Real World Effects of Medications for Chronic Obstructive Pulmonary Disease: protocol for a UK population-based non-interventional cohort study with validation against randomised trial results.

Summary of Research:

Chronic obstructive pulmonary disease (COPD) affects 3 million UK citizens. It is a progressive disease characterised by airflow obstruction, with acute exacerbations requiring urgent treatment, and often hospitalisation. Smoking cessation is the most effective intervention, but long term pharmacological treatment is also generally required. Evidence for treatment guidelines is based on randomised trial results, but there are important evidence gaps due to differences between trial patients and the intended target population; all UK patients with COPD.

Evidence of medication effects in routine clinical care is needed to understand the balance of treatment risks and benefits. Randomised trials are vital to establish treatment efficacy and safety before licensing, but treatment effects are not always the same when used in the general population and national drug licensing authorities are increasingly seeking real world evidence on which to make decisions. Additionally, large quantities of electronic health records are becoming progressively more available to researchers. The combination of demands for evidence and increased data availability means we now need rigorous methods and established best practice to translate complex data into evidence of medication effects.

This project brings together a multidisciplinary team of epidemiologists, medical statisticians, primary and secondary care physicians from international centres of excellence. We will use data from the UK Clinical Practice Research Datalink (CPRD) linked with hospital episode statistics (HES) to select a representative sample of UK patients with COPD. This world-renowned database has comprehensive medical records for a nationally representative sample of ~8% of the UK population. The main challenge for this project is to make unbiased comparisons between treatments allowing inference about treatment effects, for which we will use advanced propensity score techniques. Research objectives developed with input from people with COPD are:

1. Optimise the use of CPRD data for estimating COPD treatment effects by replicating findings of the landmark COPD trial TORCH. Using individual patient data from TORCH, we will assemble a cohort in the CPRD with similar characteristics to TORCH participants and test whether observational data can generate comparable results to trials, using cohort methodology.
2. Determine risks and benefits of COPD treatments in people excluded from TORCH. Outcomes are pneumonia, COPD exacerbation, mortality and time to treatment change. Groups to be studied include the elderly (>80 years), people with substantial comorbidity and people with and without underlying cardiovascular disease.
3. Determine the risks and benefits of currently available pharmacological COPD management in mild COPD to determine the best strategies for people in whom evidence is lacking.

Feasibility work suggests the CPRD has sufficient power to demonstrate a change in annual exacerbation rate from 0.5 to 0.4 per patient, for all study questions. By answering these questions, we will obtain important evidence about the effects of drugs for COPD, using highly efficient methods with in-built robust validation against the results of a randomised trial. This validation step will determine whether or not the observational data from the CPRD are appropriate to answer these kinds of questions. The methodological framework we will develop will also be applicable for future questions about the effectiveness of drugs as used in the NHS.

Background and Rationale

Chronic obstructive pulmonary disease (COPD) affects 3 million people in the UK [1]. The most common cause is smoking, and patients exhibit airflow obstruction that is not fully reversible. The disease is progressive, with declining lung function and a worsening of symptoms. Most troublesome are acute exacerbations manifested as a sudden worsening of symptoms e.g. severe coughing, shortness of breath and chest congestion, requiring urgent treatment, and possibly hospitalisation. Whilst smoking cessation remains the most effective intervention, the rate of exacerbation can be reduced by regular medication such as combination long acting betaadrenoceptor agonists (LABAs), and inhaled corticosteroids (ICS) or long-acting muscarinic antagonists (LAMAs).

COPD treatment guidelines are largely informed by randomised controlled trial (RCT) results [22], but we do not know if these findings apply to large patient populations not studied in trials. Fluticasone propionate + salmeterol (FP/SAL) is a LABA/ICS combination and is one of the most widely used COPD treatments. It was studied in large randomised trials (e.g. TORCH [7]), but we don't know the effects of treatment in important patient groups who were not studied. Some were excluded from trials (e.g. those >80 years and those with substantial comorbidity) and some are under represented (e.g. people with mild COPD) [7,11], meaning conclusions about these groups are difficult to make. The results we generate will aid patients, prescribers and policy makers in deciding the most appropriate treatment for COPD, based on stratified evidence, rather than assuming a one size fits all approach.

Evidence of medication effects in routine care is vital for understanding the balance of treatment risks and benefits. RCTs have unique strengths, but results from trials are not always a good guide to the effects of drugs in routine clinical practice [21]. National drug licensing authorities are now demanding better real world evidence on which to make decisions and have introduced legislation mandating studies of both effectiveness and risk to be conducted in routine clinical care rather than the narrow and optimal confines of most randomised trials [3,4]. Whilst the conduct of observational studies to investigate possible drug harms is well established, the use of these studies to estimate treatment effectiveness is in its infancy. Issues of treatment channelling and indication bias mean that measuring the intended benefit of a treatment is beset with difficulties. Whilst we and others have demonstrated it can be done in certain circumstances, we need more certainty about the methodology as it is applied in each disease area, since the issues of bias are likely to vary considerably [23]. Increasingly large quantities of electronic health records have become available to researchers recently and links between sources of data are constantly developing. Over the next few years we will see more observational studies of drug effectiveness emerging; however, rigorous, validated methodology is needed to translate these complex data into reliable evidence. For example, the availability of anonymised individual patient data from RCTs provides the potential for 'RCT-analogous' cohorts to be selected from non-interventional data sources (by matching patient records from non-interventional data to the RCT patient records on key characteristics). If subsequent analysis of a non-interventional RCT-analogous cohort generates results that are similar to those generated by the reference RCT, one could be confident in the validity of the results, and in the non-interventional methods used to obtain these results in this setting. Here we propose a template for drug effectiveness research with inbuilt validation against a randomised trial (Figure 1).

