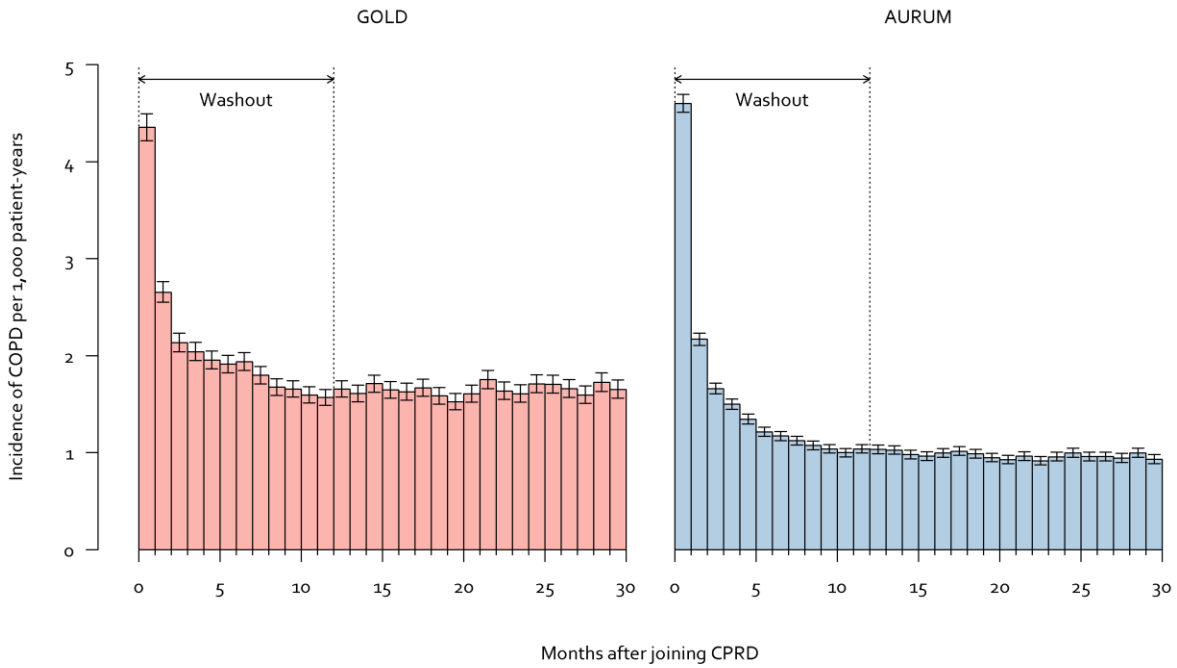
4.1 Base cohort for estimation of COPD incidence

The population of people with a history of using illicit opioids will be drawn from the Clinical Practice Research Datalink (CPRD) GOLD and AURUM databases.[14,15] A codelist for illicit opioid use ('HUPIO') has

already been developed and validated.[16] It includes patients with prescriptions of opioid agonist therapy (methadone or buprenorphine) and clinical observations such as 'heroin dependence'. An application of the phenotype in June 2020 identified 30,491 patients in GOLD and 108,270 in AURUM, with similar demographic characteristics and mortality rates to those reported in other studies of people who use illicit opioids.[16]

Patients will have a 'washout' period of 12 months. This means that patients will enter the cohort at the latest 12 months after they join CPRD and the earliest record of illicit opioid use. The purpose of the washout period is to avoid healthcare records transferred from a previous GP practice that have the wrong date. You can see this problem by reporting the incidence of COPD stratified by time after joining CPRD. This is shown in the figure below, using the method described by Lewis et al.[17]

Figure 1: Incidence of COPD in CRPD GOLD and AURUM, stratified by months after joining CPRD

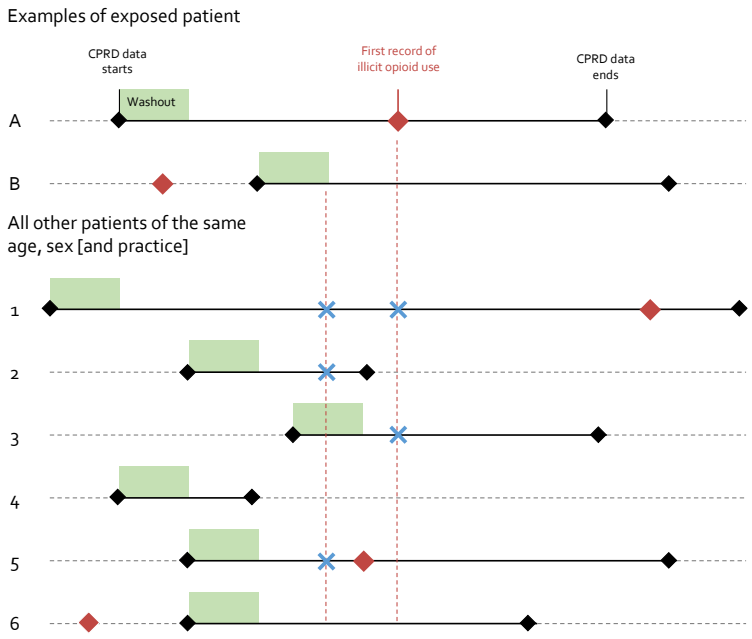


A matched comparison group of people without a history of using illicit opioids will be selected using a process called 'exposure density sampling'.[18] For each exposed individual, we will sample *m* individuals (we will set *m* to 5) at random from the patients who are unexposed at the time the exposed patient joined the cohort. Unexposed patients are assigned the same cohort entry date as their corresponding exposed patients, and may later become exposed (figure 1). Sampling is done with replacement such that unexposed patients may be included more than once. Recommended analysis with survival methods involves deduplication and cohort entry at the earliest of the sampled dates.[18]

