# Improved outcomes in ex-smokers with COPD: a UK primary care observational cohort study

Lynn Josephs [1,2], David Culliford[3], Matthew Johnson[3] and Mike Thomas[1,2,4]

**Affiliations**: [1]Dept of Primary Care and Population Sciences, Faculty of Medicine, University of Southampton, Southampton, UK. [2]NIHR Collaboration for Leadership in Applied Health Research and Care for Wessex, University of Southampton, Southampton, UK. [3]NIHR Collaboration for Leadership in Applied Health Research and Care for Wessex Methodological Hub, Faculty of Health Sciences, University of Southampton, Southampton, UK. [4]NIHR Respiratory Biomedical Research Unit, Southampton General Hospital, Southampton, UK.

**Correspondence**: Lynn Josephs, Dept of Primary Care and Population Sciences, Faculty of Medicine, University of Southampton, Aldermoor Health Centre, Aldermoor Close, Southampton SO16 5ST, UK.
E-mail: L.Josephs@soton.ac.uk

ABSTRACT Smoking cessation in chronic obstructive pulmonary disease (COPD) reduces accelerated forced expiratory volume in 1 s decline, but impact on key health outcomes is less clear. We studied the relationship of smoking status to mortality and hospitalisation in a UK primary care COPD population.

Using patient-anonymised routine data in the Hampshire Health Record Analytical Database, we identified a prevalent COPD cohort, categorising patients by smoking status (current, ex- or never-smokers). Three outcomes were measured over 3 years (2011–2013): all-cause mortality, respiratory-cause unplanned hospital admission and respiratory-cause emergency department attendance. Survival analysis using multivariable Cox regression after multiple imputation was used to estimate hazard ratios for each outcome by smoking status, adjusting for measured confounders (including age, lung function, socioeconomic deprivation, inhaled medication and comorbidities).

We identified 16 479 patients with COPD, mean±SD age 70.1±11.1 years. Smoking status was known in 91.3%: 35.1% active smokers, 54.3% ex-smokers, 1.9% never-smokers. Active smokers predominated among younger patients. Compared with active smokers (n=5787), ex-smokers (n=8941) had significantly reduced risk of death, hazard ratio (95% confidence interval) 0.78 (0.70–0.87), hospitalisation, 0.82 (0.74–0.89) and emergency department attendance, 0.78 (0.70–0.88).

After adjusting for confounders, ex-smokers had significantly better outcomes, emphasising the importance of effective smoking cessation support, regardless of age or lung function.

## Introduction

Smoking is a major cause of chronic obstructive pulmonary disease (COPD), although individuals show differing susceptibility to smoking-related airflow limitation [1, 2] because of interacting environmental and host factors [3–5]. Smoking accelerates decline in forced expiratory volume in 1 s ($FEV_1$) in susceptible subjects [2], though physiologically defined COPD can develop in subjects with normal rates of decline who start adult life with low $FEV_1$ [6]. Evidence that stopping smoking benefits $FEV_1$ trajectories comes from the Lung Health Study, with benefit persisting after 11 years in sustained "quitters" [7] and 14 year mortality benefit [8]. However, these data relate to undiagnosed smokers in a screening programme with mild to moderate airflow obstruction [7, 8], who are therefore not representative of diagnosed COPD patients in the community.

Less is known about the impact of quitting on health outcomes in patients with a clinical diagnosis of COPD. Ex-smokers have fewer lower respiratory tract infections (LRTIs) and COPD exacerbations [9], and LRTIs promote clinical and functional decline in active but not ex-smokers [10]. COPD hospitalisations may be reduced in quitters [11] and a systematic review in 2008 suggested morbidity and mortality benefits from quitting, even in severe COPD [12].

Our objective was to use routine clinical data from electronic patient records, to quantify current smoking status in a large UK primary care cohort with diagnosed COPD, and explore the relationship with all-cause mortality, unplanned respiratory-cause hospital admission and emergency department (ED) attendance, adjusting for potential confounders. Quantification of the independent effects of smoking cessation is potentially useful to clinicians advising patients to quit, as patients with early disease may perceive that quitting is unnecessary and those with severe disease, that it is too late to make any difference.

## Methods

A study protocol is publically available [13]. The study received ethical approval from the University of Southampton and governance approval from the Hampshire Health Record Information Governance Group.

We report our findings in line with STROBE [14] and RECORD [15] guidelines for observational studies using routinely collected health data (supplementary material).