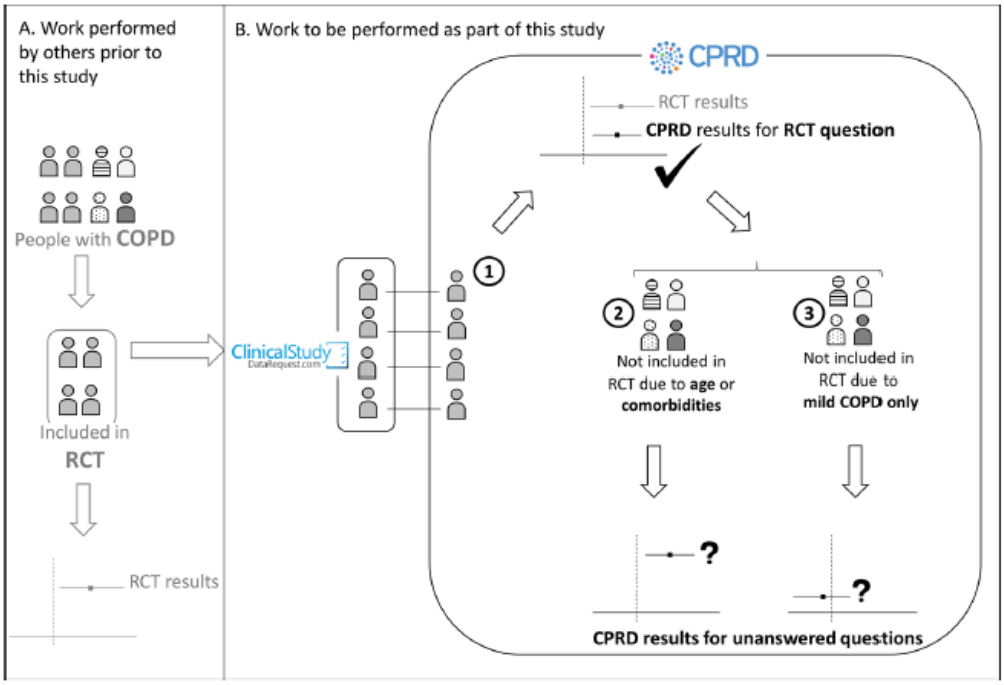

Figure 1 Overview of study objectives and sources of data for the chronic obstructive pulmonary disease (COPD) real-world medicines effects study (RCT, randomised controlled trial, CPRD, the UK Clinical Practice Research Datalink, FP/SAL, fluticasone propionate+salmeterol). (A) Work performed by others prior to this study. Of the total population of people with COPD, only a subset are included in RCTs of COPD treatments, based on the RCT inclusion/exclusion criteria. The RCT generates results that inform clinical practice, and the anonymised raw data for the study can be made available to other researchers via the Clinical Study Data Request website. For this study, the specific COPD treatment RCT of interest is the TORCH trial,[2] investigating the effect of FP/SAL on COPD exacerbations. (B) Work to be performed as part of this study. (1) Objective 1: a cohort of TORCH (RCT)-analogous patients will be selected from the CPRD, by matching people with COPD within CPRD to the records of people included in the trial. An analysis of the effect of FP/SAL on COPD exacerbations will then be performed on this TORCH-analagous CPRD cohort. If the results obtained are comparable to those obtained in the TORCH trial itself, this will serve as a validation step, showing that data from the non-inverventional ('real-world') CPRD source can reliably be used to study COPD treatment effects. (2) Objective 2: the validated analysis techniques used for objective 1 will then be used to study people in CPRD who would not have been eligible for inclusion in an RCT due to their age and the presence of other comorbidities, and for whom the effect of FP/SAL is currently unknown. (3) Objective 3: the validated analysis techniques will then be used to study people with mild COPD only, who have been under-represented in RCTs, and for whom the effect of COPD treatments is unclear.

Patient and Public Involvement

This research agenda has partially arisen following conversations between Dr Quint and people attending her COPD clinic. They have made clear they would like to be able to better understand the possible benefits and risks of COPD treatments for them, especially when they would not have been eligible for trials and the likely effects of treatments for them are unclear.

Several people in Dr Quint's clinic reviewed the plain English summary of the project for their views on 1) the project detail and 2) the wording of the summary. In particular, patients talked about the use of anonymised patient records and were encouraging of their use in this context. They confirmed the outcomes we will study are the most important outcomes for them with respect to COPD and they also made amendments to the summary to improve clarity.

Four people with COPD will be actively involved in the research through full participation with the Project Advisory Group and attendance of the group's 3 planned meetings (end of each objective). This will allow us to benefit from patient insight on the results and their interpretation. They will also work with us to determine effective and relevant ways to disseminate results more widely, e.g. through charities and Breathe Easy Groups. Formal and informal training will be provided as required, both by the study team, and through the London Farr Institute which has a well established PPI group and training materials.

Evidence explaining why this research is needed now

We did a systematic search to assess the effectiveness of FP/SAL in mild COPD compared with no treatment, for mortality, exacerbations or pneumonia over at least 12 months. Mild COPD was defined as COPD with FEV1 >60%, or MRC disease score 1 or 2, or CAT score <10, with no more than one prior exacerbation. A search of PubMed on 24 Mar 2016 had the following criteria ("Pulmonary Disease, Chronic Obstructive"[Mesh]) AND (("Fluticasone Propionate, Salmeterol Xinafoate Drug Combination"[Mesh]) OR ("Salmeterol Xinafoate"[Mesh] AND "Fluticasone"[Mesh])). 246 articles were found; following title review, 28 were retained for abstract review. Articles were rejected on the following grounds: no relevant outcomes (n=2), no untreated comparator (n=7), review article (n=6), severe COPD only or results in mild COPD not reported (n=2), <1 year follow up (n=6), TORCH papers (mild COPD results not reported; n=2), leaving 3 for review. Soriano et al [13] used GPRD to assess the effect of FP/Sal on mortality in people with COPD. They successfully measured an overall protective effect, but did not stratify by COPD severity. Nannini et al [14] was a systematic review of ICS/LABA compared with placebo for COPD, but results in mild COPD alone were not reported. One of the studies included enrolled people with mild COPD, but had short duration and no information on the outcomes of interest. One trial included people with FEV1 <70% predicted in the TRISTAN trial of FP/Sal for COPD, but did not stratify results by COPD severity and very few participants would have had mild COPD [15].

Ferguson [11] highlighted mild COPD as lacking in evidence of treatment effects. The lack of evidence of COPD treatment effects comes at an opportune time in terms of data availability and methodological advances. Our proposal exploits the recent availability of trial data to validate the use of a high quality large observational dataset in assessing the effects of COPD treatment in unstudied groups.

In summary, three factors make the proposed research both necessary and timely. 1) The lack of evidence of treatment effectiveness in large groups of people with COPD; 2) the availability of ever

increasing amounts of high quality electronic health records for research, and 3) the newly emerging need for robust methodology to generate real world evidence of drug effects.

Aims and objectives

Aim 1: To measure the association between treatments for COPD and COPD exacerbation rate, mortality, pneumonia, and time to treatment change amongst patients not included in randomised clinical trials for COPD treatments.

Aim 2: To develop a methodological framework with in built validation, for using observational electronic health records (EHR) to answer questions about drug risks and benefits.

Objectives

Objective 1. Validate methods for measuring COPD medication effectiveness in EHR data by comparing with trial results.