The main reason for this type of matching is to provide a cohort entry date for the unexposed population (which would otherwise default to the CPRD entry date), and to break the link between the observation period and the outcome. As well as cohort entry date, we will match patients by age (+/- 3 years), sex, and GP practice, which will control differences in age, sex, geographical deprivation, and clinical practice. Including these variables in the matching algorithm improves statistical efficiency, because relying on

adjustment using multivariable regression would place a lot of weight on young male participants in the unexposed group.

Figure 2: Exposure density sampling to create a comparison group of patients without a history of illicit opioid use. Blue crosses represent potential matches from which the unexposed group is sampled



After running a preliminary matching exercise, we observed that unexposed patients have a longer duration of data in CPRD prior to cohort entry. This is because the unexposed group rarely enter the cohort immediately after the washout period, while many in the exposed group do – e.g. patient B in figure 2. The unexposed group also has longer duration of follow-up after cohort entry, which may be because people who use illicit opioids are more mobile and move between practices more often. The distribution of follow-up duration before and after cohort entry in CPRD AURUM is shown in figure 3.

Different durations of data prior to cohort entry will create bias in estimates of participant characteristics at baseline derived from historical primary care records, including prevalent COPD and comorbidities. Use of linked hospital data to measure co-morbidities can help because the observation period for hospital data is similar for all individuals (starting in 1998).

Different durations of data after cohort entry will be reflected in regression models (e.g. through right-censoring), and means that matching variables need to be included in multivariable models as groups may be become unbalanced during follow-up, particularly because the unexposed group will have more follow-up at older ages.[19]

Figure 3: follow-up durations in CPRD AURUM

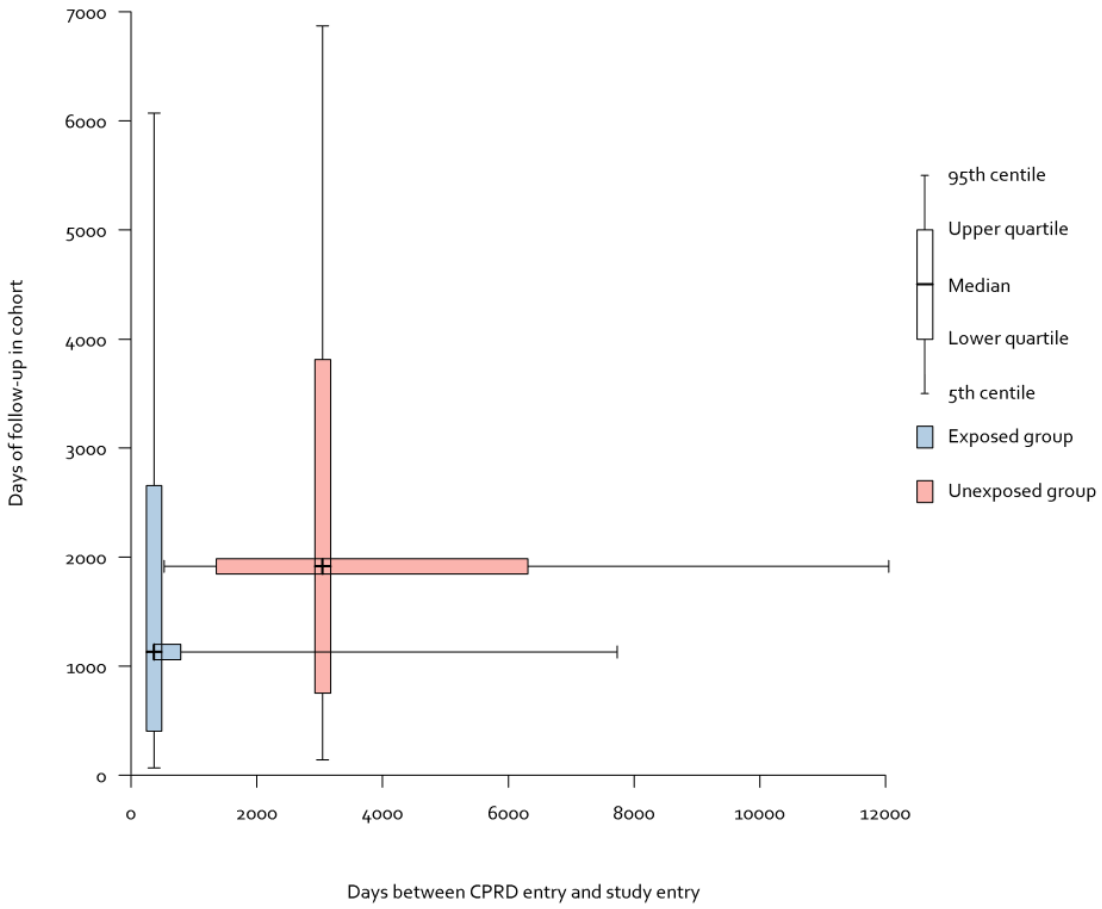


Table 1 shows the characteristics of participants from the preliminary matching exercise. Codelists for smoking and BMI and given in section 4. The exposed cohort (those with a history of using illicit opioids) have lower average BMI and a higher prevalence of underweight, and higher prevalence of current smoking.