We will obtain individual patient data from the landmark TORCH trial which measured the effect of fluticasone propionate + salmeterol (FP/SAL) on COPD exacerbations, mortality and pneumonia amongst people with COPD. Individual trial participants will be matched with similar people in the anonymised EHR database the Clinical Practice Research Datalink (CPRD). Using propensity score cohort methodology, estimates of the effect of FP/SAL on exacerbation rate, mortality and pneumonia in the CPRD cohort will be measured. The results will be compared with the TORCH findings to determine the utility of CPRD records for measuring medication effectiveness in COPD. This objective will provide a methodological framework for measuring drug effectiveness in people with COPD, using observational data from CPRD. As a secondary analysis other treatments for COPD will also be compared with no treatment in this objective. Namely:
a) long-acting beta agonist (LABA)
b) long-acting muscarinic antagonist (LAMA)
c) LABA + LAMA
d) LABA + inhaled corticosteroid (ICS) other than FP/SAL
f) LABA + LAMA + ICS

Objective 2. Extension of trial findings: Measure effectiveness in patients excluded from trials

Using the methodological template developed in Objective 1, we will determine the effect of FP/SAL on exacerbation rate, mortality, pneumonia and time to treatment change in types of people not eligible for the TORCH study (most importantly people aged >80 years or with substantial comorbidity) and look separately at important subgroups e.g. those with and without underlying cardiovascular disease. As for Objective 1, the effects of other treatments will also be measured:
a) No treatment
b) LABA
c) LAMA
d) LABA + LAMA
e) LABA + ICS
f) LABA + LAMA + ICS

Objective 3. Mild COPD: Determine treatment effectiveness in an under-studied disease stage

Whilst some patients with mild COPD could have been eligible for inclusion in TORCH, few were actually enrolled and options for disease management in addition to smoking cessation, flu and pneumococcal vaccination and potentially pulmonary rehabilitation in mild COPD are not clear. We will select a cohort of people in CPRD with mild COPD. Using the methodological template developed in Objective 1, we will compare exacerbation rate, mortality, pneumonia and time to treatment change based on prescribed treatments:

a) No treatment
b) LABA
c) LAMA
d) LABA + LAMA
e) LABA + ICS
f) LABA + LAMA + ICS

We will also perform a sensitivity analysis where we allow the group of people with FEV1 >60% predicted who had a maximum of one exacerbation within 1 year post-COPD diagnosis to be included.

Research Plan

Data Sources

This project will combine data from a randomised clinical trial (TORCH) and prospectively collected routine electronic healthcare data from the Clinical Practice Research Datalink.

TORCH

TORCH was a placebo controlled randomised trial of the combined inhaler fluticasone propionate (FP) + salmeterol (SAL) for the treatment of COPD. Patients were randomised to receive FP+SAL, FP alone, SAL alone or placebo [7] and the primary comparison of interest was between FP+SAL and placebo. Key outcomes were expected benefits (rate of COPD exacerbation and mortality) and an expected harm due to the immunosuppressive action of the corticosteroid FP (pneumonia). Whilst findings for the primary endpoint of mortality were null, this was thought to be due to poor statistical power as a result of a lower than anticipated mortality rate. Nonetheless, a lower rate of exacerbations was seen with FP/SAL, and a higher rate of pneumonia was observed. As one of the largest trials in COPD, and with three year follow up, TORCH is a landmark study, providing a validation point for our study. We will obtain individual patient data from the TORCH study via www.clinicalstudydatarequest.com for use in Objective 1 (see below)

CPRD

The CPRD is an exceptional source of anonymised UK population-based electronic health records. The primary care records comprise ~8-10% of the UK population and contain comprehensive information on clinical diagnoses, prescribing, referrals, tests and demographic/lifestyle factors [5]. Data quality/validity are high and the data are nationally representative [5, 6]. A patient starts contributing follow-up time to the database at the date they join an 'up-to-standard' practice (or the date that their practice starts contributing up-to-standard data), and stop contributing follow-up time on either their death date, their transfer out date (the date that they leave the database due to reasons other than death) or on the last collection date for their practice. Linkage between the primary care records in CPRD and hospital episode statistics (HES) is well established for >60% of practices in the CPRD, providing a data set augmented with detailed secondary care diagnostic and procedural records. Our group has many years experience in the use of these data to conduct high quality studies, and we have established validated algorithms to identify COPD, COPD exacerbations and pneumonia (both hospital and primary care managed) in CPRD/HES linked data [6,8,9].

People with COPD, registered in CPRD are eligible for this study. We will not make changes to treatment people receive as part of this purely observational study; we will compare outcomes between people receiving treatments for COPD, and those receiving no treatment.

The outcomes we will study are COPD exacerbation, mortality, pneumonia and change in COPD treatment as they are key endpoints of significance for people with COPD and treating clinicians [22]. All are readily identifiable and validated in CPRD/HES data. We will measure absolute rates of each outcome and relative rates comparing COPD treatment with no treatment.

Analysis Plan:

Objective 1: Validate methods for measuring COPD medication effectiveness in EHR data by comparing with trial results.

Individual patient data for TORCH will be obtained via Clinicalstudydatarequest.com. Patients registered in the CPRD who also meet TORCH inclusion/exclusion criteria [7] will be selected from between the dates of 1st January 2004 and 1st January 2017. The following TORCH Inclusion criteria will be applied first: a diagnosis of COPD, age 40-80 years, a history of smoking, FEV_1<60% predicted and FEV_1/FVC<70%. An eligible-for-inclusion date will then be assigned as the date that all of the above inclusion criteria were met for the indivdual, before excluding those with any of the following exclusion criteria prior to this date: a diagnosis of asthma (within the previous 5 years), a diagnosis for any (non-COPD) respiratory disorder, a record of lung surgery, a diagnosis of alpha-1 antitrypsin deficiency, evidence of of drug or alcohol abuse, a record of having received long-term oxygen therapy, diagnoses for conditions likely to interfere with the TORCH trial or cause death within 3 years, current use of oral corticosteroid therapy (defined as continuous use for >6 weeks, with courses of oral corticosteroids separated by a period of <7 days considered as continuous use), any any exposure to any of the TORCH study drugs within the previous 4 weeks. Finally, in-line with the TORCH trial approach, anyone who has an exacerbation requiring oral corticosteroid therapy or hospitalisation during the run-in period (the 2-week period following eligibility) will also be excluded. Given the limited information on how asthma exclusions were applied in the TORCH study, we will perform a sensitivity analysis in which the asthma exclusion is a diagnosis within the previous 1 year, rather than 5 years as specified above.Feasibility counts in the CPRD indicate ~54,000 CPRD patients meet these inclusion/exclusion criteria. Initially, each TORCH participant (n=6,112) will be individually matched with up to 2 CPRD patients who were not treated for COPD at the time they attained TORCH eligibility. Feasibility work shows ~20,000 CPRD patients meet these criteria. This will create a CPRD population of people with untreated COPD analogous to the TORCH population at baseline (CPRD TORCH analogous untreated group, n~12,000). Next, all CPRD patients who were TORCH eligible AND who received FP/SAL will be selected (n>20,000; CPRD TORCH eligible treated group). Logistic regression modelling will then be used to calculate a propensity score (PS) for FP/SAL treatment among the combined CPRD TORCH analogous untreated and CPRD TORCH eligible treated groups (n>30,000). A wide range of clinical/demographic factors will be used in the model based on established practice for optimising propensity score modelling [24, 25]. The propensity score will be constructed using the principle that predictors of the exposure and outcome, or outcome only (mortality) should be included. We will consider a wide range of factors for inclusion, such as: age, sex, body mass index, alcohol consumption, and a wide range of comorbidities (e.g. type 2 diabetes, coronary heart disease, cerebrovascular disease, peripheral vascular disease, heart failure, hypertension, renal disease, cancer). We will further adjust for healthcare utilization intensity (number of prior visits, hospitalizations, number of distinct medications used, number of

procedures, etc.) as these are generic correlates of disease state and the likelihood of recording completeness. We have substantial prior experience of building propensity models, and Dr Williamson (Co-I) is leading an MRC funded project to determine optimal propensity score methods for use with missing data.

Every untreated patient will then be PS matched 1:1 with the FP/SAL treated patient with the closest propensity score, resulting in a TORCH analogous cohort of FP/SAL treated and untreated CPRD patients (n~24,000).

Comparisons will be made according to FP/Sal status for rate of COPD exacerbation, pneumonia and mortality over 3 years. All analyses will be performed according to the 'intention-to-treat' principle (as was done in the TORCH study), meaning that if a participant enters the study as either an exposed or unexposed participant, they will remain assigned to that exposure category for the entire duration
of their follow-up (irrespective as to whether their true exposure status changes). For exacerbations, a negative binomial model will be used, accounting for variability between patients in the number and frequency of exacerbations, with the number of exacerbations as the outcome and the log of treated time as an offset variable. Time to mortality and first pneumonia will be analysed using Cox proportional hazards regression. This mirrors TORCH endpoints of major benefit and harm. We anticipate the propensity matching process will allow us to assemble treated and untreated groups that are very similar with respect to baseline characteristics except FP/SAL treatment status. However, this will be tested by assessing standardised differences for each baseline variable. If substantial differences are noted for important variables, it may be necessary to further adjust the statistical models [26]. This could also include examining the effect of using greedy versus optimum matching approaches in order to obtain the closest propensity score match and/or matching at a ratio other than 1:1.

The research team has substantial experience in defining disease phenotypes using data from the CPRD and will largely use well established algorithms and code lists for this purpose. Specifically for COPD exacerbations we will be applying algorithms developed by Dr Quint who has recently published work validating the recording of acute exacerbations using CPRD data and found a combination of appropriate diagnostic and therapy codes gives a positive predictive value of 86% [28]. Furthermore, Dr Quint's team have now extended this work to include HES data showing improvements in ascertainment when linked primary and secondary care data are used.

Specifically for cardiovascular disease we will ascertain individual patient status for coronary heart disease, cerebrovascular disease, peripheral vascular disease, heart failure and hypertension, again using well established algorithms developed by our group, making use of direct primary care expertise.

We will validate our findings against TORCH by determining whether results of the CPRD analysis are compatible with the TORCH rate ratio for exacerbations (0.75; 95% CI 0.69-0.81). This outcome has been selected as it is an outcome of key significance for people with COPD [22] and the result in TORCH shows a clear benefit with 95% confidence intervals below 1. We have set two criteria that must be met for us to conclude results are consistent. First the effect size must be clinically comparable with TORCH findings; the rate ratio for exacerbations in CPRD must be between 0.65 and 0.9. This range is deliberately not symmetrical around the TORCH estimate of 0.75 as we would anticipate the treatment effect in routine clinical care may be weaker than that seen in the optimised setting of a randomised trial. We recognise this rule could be met with a poorly powered,

inconclusive result, so a second criterion is that the 95% confidence interval for the rate ratio must exclude 1. If the results for our FP/SAL versus no treatment comparison are not consistent with the TORCH FP/SAL versus placebo results, we will perform additional analyses where instead of using a no-treatment comparator group, our objective 1 comparator group will be people exposed to SAL, one of the other comparator groups from the TORCH trial. If we go on to compare FP/SAL with SAL alone (see the 'Exposures, outcomes and covariates' section, 'Exposures' subheading), the 95% CI would also need to exclude 1, and the rate ratio would need to be between 0.81 and 0.95 (compared with the TORCH FP/SAL vs SAL result of 0.88, 95% CI 0.81 to 0.95).

Adherence to issued prescribing in general practice is likely to vary according to the treatment issued e.g. short course antibiotic treatment is notoriously not well adhered to, whereas long term life-saving treatment such as antiretroviral medication is more likely to be taken as prescribed. Whilst we do not have figures for adherence for COPD medication in UK general practice, we are able to estimate the proportion of time covered by prescribing as a proxy for adherence and will account for this in our analyses. Moreover, our intention is to estimate the effect of prescribing at the population level, and to some extent, the clinical effects we will measure are in part due to pharmacological effects, and in part the way the treatment is taken which includes adherence. Also of note, prescribing for COPD in the UK is predominantly through GPs and so we will not be missing prescribing information from other potential sources of treatment.

The data analysis for adherence will necessarily be a significant element of the work to be done for this study. However, we have reviewed the records for a random sample of 30 people with COPD starting treatment with FP/SAL to look at adherence patterns over the course of a year. Of the 30 patients, 20 (67%) were still receiving FP/SAL one year after starting treatment. Of the 20 who received FP/SAL for a full year, 15 (75%) received sufficient prescriptions to suggest at least 50% adherence over the year, and 8 (40%) had sufficient prescriptions to suggest 80% adherence or higher. As expected, this suggests two things: Firstly adherence is likely to be poorer in routine clinical care than in the trial population; in TORCH 80% of participants were estimated to have adherence at 80% or higher. Secondly there is a wide range of adherence in routine care. This will allow us to estimate both the population level effects of treatment as actually used in routine care, but also to estimate the treatment effect in patients with more similar levels of adherence to trial participants. Whilst we acknowledge that prescribing can only be a proxy for used medication, we believe it is not an unreasonable assumption that the amount of medication prescribed is correlated with the amount consumed. We plan to assess adherence for the cohort that we select for objective 1 beyond 1 year and report the findings.

In the event that Objective 1 detects a null or poorer treatment effect than anticipated (rate ratio > 0.9), we will conduct a sensitivity analysis restricted to people estimated to be covered by FP/SAL treatment for 80% of their follow up.

For the study phase replicating TORCH findings, we will limit included patients to those receiving FP/SAL 500/50, the dose used in TORCH. This information is recorded for all prescriptions of FP/SAL and this dose is the only currently approved dose for COPD in the UK (though we recognise some prescribing may not follow the licensed indication).

For further study phases we will again utilise recorded prescribing information to determine the dose received. We will be reliant mostly on the strength of each individual drug which is recorded automatically against each product and does not require GPs to enter this data, ensuring completeness. We will then be able to stratify analyses based on the dose prescribed

As a secondary analysis, the effects of other COPD treatments on the same outcomes will also be determined.

Objective 2: Extension of trial findings: Measure effectiveness in patients excluded from trials.

If the results of Objective 1 confirm the utility of CPRD data for this research we will assemble a cohort with COPD not meeting TORCH entry criteria due to age (>80 years), additional lung disease, and substantial comorbidity. TORCH required people to be excluded from the study if they had serious uncontrolled disease with a likelihood of causing death within 3 years. It is possible this criterion affected participant selection and led to the lower overall rate of death than originally anticipated, although we recognise that this criterion is subjective. During Objective 1 we will be able to select groups of people in the CPRD who although eligible for inclusion in TORCH tended not to be included, most likely because of this subjective exclusion criterion. We anticipate this will be people with substantial comorbidity e.g. serious vascular disease, as well as people who had an asthma diagnosis at any time prior to inclusion Status for such diseases is readily identified in both the CPRD data and in the TORCH baseline data.