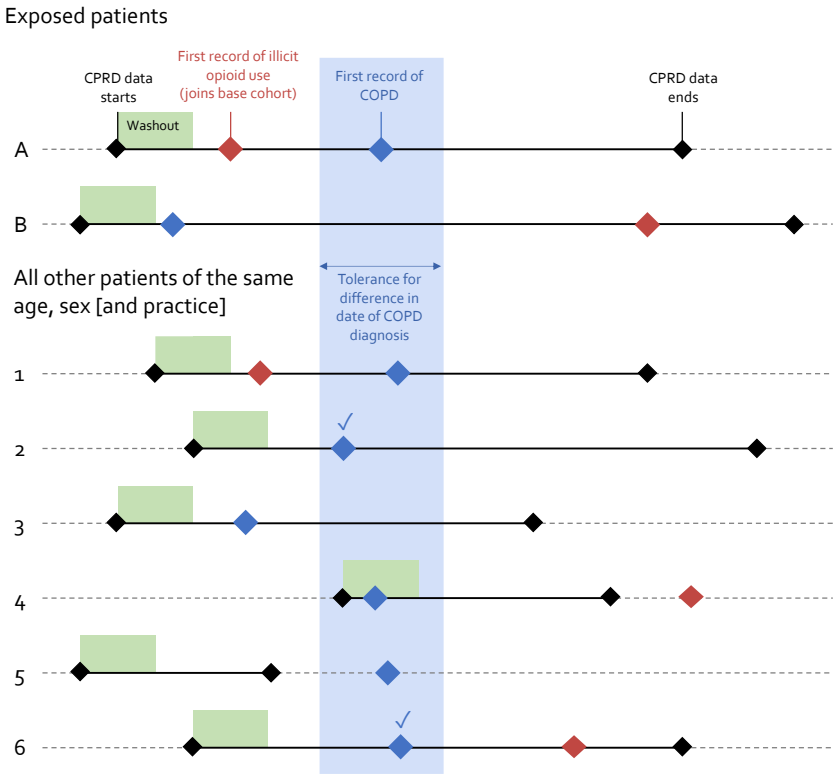
Table 1: baseline characteristics. Age, sex, region, and year of cohort entry are matched.

		AURUM		GOLD	
Variable	Level	Exposed	Unexposed	Exposed	Unexposed
Total		88,879 (100.0)	444,367 (100.0)	24,683 (100.0)	123,415 (100.0)
Follow-up duration	Median (IQR)	3.1 (1.1-7.3)	5.3 (2.1-10.5)	2.9 (1.1-6.5)	4.7 (1.9-8.9)
	Mean (sd)	5.0 (5.2)	6.9 (5.8)	4.4 (4.3)	5.9 (4.7)
Age at cohort entry	Under 18	-	2,677 (0.6)	-	956 (0.8)
	18-24	9,868 (11.1)	47,672 (10.7)	3,411 (13.8)	16,247 (13.2)
	25-34	32,993 (37.1)	163,153 (36.7)	9,828 (39.8)	48,415 (39.2)
	35-44	28,421 (32.0)	142,063 (32.0)	7,341 (29.7)	36,998 (30.0)
	45-54	13,347 (15.0)	67,200 (15.1)	3,077 (12.5)	15,654 (12.7)
	55-64	4,250 (4.8)	20,668 (4.7)	1,026 (4.2)	4,881 (4.0)
	Over 65	-	934 (0.2)	-	264 (0.2)
	Median (IQR)	35.4 (29.2-42.9)	35.5 (29.2-43.0)	34.1 (28.1-41.4)	34.2 (28.1-41.5)
	Mean (sd)	36.6 (9.8)	36.6 (9.9)	35.4 (9.6)	35.4 (9.8)
Sex	Male	61,317 (69.0)	306,564 (69.0)	17,083 (69.2)	85,415 (69.2)
	Female	27,562 (31.0)	137,803 (31.0)	7,600 (30.8)	38,000 (30.8)
Region	East Midlands	1,543 (1.7)	7,715 (1.7)	820 (3.3)	4,100 (3.3)
	East of England	2,743 (3.1)	13,715 (3.1)	2,143 (8.7)	10,715 (8.7)
	London	12,949 (14.6)	64,745 (14.6)	2,660 (10.8)	13,300 (10.8)
	North East	5,208 (5.9)	26,040 (5.9)	870 (3.5)	4,350 (3.5)
	North West	18,304 (20.6)	91,520 (20.6)	5,419 (22.0)	27,095 (22.0)
	South Central	7,864 (8.8)	39,320 (8.8)	2,395 (9.7)	11,975 (9.7)
	South East Coast	3,854 (4.3)	19,270 (4.3)	2,138 (8.7)	10,690 (8.7)
	South West	17,367 (19.5)	86,835 (19.5)	3,961 (16.0)	19,805 (16.0)
	West Midlands	14,915 (16.8)	74,547 (16.8)	3,221 (13.0)	16,105 (13.0)
	Yorkshire & The Humber	4,082 (4.6)	20,410 (4.6)	1,056 (4.3)	5,280 (4.3)
	Missing	50 (0.1)	250 (0.1)	-	-
Year of cohort entry	1998-2001	12,759 (14.4)	63,789 (14.4)	4,152 (16.8)	20,760 (16.8)
	2002-2005	13,458 (15.1)	67,290 (15.1)	5,207 (21.1)	26,035 (21.1)
	2006-2009	16,928 (19.0)	84,635 (19.0)	6,333 (25.7)	31,665 (25.7)
	2010-2013	17,253 (19.4)	86,248 (19.4)	6,018 (24.4)	30,090 (24.4)
	2014-2017	18,145 (20.4)	90,725 (20.4)	2,698 (10.9)	13,490 (10.9)
	2018-2020	10,336 (11.6)	51,680 (11.6)	275 (1.1)	1,375 (1.1)
BMI	Underweight (<18.5)	4,552 (5.1)	9,626 (2.2)	1,288 (5.2)	2,592 (2.1)
	Healthy (18.5 – 24.9)	36,760 (41.4)	146,084 (32.9)	10,385 (42.1)	41,296 (33.5)
	Overweight (25 – 29.9)	17,383 (19.6)	116,549 (26.2)	4,476 (18.1)	32,176 (26.1)
	Obese (30 – 39.9)	9,636 (10.8)	68,519 (15.4)	2,444 (9.9)	18,204 (14.8)
	Severely obese (40+)	1,600 (1.8)	10,344 (2.3)	374 (1.5)	2,578 (2.1)
	Missing	18,948 (21.3)	93,245 (21.0)	5,716 (23.2)	26,569 (21.5)
	Median (IQR)	23.8 (21.0-27.6)	25.6 (22.7-29.4)	23.5 (20.8-27.2)	25.5 (22.6-29.2)
	Mean (sd)	25.0 (5.7)	26.6 (5.7)	24.7 (5.6)	26.4 (5.6)
Smoking	Never	6,334 (7.1)	202,334 (45.5)	1,617 (6.6)	57,988 (47.0)
	Ex-smoker	6,189 (7.0)	54,917 (12.4)	1,619 (6.6)	16,489 (13.4)
	Current smoker	69,797 (78.5)	153,342 (34.5)	18,674 (75.7)	36,816 (29.8)
	No records	6,559 (7.4)	33,774 (7.6)	2,773 (11.2)	12,122 (9.8)

4.2 Population for analysis of outcomes after a new diagnosis of COPD

Patients with a new diagnosis of COPD and a history of using illicit opioids (i.e. the first record of COPD comes after the first record of illicit opioid use) will be included in a second stage of analysis. These patients will be matched to a comparison group of patients with a new diagnosis of COPD but no history of illicit opioid use (figure 2). Matching will be by age (+/- 3 years), sex, and date of COPD diagnosis (+/- 12 months). We will also aim to match by GP practice if there are sufficient eligible patients.

Figure 4: Exposure density sampling to create a comparison group of people with a COPD diagnosis but no history of illicit opioid use. In this example, patient A has a new diagnosis after cohort entry, while patient B has prevalent COPD at cohort entry and is excluded. Ticks represent potential matches from which the unexposed group for parient A is sampled.



The table below shows characteristics of participants after a preliminary matching exercise. Phenotyping algorithms used to define key are described in section 4. Participants with incident COPD are older than the underlying cohort (which is partly because COPD is age-related and partly because the incident COPD diagnoses are after cohort entry by design), more likely to be underweight, and more likely to be current smokers (particularly in the unexposed group, as smoking prevalence is already very high in the underlying exposed cohort). Comparing COPD patients with and without a history of illicit opioid use, those with illicit opioid use were more likely to be underweight, more likely to be current smokers, and more likely to have severe COPD based on spirometry and the MRC breathlessness scale. COPD severity has a large amount of missing data (approximately one-third of participants for both variables).

Table 2: baseline characteristics for patients with incident COPD, stratified by history of illicit opioid use. Number (%)