Feasibility counts indicate >40,000 patients meet these criteria based on age alone, with 25% exposed to FP/SAL. FP/SAL treatment status will be determined for all eligible patients for this objective and a PS will be calculated based on logistic modelling using the variables determined in Objective 1. FP/SAL treated and untreated people will be matched 1:1 on PS and their risk of COPD exacerbation, mortality, pneumonia and time to treatment switch will be compared using the methodology outlined in Objective 1. We will define switching as stopping receiving prescriptions for the active substance of interest and starting treatment with another active substance, based on prescribing patterns**.** Results will be stratified by key TORCH exclusions (e.g. age, and comorbidity) and by underlying vascular disease status. As a secondary analysis, the association between other treatments for COPD will also be measured for these outcomes; a) long-acting beta agonist (LABA), b) long-acting muscarinic antagonist (LAMA), c) LABA + LAMA, d) non FP/SAL LABA + inhaled corticosteroid (ICS), e) LABA + LAMA + ICS. Participants for each of the objective 2 cohorts will be selected in a similar fashion to the objective 1 cohort, with the amended eligibility criteria specified above applied (ie, step 1 will be modified for selection of each of the objective 2 cohorts). As for objective 1, each participant
will be allowed to have multiple FP/SAL exposed and unexposed eligibility periods in their record, as described in figure 3. In contrast to objective 1, there will be no matching of unexposed patients to TORCH patients, as we do not require a TORCH-analagous cohort for this analysis (ie, no step 3). All other selection steps will be as applied for objective 1, including the use of propensity score matching in order to obtain comparable unexposed and exposed groups for analysis.

Objective 3: Mild COPD - Determine treatment effectiveness in an under-studied disease stage.

We will select people with mild COPD (a COPD diagnosis, >60% predicted FEV1 (or >50% plus MRC breathlessness scale 1 or 2, or >50% plus COPD Assessment Test (CAT) score <10), and a maximum of 1 exacerbation in the year post COPD diagnosis. We will compare outcomes for those subsequently receiving COPD medications and those receiving none. As recognised by NICE and others, people with mild COPD are routinely not included in many randomised trials [11,22] and as a result, there is currently little evidence of treatment effects in this group, particularly on the use of triple therapy (LABA + LAMA + ICS). We are not advocating that triple therapy should be the treatment of choice in this group, but clinical practice shows it is prescribed for this group, and yet we do not know how it affects patient outcomes.

Feasibility work in the CPRD shows that the parameters needed to identify mild COPD are recorded for ~65-70% of COPD patients since 2004 due to QoF requirements, meaning we will readily be able to include a large patient group for this objective. Preliminary work indicates >40,000 patients in CPRD will meet the entry criteria for this objective and ~30% were exposed to FP/SAL. Outcomes will be compared according to treatment status, comparing the following treatment options against no treatment: a) LABA, b) LAMA, c) LABA + LAMA, d) LABA + ICS, e) LABA + LAMA + ICS. Methodology will be as described for Objective 2, using PS matching to assemble comparable groups. As for objective 2, the selection steps will be similar to objective 1, with modified criteria for step 1 and the removal of the TORCH-matching step (step 3).

To deal with multiple tests we will use 95% confidence intervals for the primary analysis comparing COPD exacerbation in the TORCH analogous CPRD population and 99% confidence intervals for all other analyses.

Health technologies being assessed:

This project will assess pharmaceutical treatments for COPD. The specific types of treatments to be assessed are:

a) long-acting beta agonist (LABA), b) long-acting muscarinic antagonist (LAMA), c) LABA + LAMA, d) non FP/SAL LABA + inhaled corticosteroid (ICS), e) LABA + LAMA + ICS

There is no plan to conduct economic evaluation in this proposal. Depending on the findings, a further proposal may be forthcoming which would include economic evaluation.

Design and theoretical/conceptual framework

All three objectives will be answered using cohort study designs, comparing the outcomes of people treated for COPD with the outcomes of those not treated. Propensity score matching will be used as detailed in the Research Plan, enabling us to assemble groups of treated/untreated people who are otherwise similar, so we can estimate the effect of the treatment itself on the outcomes of interest. Since the research is observational rather than randomised, the possibility that the associations we measure are not caused by the treatment remains a concern. Other differences between groups of people being compared could also explain any differences seen (e.g. unmeasured differences in COPD severity). A key strength of our proposal is the validation step in Objective 1 where we measure the association between FP/SAL and COPD exacerbation rate and directly compare our results against those found in the TORCH randomised trial. There is also considerable experience of treating COPD in the team, both in primary and secondary care, which will facilitate our understanding of any unanticipated findings.

Following Objective 1 we will use the same methodological framework to measure the absolute and relative effects of COPD treatment in people not included in randomised trials. Key outcomes of major benefit and risk have been selected (COPD exacerbation, mortality, time to treatment change and pneumonia). We do not know whether the relative and absolute effects of COPD treatment are the same in these patients, as treatment effects often vary considerably in routine clinical use compared with the randomised trial setting [21].

Target population:

The target population is all people with COPD in the United Kingdom. The project focuses specifically on the types of people who are not usually enrolled in randomised trials, but who form a large proportion of the population with COPD. This is important as we aim to deliver evidence about the effects of treatments in this generally under studied, but large population. The data we will use are nationally representative meaning the results we obtain will generalise fully to the wider population with COPD.

Inclusion/Exclusion Criteria

We intend to include patients registered in the CPRD for analyses as outlined below. All criteria are readily identified in the CPRD based on routine recording of demographics, disease history and clinical outcomes. In particular, lung function is well recorded for most patients with COPD due to expectations under the QoF.

For Objective 1, we will include all patients who would have been eligible for the TORCH study based on the following key criteria:

- a diagnosis of COPD,
- age 40-80 years,
- lung function (FEV1<60% predicted, FEV1/FVC ratio <70%),
- smoking history,
- no history of asthma,
- no history of lung surgery,
- no requirement for long-term oxygen therapy,
- no diagnosed alpha-1 antitrypsin deficiency,
- no evidence of drug/alcohol abuse.

Some of the TORCH inclusion criteria will not be fully assessable using CPRD data (e.g. we will be able to assess whether patients are smokers but will not always know their pack year history). Hence the inclusion/exclusion criteria are analogous with TORCH criteria but we acknowledge they are not identical. Identification of criteria will be done based on algorithms already determined [8] and by the identification of clinical codes in the CPRD.

From this TORCH eligible group, patients not treated with FP/SAL will then be matched 2:1 with TORCH participants on baseline characteristics, providing a pool of ~12,000 untreated patients. FP/SAL treated patients in the TORCH eligible group will be selected and combined with the ~12,000 untreated group. Propensity scores will be calculated for FP/SAL treatment amongst this combined group of >30,000 patients. Each FP/SAL untreated patient will then be matched with the treated patient with the closest propensity score, providing an analysis population of ~24,000.

Recording patterns for diseases can change over calendar time and for this reason we will match exposed and unexposed groups on calendar time i.e. the start date for an unexposed person will be the same as for an exposed person, necessitating presence in the CPRD at the same time.

To closely mirror the process followed for TORCH, we will include new users of each treatment of interest but possible prior users of other treatments. In TORCH, patients had to not be long term oral corticosteroid users but could be on other treatments.

For each study question we will be looking at new users of each treatment of interest with no prior record of receiving that drug. Prior treatment with other medications before cohort entry is recorded in CPRD and will be adjusted in the propensity score model.

Substantial preparatory work has been conducted by Dr Quint's group to explore the completeness and utility of respiratory parameters recorded in CPRD for people with COPD. Over 80% of people with COPD in the CPRD have a valid and usable record for FEV1 in a 15 month time window, suggesting we will have up to date measures on the majority of potential participants. Similar levels of recording have also been found for MRC scores.

On the wider issue of missing data of relevance for this study, CPRD data are shown to be almost complete for drug prescribing and mortality (partly through ONS linkage). Smoking history tends to be very well recorded for people with COPD and missingness is likely to be minimal [28]. Information on important comorbidity is also well recorded in CPRD. We will conduct both complete case analyses and use multiple imputation where appropriate assumptions hold, applying findings from methodological work led by Dr Williamson into the use of multiple imputation in propensity score modelling.

Long term oxygen therapy is arranged through GPs and preliminary work has allowed us to identify product codes for oxygen within the CPRD product dictionary, and which we will use to identify people on oxygen therapy. We have previously used recorded alpha-1 antitrypsin (A1AT) in patient clinical files (using a specific Readcode) to identify this condition. A1AT is likely to be recorded where known but we accept that it will not be measured in all potential participants. We also acknowledge that alcohol and drug abuse will not be perfectly ascertained in CPRD; whilst we will use Readcodes and prescribing (e.g. disulfiram) that suggests alcohol and drug abuse, we will not correctly identify abuse in people for whom no records exist. These limitations are in line with our aim to select a population as similar to the TORCH population as possible, whilst accepting that some differences will remain.

For Objective 2, we will include patients with a valid COPD diagnosis in the CPRD [8] meeting these additional criteria

- age >80 years,
  OR
- history of lung surgery
  OR
- History of long term oxygen therapy
  OR
- Evidence of drug/alcohol abuse
  OR
- Substantial comorbidity: TORCH required people to be excluded from the study if they had serious uncontrolled disease with a likelihood of causing death within 3 years. It is likely this criterion affected participant selection and led to a lower overall rate of death than originally anticipated, although we recognise this criterion is subjective. During Objective 1 we will be able to find groups of people who were generally not included despite being eligible, most likely because of this subjective exclusion criterion. We anticipate this will be people with substantial comorbidity e.g. serious vascular disease. Status for such diseases is readily

identified in both the CPRD data and in the TORCH baseline data. We will only be able to specify this criterion in detail after we have completed Objective 1.

For Objective 3, we will include patients with a valid COPD diagnosis in the CPRD [8] meeting these additional criteria

- >60% predicted FEV1 (or >50% plus MRC breathlessness scale 1 or 2, or >50% plus COPD Assessment Test (CAT) score <10)
  AND
- a maximum of 1 exacerbation in the year post COPD diagnosis.
- no evidence of asthma

Setting/context:
This project will be conducted using primary care and linked secondary care data from the Clinical Practice Research Datalink and Hospital Episode Statistics. Data are derived from the routinely collected anonymised electronic health records of patients from >600 UK general practices.

Sampling:
We will include all eligible patients registered in the CPRD and who meet the criteria described in the Research plan. Assuming a baseline conservative exacerbation rate of 0.5 per patient per year [28], we would only require a sample of 408 patients per treatment group to detect a reduction in annual exacerbation rate to 0.4 per year, with 80% power and 5% significance. The estimated sample size is >20,000 for each objective which will provide ample power for the main outcomes of interest, but also allow stratification by patient characteristics to determine stratified results, and will also be ample for the secondary analyses where we will use 99% confidence intervals. For example, to detect a reduction from 0.5 to 0.4 exacerbations per year with 80% power and 1% significance we would need ~600 people in each treatment group. There is no additional data access cost associated with including such a large population in the study.

Misclassification of drug exposure periods and outcome status:
It is possible that an individual may still be exposed to FP/SAL for some time after a prescription has finished, for example, if they have medication at home that they have not used from a previous prescription. This would mean that people may become eligible for inclusion in the unexposed group while they are actually still exposed. If our result differs from the TORCH results (eg, a rate ratio <0.65 or >0.9), we will conduct a sensitivity analysis in which we include an additional (grace) exposed
period equivalent to the length of a single prescription at the end of each actual exposed period, and only classify individuals as eligible for inclusion as unexposed at the end of this additional period. Our results could also be impacted by misclassification of outcome, given the routine nature of the data. Our initial approach for detection of COPD exacerbations is to use a validated case definition from previous work that maximises positive predictive value while maintaining a relatively high sensitivity.12 If our result differs from the TORCH results, we will consider performing a sensitivity
analysis in which we assess the impact of applying alternative case definitions for COPD exacerbations (see online supplementary material for an overview of articles relating to the case definitions we plan to use, including any validity measurements provided).

Data collection:
Data will be obtained from:

- Clinicalstudydatarequest.com – individual patient data for TORCH participants will be obtained. We have already begun the process of applying for these data which are listed as freely available to researchers.

- CPRD linked with HES data. Our research team has considerable experience in obtaining data from the CPRD for research of this kind, and does not anticipate any difficulties in obtaining the required permission. The key exposures, outcomes and covariates required for this study are well characterised in these data, and some have been validated by our research group [5,6,8,9].

Data analysis
Please see Research Plan above

Dissemination and projected outputs
Dissemination of findings will be via a combination of channels. The work will be published in high ranking peer reviewed journals and we anticipate 3 publications to arise directly from the planned work. Findings will also be presented at relevant scientific conferences: the British Thoracic Society Conference (year 2) and the European Respiratory Society International Congress (year 3).
The dissemination phase is where public and patient engagement is likely to have the biggest impact on this programme of work. We will engage with patients already identified by Dr Quint from her clinic and Breathe Easy Groups and with relevant charities such as the British Lung Foundation to determine the most relevant ways to disseminate results directly to patients in an accessible manner, and to help our understanding of the likely impact of results to specific groups of patients.

The project will be featured on the LSHTM Electronic Health Records research group web page as ongoing research and presented in a public/patient friendly way. This page will be updated with progress as the programme develops.

We will communicate directly with NICE to ensure they are kept informed of results that are of direct relevance to the guidance they have issued on COPD, and with the Medicines and Healthcare Products Regulatory Agency if it appears that findings may impact the risk/benefit profile of COPD treatments.

Patients, healthcare professionals and policy makers will benefit immediately from the research. It will generate new evidence about the risks and benefits of treatments for COPD in the general population, and improve clarity about disease stage and optimum treatment options. This evidence will feed into informed decision making for patients and doctors trying to decide on the likely risks and benefits of COPD treatments for individual patients.