		AURUM		GOLD	
Variable	Level	Exposed	Unexposed	Exposed	Unexposed
Total		3,318 (100.0)	16,590 (100.0)	585 (100.0)	2,925 (100.0)
Follow-up	Median [IQR]	3.1 [1.3-5.9]	3.7 [1.7-6.9]	2.5 [1.0-4.9]	3.2 [1.5-5.7]
	Mean [sd]	4.1 [3.6]	4.7 [3.9]	3.4 [3.0]	4.0 [3.2]
Age	18-24	4 (0.1)	20 (0.1)	1 (0.2)	7 (0.2)
	25-34	112 (3.4)	534 (3.2)	20 (3.4)	92 (3.1)
	35-44	935 (28.2)	4,482 (27.0)	169 (28.9)	846 (28.9)
	45-54	1,490 (44.9)	7,460 (45.0)	259 (44.3)	1,274 (43.6)
	55-64	777 (23.4)	3,888 (23.4)	136 (23.2)	672 (23.0)
	65+	-	206 (1.2)	-	34 (1.2)
	Median [IQR]	48.8 [43.5-54.4]	49.1 [43.7-54.9]	48.3 [43.1-54.5]	48.6 [43.6-54.8]
	Mean [sd]	48.9 [7.8]	49.2 [7.9]	48.7 [7.8]	49.0 [7.9]
Sex	Male	2,132 (64.3)	10,660 (64.3)	375 (64.1)	1,875 (64.1)
	Female	1,186 (35.7)	5,930 (35.7)	210 (35.9)	1,050 (35.9)
Region	East Midlands	51 (1.5)	376 (2.3)	8 (1.4)	56 (1.9)
	East of England	101 (3.0)	729 (4.4)	52 (8.9)	261 (8.9)
	London	549 (16.5)	2,737 (16.5)	80 (13.7)	450 (15.4)
	North East	144 (4.3)	1,195 (7.2)	22 (3.8)	69 (2.4)
	North West	1,066 (32.1)	3,385 (20.4)	189 (32.3)	658 (22.5)
	South Central	244 (7.4)	1,583 (9.5)	44 (7.5)	293 (10.0)
	South East Coast	124 (3.7)	1,073 (6.5)	42 (7.2)	425 (14.5)
	South West	507 (15.3)	2,132 (12.9)	78 (13.3)	332 (11.4)
	West Midlands	432 (13.0)	2,676 (16.1)	60 (10.3)	298 (10.2)
	Yorkshire & The Humber	100 (3.0)	703 (4.2)	10 (1.7)	83 (2.8)
	Missing	0 (0.0)	1 (0.0)	-	-
Year of diagnosis	1998-2001	32 (1.0)	175 (1.1)	4 (0.7)	25 (0.9)
	2002-2005	173 (5.2)	878 (5.3)	51 (8.7)	249 (8.5)
	2006-2009	347 (10.5)	1,704 (10.3)	129 (22.1)	644 (22.0)
	2010-2013	752 (22.7)	3,721 (22.4)	209 (35.7)	1,049 (35.9)
	2014-2017	1,206 (36.3)	6,181 (37.3)	168 (28.7)	824 (28.2)
	2018-2020	808 (24.4)	3,931 (23.7)	24 (4.1)	134 (4.6)
BMI	Underweight (<18.5)	369 (11.1)	620 (3.7)	65 (11.1)	119 (4.1)
	Healthy (18.5 – 24.9)	1,388 (41.8)	5,532 (33.3)	235 (40.2)	941 (32.2)
	Overweight (25 – 29.9)	751 (22.6)	4,996 (30.1)	129 (22.1)	878 (30.0)
	Obese (30 – 39.9)	581 (17.5)	4,203 (25.3)	108 (18.5)	779 (26.6)
	Severely obese (40+)	121 (3.6)	886 (5.3)	25 (4.3)	139 (4.8)
	Missing	108 (3.3)	353 (2.1)	23 (3.9)	69 (2.4)
	Median [IQR]	24.2 [20.5-29.1]	26.7 [23.0-31.3]	24.2 [20.8-29.7]	26.8 [23.0-31.4]
	Mean [sd]	25.5 [6.9]	27.8 [6.7]	25.7 [6.9]	27.8 [6.7]
Smoking	Never	49 (1.5)	1,919 (11.6)	12 (2.1)	357 (12.2)
	Ex-smoker	369 (11.1)	3,354 (20.2)	64 (10.9)	752 (25.7)
	Current smoker	2,855 (86.0)	11,027 (66.5)	504 (86.2)	1,776 (60.7)
	Missing	45 (1.4)	290 (1.7)	5 (0.9)	40 (1.4)
COPD stage	Mild (FEV1 >=80% predicted)	146 (4.4)	261 (1.6)	45 (7.7)	79 (2.7)
	Moderate (FEV1>=50% predicted)	498 (15.0)	1,608 (9.7)	107 (18.3)	358 (12.2)
	Severe (FEV1>=30% predicted)	1,067 (32.2)	6,347 (38.3)	231 (39.5)	1,307 (44.7)
	Very severe (FEV1<30% predicted)	524 (15.8)	3,800 (22.9)	87 (14.9)	676 (23.1)
	No data	1,083 (32.6)	4,574 (27.6)	115 (19.7)	505 (17.3)
	Median [IQR]	64.0 [47.0-79.0]	71.0 [57.0-84.0]	61.1 [44.4-76.2]	68.8 [54.9-81.6]
	Mean [sd]	63.0 [21.7]	70.3 [20.1]	60.6 [22.1]	68.2 [19.8]
MRC breathlessness grade	1 (least severe)	292 (8.8)	3,664 (22.1)	39 (6.7)	535 (18.3)
	2	820 (24.7)	4,542 (27.4)	137 (23.4)	736 (25.2)
	3	639 (19.3)	2,057 (12.4)	108 (18.5)	327 (11.2)
	4	321 (9.7)	665 (4.0)	56 (9.6)	118 (4.0)
	5 (most severe)	61 (1.8)	101 (0.6)	12 (2.1)	13 (0.4)
	No data	1,185 (35.7)	5,561 (33.5)	233 (39.8)	1,196 (40.9)

5 Phenotyping algorithms

5.1 Incident COPD

New diagnoses of COPD in COPD GOLD will be based on a validated codelist,[20] which has an estimated positive predictive value of 87%. We created a new phenotype algorithm for AURUM derived by applying the following search terms in the AURUM SNOMED data dictionary: "copd", "chronic obstruct*", "bronchitis", "emphysema". The list of SNOMED concepts is available at: https://github.com/danlewer/hupio/blob/main/codelists/aurum_copd.csv/ (concept identifiers are prefixed with 'x' to avoid data type conversion). The AURUM phenotype appears less sensitive than the GOLD phenotype, as the incidence of COPD using these phenotypes is lower in AURUM (figure 1).

5.2 Outcomes after incident COPD

Secondary prevention: Where an individual is diagnosed with COPD, NICE guidance[21] recommends support with smoking cessation (including psychological support and/or prescription of NRT or a smoking cessation drug such as bupropion), seasonal influenza vaccine, pneumococcal vaccine, referral to pulmonary rehab, and COPD-specific drugs (inhaled corticosteroids and/or bronchodilators; SAMA/SABA/LAMA/LABA).

Adverse outcomes: Unplanned hospital admissions due to acute exacerbations of COPD, unplanned hospital admissions with a primary respiratory cause, all-cause mortality, and death with an underlying respiratory cause.

Table 3: outcomes for participants with incident COPD

Outcome	Timeframe	Exclusions	Model type	Derivation in GOLD	Derivation in AURUM
Secondary prevention (healthcare)					
Smoking cessation support	Within 12 months of diagnosis	No records of current smoking	Poisson with binary outcome	Medcodes and prodcodes for relevant prescriptions (NRT, varenicline, or bupropion), or psychological support - https://github.com/danlewer/hupio/blob/main/codelists/smoking_cessation.csv	
Seasonal influenza vaccine	Separate follow-up in each flu season after diagnosis	None – participants vaccinated prior to diagnosis in the first season will be considered vaccinated	Poisson with binary outcome	Immunisation table, immstype 4, 84, 85, 89, 97, 100, 101, 102, 105, 106, 116-126	Medcodes for relevant clinical observations and prescriptions: https://github.com/danlewer/hupio/blob/main/codelists/aurum_flu_vaccine.csv
Pneumococcal vaccine	Within 12 months of diagnosis	Pneumococcal vaccine before COPD diagnosis	Poisson with binary outcome	Immunisations table, immstype 13, 18, 28, 82, 115	Medcodes for relevant clinical observations and prescriptions: https://github.com/danlewer/hupio/blob/main/codelists/aurum_pneumo_vaccine.csv
Referral to pulmonary rehab	Within 12 months of diagnosis	Prior diagnosis of rheumatoid arthritis, depression, heart failure, or stroke[22]	Poisson with binary outcome	Read codes 8FA2.00, 8FA..00, 8H7u.00, Z678.00, 8FA1.00, 8FA0.00	Medcodes 25607012, 1485151019, 1476256018, 619110000000113, 1476257010
COPD-specific drugs	Within 12 months of diagnosis	None	Poisson with binary outcome	Prodcodes for prescriptions: https://github.com/danlewer/hupio/blob/main/codelists/copd_meds.csv	
Adverse outcomes					
Acute exacerbation of COPD	Until end of follow-up	None	Cox proportional hazards (time until first event)	Unplanned hospital admission with a primary cause of ICD-10 J41-44, or J44.0/1 in any position23, from linked Hospital Episode Statistics	
Unplanned hospital admission with primary respiratory cause	Until end of follow-up	None	Cox proportional hazards (time until first event)	Unplanned hospital admission with a primary cause of ICD-10 J00-J99, from linked Hospital Episode Statistics	
All-cause death	Until end of follow-up	None	Cox proportional hazards (time until first event)	From linked ONS mortality data	
Death with underlying respiratory cause	Until end of follow-up	None	Cox proportional hazards (time until first event)	Underlying cause is ICD-10 J00-J99, from linked ONS mortality data	

5.3 Covariates

Covariates will include disease severity measured by FEV1/predicted FEV[24] and the MRC breathlessness grade, BMI, smoking status, and comorbidities.

Table 4: covariates for analysis of secondary prevention and adverse outcomes among people with incident COPD

Covariate	Timeframe	Derivation in GOLD	Derivation in AURUM
Forced Exhaled Volume in 1 second (FEV1) / predicted FEV1	Most recent data in 12 months before diagnosis (except height and ethnicity which will be taken from any record)	FEV1 from Test table, enttype 394 Predicted derived from sex, height (Clinical table, enttype 14), age, and ethnicity (https://www.caliberresearch.org/portal/phenotypes/ethnicity) Algorithm: https://gist.github.com/danlewer/dcc13f0d01d2a0dd4c8266690927b9fa	Medcode 457081010
MRC breathlessness grade	Most recent data in 12 months before diagnosis	Medcodes: Stage 1: 19432 Stage 2: 19427 Stage 3: 19426 Stage 4: 19430 Stage 5: 19429	Medcodes: Stage 1: 1485144011 Stage 2: 1485147016 Stage 3: 1485148014 Stage 4: 1485149018 Stage 5: 1485150018
BMI	Most recent data in 12 months before diagnosis	Clinical table enttype 13	Medcode 100716012
Smoking status	Most recent data in 12 months before diagnosis	Existing phenotype: https://www.caliberresearch.org/portal/show/smoking_status_gprd	Codes for relevant clinical observations: https://github.com/danlewer/hupio/blob/main/codelists/aurum_smoking.csv
Comorbidities	Three years before diagnosis	1. Number of unique ICD-10 chapters recorded in any diagnostic position in Finished Consultant Episodes (from Hospital Episode Statistics Admitted Patient Care) starting in the three years prior to cohort entry, from chapters 2–14 and 17, excluding chapters such as infections where an admission may not represent a long-term condition. 2. Charlson Comorbidity Index[25] based on ICD-10 diagnoses recorded in Finished Consultant Episodes (from Hospital Episode Statistics Admitted Patient Care) starting in the three years prior to cohort entry.	