The findings of this work will feed directly into the teaching programmes at LSHTM on electronic health records and pharmacoepidemiology

We anticipate 3 publications to arise from this project, and we have a track record of producing high impact observational research on the use of medicines and in respiratory medicine [8,9,16,17]. Papers will be based on 1) the validation study comparing CPRD data with the TORCH study findings, 2) the effects of FP/Sal for the treatment of people excluded from TORCH, 3) the effects of COPD medications in people with mild COPD. Our PPI programme ensures we will engage with patient groups and charities to make key results and conclusions available in patient/public friendly format. It is possible our results will impact on the risk/benefit profile of medicines used for COPD and we will therefore engage with both the MHRA and NICE, dependent on the findings. We will also engage with media outlets to explain the findings and how they impact on the health of people with COPD.

By quantifying outcomes associated with COPD treatment, and stratifying by important patient characteristics we will provide information needed by patients, healthcare providers and policy

makers that will aid decision making around treatment choices. If groups of patients are found to have better chances of achieving positive outcomes, this could influence personal decisions about the suitability of treatment, especially amongst patients with so far understudied mild COPD.

Plan of investigation and timetable

Please see flow diagram for further detail of the timetable. The project will begin with applying for and obtaining individual patient data from the TORCH study. This process has already begun and we anticipate will take ~6-9 months for data delivery. In parallel with this, we will apply for approval to use CPRD data for the project, through the MHRA Independent Scientific Advisory Committee (ISAC), which we anticipate will take 2 months to complete. CPRD data will then be prepared for all patients who would have been eligible for TORCH. When both CPRD and TORCH data are available, individual patient matching will take place and CPRD data will be analysed for Objective 1 (months 12-18). The Project Advisory Group will meet at this stage to determine whether the results meet the criteria to continue with the project before moving on to Objective 2 between months 18-24. Objective 3 will be completed in months 24-30 and the remaining time will be used for dissemination, including preparation and submission of manuscripts to peer reviewed journals. The Project Advisory Group including patient representatives will be consulted when results are available for each objective ~months 18, 24 and 30. Dissemination of results via conferences to obtain valuable input from the wider research community will be in years 2 (British Thoracic Society) and 3 (European Respiratory Society).

Project management

The project will be managed by the Principal Investigator who is very experienced in conducting and managing projects of this type. Dedicated meetings will be held with the post -holder, initially weekly to ensure project momentum, and then bi-weekly when the project is running smoothly. The project will be conducted entirely at LSHTM, with project team members elsewhere kept fully informed with monthly updates, and consulted at key stages for their input as required. For example we anticipate working closely with statistical experts for Objective 1 to ensure optimal use of the observational data to replicate the TORCH analysis in CPRD. This will involve Skype/teleconference meetings with Prof Schneeweiss in Harvard, and face to face meetings with the UK based collaborators (all in London).

The Project Advisory Group will be comprised of the main investigators (ID, LS, JQ, EW, the post -holder to be appointed), and the COPD patient representatives and will be independently chaired by Prof Ian Wong from University College London. Prof Wong has extensive experience in the use of electronic health records for research into the effects of drugs and has kindly agreed to this role. The group will meet after each objective to discuss the results, implications, and dissemination.

Approval by ethics committees

Ethical approval will be sought from the London School of Hygiene & Tropical Medicine Ethics Committee, if the project is funded. Based on previous work using anonymised electronic health record datasets we do not anticipate barriers to approval and approval is normally granted within days/weeks.

Scientific approval will also be obtained from the Independent Scientific Advisory Committee (ISAC) of the Medicines and Healthcare Products Regulatory Agency, for use of the CPRD data. To do this, we will need to submit a structured protocol, based largely on the application we present here, and so conversion to a full protocol will be readily achieved. Based on experience a protocol is likely to be quickly approved (within weeks of application) and we do not anticipate any barriers. CPRD data are already approved via a National Research Ethics Committee for purely observational research of this type.

Expertise

Ian Douglas has extensive experience using electronic health records (EHRs) to determine the effects of drug treatments. He is a member of the MHRA Pharmacovigilance Expert Advisory Group, giving advice on all aspects of drug risk/benefit evaluation and methodology and directs the LSHTM programme of teaching in pharmacoepidemiology. Ian will lead all aspects of the proposed research; design, conduct, reporting and dissemination.

Elizabeth Williamson is a leading statistician in propensity score techniques for EHRs, particularly with missing data, and will provide advice on design/interpretation of all analyses.

James Carpenter is a statistician with an international reputation based both at LSHTM and the MRC Clinical Trials Unit. His dual expertise in both clinical trial analysis and approaches to missing data will be invaluable for this project and he will provide advice on design/interpretation of all analyses.

Jennifer Quint is an excellent respiratory clinical epidemiologist undertaking hands on clinical work as a consultant and leading a team of observational researchers. She will provide input to all aspects of study design, and the interpretation and dissemination of results. Dr Quint's clinical work has also identified the patients to be involved in PPI activities and she will be fully engaged with the patient representatives along with Ian Douglas.

Sebastian Schneeweiss is an international leader in observational epidemiology to investigate drug effects and has published extensively on the use of propensity scores and also the utility of randomised trials for informing observational studies. He will provide advice on design/interpretation of all analyses.

Liam Smeeth is a clinical epidemiologist with an outstanding record in the use of EHRs, and also a practising GP. He is also a Wellcome Senior Fellow and deputy director of the London Farr Institute. He will provide expertise and advice on all aspects of study design, interpretation and dissemination.

References

1. Healthcare Commission (2006) Clearing the air: a national study of chronic obstructive pulmonary disease. London: Healthcare Commission.
2. Rodriguez-Roisin R. Toward a consensus definition of COPD exacerbations. Chest2000;117:398S-401S.
3. REGULATION (EU) No 1235/2010 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
4. Guidance for Industry Postmarketing Studies and Clinical Trials — Implementation of Section 505(o)(3) of the Federal Food, Drug, and Cosmetic Act, US FDA, April 2011
5. Herrett, E., Gallagher, A. M., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., & Smeeth, L. (2015). Data Resource Profile: Clinical Practice Research Datalink (CPRD). International Journal of Epidemiology, 44(3), 827–36. doi:10.1093/ije/dyv098
6. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol. 2010 Jan;69(1):4-14. doi: 10.1111/j.1365-2125.2009.03537.x.