6 Analysis

6.1 Incidence of COPD

We will report the number of patients with prevalent COPD (i.e. a COPD code prior to cohort entry in either primary care or Hospital Episode Statistics Admitted Patient Care data) in the exposed and unexposed groups. We will then exclude these patients and use a left-truncated Cox proportional hazards model[18] to estimate the hazard ratio of COPD comparing patients with and without a history of illicit opioid use. Age group and opioid use will be included as time-varying covariates. A second model will adjust for smoking status at baseline.

6.2 Mortality due to COPD

We will use a left-truncated Cox proportional hazards model to compare the risk of death due to COPD (defined as deaths with an underlying cause of ICD-10 J41-44) in the whole cohort. We will not exclude prevalent cases of COPD from this cohort. Participants will be right-censored at death due to other causes (i.e. ignoring competing risks), with sensitivity analysis where death due to all other causes is considered a competing risk using the 'proportional subdistribution hazards' model described by Fine and Gray[26] and implemented in the R package 'cmprisk'. Age group and opioid use will be time-varying.

Among people who died with an underlying cause of COPD, we will describe the time between diagnosis in primary care and death. A smaller proportion of diagnoses (or a shorter time between diagnosis and death) among people with a history of using illicit opioids may suggest later diagnosis or poor healthcare access.

6.3 Outcomes after a new diagnosis of COPD

Healthcare-related outcomes (smoking cessation, immunisations, pulmonary rehabilitation, and COPD medications) will be analysed as binary outcomes, where the outcomes shows whether the intervention was provided at least once in the 12 months after COPD diagnosis (including the day of diagnosis). We will estimate risk ratios using Poisson regression.

Participants are eligible for seasonal flu immunisation every flu season after diagnosis. Participants will be given separate follow-up periods for each flu season (defined as 1 September to 31 March) after diagnosis. The calendar year of the season and the number of seasons after diagnosis will be included as independent variables. Where participants exited the cohort before the end of the flu season, that flu season will be excluded from analysis.

Some outcomes are not relevant for all participants. For example, support to stop smoking will not be offered to people who do not smoke. These participants will be excluded. This may lead to the exposed and unexposed groups may become unbalanced in terms of the matching variables, which means that analyses of these outcomes must be adjusted by the matching variables (including the date/year of cohort entry or diagnosis).

Adverse outcomes (hospitalization and mortality) will be analysed as time-to-first-event, using a left-truncated Cox proportional hazards model. Follow-up will continue until the participant experiences the event or exits the cohort.

For each outcome, we will fit the following models:

- An unadjusted model estimating the relative risk of each outcome
- A model adjusted for participant characteristics and disease stage/type and co-morbidities at baseline.
- Models of adverse outcomes among participants with incident COPD will additionally adjusted for healthcare (immunisation, pulmonary rehab, etc.), to test whether healthcare access could explain differences in outcomes.

The primary analysis will use 'missing' categories for variables with missing data (to allow missingness to be independently associated with outcomes). We will also conduct an analysis using multiple imputation to explore potential biases resulting from missing data.

7 Limitations

- COPD is a heterogeneous disease, and is difficult to characterize, especially using electronic health records. It is possible that differences in outcomes could be partially explained by unmeasured differences in the type of disease.
- There does not appear to be a well-recorded measure of smoking intensity / duration (that might allow calculation of pack-years, for example) in GP records.
- People with a history of illicit opioid use in this study will differ to people with a history of illicit opioid use not included in this study (i.e. selection bias). This is particularly important for the analysis of COPD incidence, because inclusion in the study may be associated with poor health (as those with more GP appointments are likely to have more opportunity to disclose drug use). The direction of bias is therefore likely to be that incidence differences are overstated.
- The study of healthcare and adverse outcomes among incident cases is a study of care after diagnosis, while many cases are likely to be undiagnosed.

9 Further information

9.1 Ethics

The study was approved by the MHRA (UK) Independent Scientific Advisory Committee and 19_142R, under Section 251 (NHS Social Care Act 2006). This study is based in part on data from the Clinical Practice Research Datalink obtained under license from the UK Medicines and Healthcare products Regulatory Agency. The data are provided by patients and collected by the NHS as part of their care and support. The interpretation and conclusions contained in this study are those of the authors alone.

9.2 Funding

DL is funded by a National Institute of Health Research Doctoral Research Fellowship [DRF-2018-11-ST2-016]. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

10 References

1 Furlong J, Peat R, Young R, *et al.* Prevalence of COPD in heroin smokers screened at substance misuse clinics. *European Respiratory Journal* 2016; **48**. DOI:10.1183/13993003.congress-2016.PA4178.

2 Burhan H, Young R, Byrne T, *et al.* Screening Heroin Smokers Attending Community Drug Services for COPD. *Chest* 2019; **155**: 279–87. DOI:10.1016/j.chest.2018.08.1049.

3 Grischott T, Falcato L, Senn O, Puhan MA, Bruggmann P. Chronic obstructive pulmonary disease (COPD) among opioid-dependent patients in agonist treatment. A diagnostic study. *Addiction* 2019; **114**: 868–76. DOI:10.1111/add.14559.

4 Hulin J, Brodie A, Stevens J, Mitchell C. Prevalence of respiratory conditions among people who use illicit opioids: a systematic review. 2019; : 18.