7. Calverley, P. M., Anderson, J. A., Celli, B. R., Ferguson, G. T., Jenkins, C., Jones, P. W., … Vestbo, J. (2007). Salmeterol and Fluticasone Propionate and Survival in Chronic Obstructive Pulmonary Disease. New England Journal of Medicine, 356(8), 775–789.
8. Quint, J., Müllerova, H., DiSantostefano, R., Forbes, H., Eaton, S., Hurst, J., … Smeeth, L. (2014, August 7). Validation of chronic obstructive pulmonary disease recording in the Clinical Practice Research Datalink (CPRD-GOLD). BMJ Open. BMJ Publishing Group. doi:10.1136/bmjopen-2014-005540
9. Millett ER, Quint JK, Smeeth L, Daniel RM, Thomas SL Incidence of community-acquired lower respiratory tract infections and pneumonia among older adults in the United Kingdom: a population-based study. PLoS One. 2013 Sep 11;8(9):e75131. doi: 10.1371/journal.pone.0075131. eCollection 2013.
10. https://www.clinicalstudydatarequest.com/
11. Ferguson, G. T. (2011). Maintenance pharmacotherapy of mild and moderate COPD: what is the evidence? Respiratory Medicine, 105(9), 1268–74. doi:10.1016/j.rmed.2011.02.005
12. Freemantle, N., Marston, L., Walters, K., Wood, J., Reynolds, M. R., & Petersen, I. (2013). Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. BMJ (Clinical Research Ed.), 347, f6409. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24217206
13. Soriano, J. B., Vestbo, J., Pride, N. B., Kiri, V., Maden, C., & Maier, W. C. (2002). Survival in COPD patients after regular use of fluticasone propionate and salmeterol in general practice. The European Respiratory Journal, 20(4), 819–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12412670
14. Nannini, L. J., Poole, P., Milan, S. J., Holmes, R., & Normansell, R. (2013). Combined corticosteroid and long-acting beta$_2$-agonist in one inhaler versus placebo for chronic obstructive pulmonary disease. The Cochrane Database of Systematic Reviews, 11, CD003794. doi:10.1002/14651858.CD003794.pub4
15. Calverley, P., Pauwels, R., Vestbo, J., Jones, P., Pride, N., Gulsvik, A., … Maden, C. (2003). Combined salmeterol and fluticasone in the treatment of chronic obstructive pulmonary disease: a randomised controlled trial. Lancet (London, England), 361(9356), 449–56. doi:10.1016/S0140-6736(03)12459-2
16. Rothnie, K. J., Smeeth, L., Herrett, E., Pearce, N., Hemingway, H., Wedzicha, J., … Quint, J. K. (2015). Closing the mortality gap after a myocardial infarction in people with and without chronic obstructive pulmonary disease. Heart (British Cardiac Society), 101(14), 1103–10. doi:10.1136/heartjnl-2014-307251
17. Quint, J. K., Herrett, E., Bhaskaran, K., Timmis, A., Hemingway, H., Wedzicha, J. A., & Smeeth, L. (2013). Effect of β blockers on mortality after myocardial infarction in adults with COPD: population based cohort study of UK electronic healthcare records. BMJ (Clinical Research Ed.), 347, f6650. doi:10.1136/bmj.f6650
18. Seemungal, T. A. R., Hurst, J. R., & Wedzicha, J. A. (2009). Exacerbation rate, health status and mortality in COPD--a review of potential interventions. International Journal of Chronic Obstructive Pulmonary Disease, 4, 203–23. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2699821&tool=pmcentrez&rendertype=abstract
19. Piaggio, G., Elbourne, D. R., Pocock, S. J., Evans, S. J. W., & Altman, D. G. (2012). Reporting of noninferiority and equivalence randomized trials: extension

of the CONSORT 2010 statement. JAMA, 308(24), 2594–604. doi:10.1001/jama.2012.87802.

20. Vestbo, J., Anderson, J. A., Calverley, P. M. A., Celli, B., Ferguson, G. T., Jenkins, C., … Jones, P. W. (2009). Adherence to inhaled therapy, mortality and hospital admission in COPD. Thorax, 64(11), 939–43. doi:10.1136/thx.2009.113662
21. van Staa TP1, Leufkens HG, Zhang B, Smeeth L. (2009) A comparison of cost effectiveness using data from randomized trials or actual clinical practice: selective cox-2 inhibitors as an example. PLoS Med. 2009 Dec;6(12):e1000194. doi: 10.1371/journal.pmed.1000194. Epub 2009 Dec 8.
22. NICE (2010) Chronic obstructive pulmonary disease in over 16s: diagnosis and management; nice.org.uk/guidance/cg101
23. Smeeth L, Douglas I, Hall AJ, Hubbard R, Evans S. Effect of statins on a wide range of health outcomes: a cohort study validated by comparison with randomized trials. Br J Clin Pharmacol 2009;67:99-109.
24. Wyss R, Girman CJ, LoCasale RJ, Brookhart MA, Stürmer T. Variable Selection for Propensity Score Models When Estimating Treatment Effects on Multiple Outcomes: a Simulation Study. Pharmacoepidemiology and drug safety. 2013;22(1):77-85. doi:10.1002/pds.3356.
25. Brookhart MA1, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T.Variable selection for propensity score models. Am J Epidemiol. 2006 Jun 15;163(12):1149-56. Epub 2006 Apr 19
26. Austin PC, (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies Multivariate Behav Res. 2011 May; 46(3): 399–424.
27. Berkowitz, S. A., Krumme, A. A., Avorn, J., Brennan, T., Matlin, O. S., Spettell, C. M, Choudhry, N. K. (2015). Initial Choice of Oral Glucose-Lowering Medication for Diabetes Mellitus: A Patient-Centered Comparative Effectiveness Study. JAMA Internal Medicine. oi:10.1001/jamainternmed.2014.5294
28. Rothnie, K. J., Müllerová, H., Hurst, J. R., Smeeth, L., Davis, K., Thomas, S. L., & Quint, J. K. (2016). Validation of the Recording of Acute Exacerbations of COPD in UK Primary Care Electronic Healthcare Records. PloS One, 11(3), e0151357. http://doi.org/10.1371/journal.pone.0151357

Overview of algorithms to be used for detecting COPD, COPD exacerbations and pneumonia

Condition	Paper (author, year)	Algorithm description[1]	Validity[2]	Other notes
COPD	Quint et al, 2014	- CPRD[3] diagnostic (Read) code for COPD	PPV[4]: 87% (78 – 92)	1. Comparison with gold standard of respiratory physician review of information obtained by questionnaire from GPs 2. 8 algorithms presented in total, PPVs ranging from 12 to 89
COPD exacerbation	Rothnie et al, 2016	- CPRD diagnostic (Read) code for LRTI or Acute Exacerbation COPD (AECOPD) **OR** - A prescription of a COPD-specific antibiotic combined with OCS for 5-14 days **OR** - A record (Read code) of two or more respiratory symptoms of AECOPD with a prescription of COPD-specific antibiotics and/or OCS on the same day	PPV: 86% (83 – 88) Sensitivity: 63% (55 – 70)	1. Comparison with gold standard of respiratory physician review of information obtained by questionnaire from GPs 2. 15 algorithms presented in total, PPVs ranging from 61% – 100%, sensitivities ranging from 1.6% – 63%
Pneumonia	Millet et al, 2013	- CPRD diagnostic (Read) codes and HES[5] diagnostic (ICD-10) codes for pneumonia (identified as a subset of an initial search for LRTI codes) - Records in both database within the 28 days considered the same illness-episode	No validation performed	

Note 1: Main algorithm presented in article and to be applied initially in COPD medications real-world effects study (details on other algorithms presented in paper provided in the "Other notes" column where appropriate). **Note 2**: Validity=measure of validity presented in article:result obtained (95% CI). **Note 3**: CPRD=UK Clinical Practice Research Datalink **Note 4**: PPV=positive predictive value **Note 5:** HES=Hospital Episode Statistics