5 Pierce M, Bird SM, Hickman M, Millar T. National record linkage study of mortality for a large cohort of opioid users ascertained by drug treatment or criminal justice sources in England, 2005–2009. *Drug and Alcohol Dependence* 2015; **146**: 17–23. DOI:10.1016/j.drugalcdep.2014.09.782.

6 Lewer D, Tweed EJ, Aldridge RW, Morley KI. Causes of hospital admission and mortality among 6683 people who use heroin and crack cocaine, comparing relative and absolute risks: a population-based cohort study. *Drug and Alcohol Dependence* 2019. DOI:10.1016/j.drugalcdep.2019.06.027.

7 Restrepo CS, Carrillo JA, Martínez S, Ojeda P, Rivera AL, Hatta A. Pulmonary Complications from Cocaine and Cocaine-based Substances: Imaging Manifestations. *RadioGraphics* 2007; **27**: 941–56. DOI:10.1148/rg.274065144.

8 Cruz R, Davis M, O'Neil H, *et al.* Pulmonary manifestations of inhaled street drugs. *Heart and Lung* 1998; **27**: 297–305. DOI:10.1016/S0147-9563(98)90049-8.

9 Neale J, Tompkins C, Sheard L. Barriers to accessing generic health and social care services: a qualitative study of injecting drug users: Drug injectors and barriers to service use. *Health & Social Care in the Community* 2007; **16**: 147–54. DOI:10.1111/j.1365-2524.2007.00739.x.

10 Ross LE, Vigod S, Wishart J, *et al.* Barriers and facilitators to primary care for people with mental health and/or substance use issues: a qualitative study. *BMC Family Practice* 2015; **16**. DOI:10.1186/s12875-015-0353-3.

11 Mclaughlin D, Mckenna H, Leslie J, Moore K, Robinson J. Illicit drug users in Northern Ireland: perceptions and experiences of health and social care professionals. *Journal of Psychiatric and Mental Health Nursing* 2006; **13**: 682–6. DOI:10.1111/j.1365-2850.2006.01015.x.

12 Weiss L, McCoy K, Kluger M, Finkelstein R. Access to and use of health care: perceptions and experiences among people who use heroin and cocaine. *Addiction Research & Theory* 2004; **12**: 155–65. DOI:10.1080/16066350310000155099.

13 van Boekel LC, Brouwers EPM, van Weeghel J, Garretsen HFL. Stigma among health professionals towards patients with substance use disorders and its consequences for healthcare delivery: Systematic review. *Drug and Alcohol Dependence* 2013; **131**: 23–35. DOI:10.1016/j.drugalcdep.2013.02.018.

14 Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology* 2015; **44**: 827–36. DOI:10.1093/ije/dyv098.

15 Wolf A, Dedman D, Campbell J, *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *International Journal of Epidemiology* 2019; **48**: 1740–1740g. DOI:10.1093/ije/dyz034.

16 Lewer D, Padmanathan P, Qummer ul Arfeen M, *et al.* Healthcare use by people who use illicit opioids (HUPIO): development of a cohort based on electronic primary care records in England [version 1; peer review: 1 approved]. *Wellcome Open Research* 2020; **5**. DOI:10.12688/wellcomeopenres.16431.1.

17 Lewis JD, Bilker WB, Weinstein RB, Strom BL. The relationship between time since registration and measured incidence rates in the General Practice Research Database. *Pharmacoepidem Drug Safe* 2005; **14**: 443–51. DOI:10.1002/pds.1115.

18 Ohneberg K, Beyersmann J, Schumacher M. Exposure density sampling: Dynamic matching with respect to a time-dependent exposure. *Statistics in Medicine* 2019; **38**: 4390–403. DOI:10.1002/sim.8305.

19 Sjölander A, Greenland S. Ignoring the matching variables in cohort studies - when is it valid and why? *Statist Med* 2013; **32**: 4696–708. DOI:10.1002/sim.5879.

20 Quint JK, Mullerova H, DiSantostefano RL, *et al.* Validation of chronic obstructive pulmonary disease recording in the Clinical Practice Research Datalink (CPRD-GOLD). *BMJ Open* 2014; **4**: e005540–e005540. DOI:10.1136/bmjopen-2014-005540.

21 National Institute for Health and Care Excellence. NICE guideline [NG115]: Chronic obstructive pulmonary disease in over 16s: diagnosis and management. 2018. https://www.nice.org.uk/guidance/ng115 (accessed Jan 18, 2021).

22 Moore E, Newson R, Joshi M, *et al.* Effects of Pulmonary Rehabilitation on Exacerbation Number and Severity in People With COPD. *Chest* 2017; **152**: 1188–202. DOI:10.1016/j.chest.2017.05.006.

23 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. DOI:10.1371/journal.pone.0151357.

24 Global Initiative for Chronic Obstructive Lung Disease. Pocket Guide to COPD Diagnosis, Management and Prevention. 2018. https://goldcopd.org/wp-content/uploads/2018/02/WMS-GOLD-2018-Feb-Final-to-print-v2.pdf (accessed June 18, 2019).

25 Quan H, Li B, Couris CM, *et al.* Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. *American Journal of Epidemiology* 2011; **173**: 676–82. DOI:10.1093/aje/kwq433.

26 Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* 1999; **94**: 496–509. DOI:10.1080/01621459.1999.10474144.