### Setting

This was a retrospective observational cohort study of patients with a primary care diagnosis of COPD, with outcomes over 3 years (2011–2013) or until death, using individual patient-anonymised routinely collected clinical data held in Hampshire Health Record Analytical Database (HHRA), an electronic UK regional linked primary–secondary care database of around 1.4 million patients living in Hampshire, UK (supplementary material).

### Participants

We identified a prevalent COPD cohort using Read diagnostic codes as at December 31, 2010. Included patients had 3 years' continuous data from January 1, 2011 (or until death) and were over 25 years old.

### Explanatory and outcome variables

Patients were characterised by baseline parameters from routine clinical records: current smoking status, age, sex, socioeconomic deprivation (index of multiple deprivation (IMD)), body mass index (BMI), $FEV_1$/forced vital capacity (FVC) %, $FEV_1$ % predicted, use of inhaled long-acting bronchodilators and corticosteroids and by the presence or absence of comorbidities. Smoking status was categorised as active smoker, ex-smoker or never-smoker, based on the most recent code prior to January 1, 2011; time from baseline smoking code to study start was quantified.

Never-smokers were re-assigned as ex-smokers if there were contradictory preceding smoking codes.

Duration of COPD was estimated from the interval between first COPD code and the study start.

Comorbidities were defined using diagnostic codes for anxiety/depression, asthma, bronchiectasis, cerebrovascular disease, chronic kidney disease, connective tissue disease, dementia, diabetes mellitus, gastro-oesophageal reflux, heart failure, hyperlipidaemia, hypertension, ischaemic heart disease, lung cancer, obstructive sleep apnoea, osteoporosis, peripheral vascular disease and rhino-sinusitis.

Outcomes were based on coded clinical entries: all-cause mortality, respiratory-cause unplanned hospital admissions (based on primary diagnosis) and respiratory-cause ED attendances.

Additional details and all code lists are available in the supplementary material.

*Statistical methods*

Details of data handling within HHRA (data access, composition and cleaning) are described in the supplementary material.

Summary measures were used to describe patient characteristics: mean and standard deviation for all except IMD, duration of COPD diagnosis and time from baseline smoking status to study start, where median and interquartile range were used. The main analysis used Cox regression to estimate the effect of baseline smoking status on time to the occurrence of each of the three outcomes: death from any cause, first unplanned respiratory-cause hospital admission and first respiratory-cause ED attendance, with event times censored after 3 years. Patients with unknown smoking status were excluded from this analysis. The effects of smoking status are reported as hazard ratios, with full adjustment for the following potential confounders: age, gender, BMI, IMD rank decile, $FEV_1$ % predicted, baseline inhaled medication and presence of comorbidities listed above. Age, BMI and $FEV_1$ % predicted were modelled as continuous variables but all other variables were categorical. To help interpret adjusted hazard ratios for each outcome, we calculated the population attributable risk (PAR) fraction [16] for active smokers to estimate the proportion of deaths and hospitalisations that could be avoided if all active smokers became ex-smokers (supplementary material).

Additional sensitivity analyses were performed, in which $FEV_1$/FVC % and then time from COPD diagnosis to study start (modelled as continuous variables) were included as further confounding factors. Smoking pack-years were not included in our *a priori* analysis plan, because of poor recording in UK routine clinical data (see discussion), but are included in additional analyses in the supplementary material.

Multiple imputation using chained equations with fully conditional specification [17] was used to impute missing data among the adjustment variables, based on an assumption that data were "missing at random" [18]. Five imputations were used, and all the adjustment variables listed in the main analysis above were included in the imputation model, along with the three time-to-event outcomes and their event/censoring indicators. Smoking status was not imputed, as it was the key explanatory variable and was observed for over 90% of the cohort.

All parameter estimates are presented with 95% confidence intervals. All tests were conducted as two-sided, at the 5% significance level. Analyses were conducted using the statistical software packages SAS v9.4 (SAS Institute, Cary, NC, USA), SPSS v22 (IBM Corp, Armonk, NY, USA) and R v3.1 (R Core Team, Vienna, Austria).

## Results

We identified 16 479 patients with a primary care diagnosis of COPD and continuous data: 53.7% male, mean ±SD age 70.1±11.1 years. Table 1 summarises baseline demographic and clinical characteristics. Smoking status was recorded in 91.3% (n=15 040): active smokers 38.5% (n=5787), ex-smokers 59.4% (n=8941) never-smokers 2.1% (n=312). Smoking status was based on data close to the study start: median (interquartile range (IQR)) 3 (1–8) months for active smokers, 5 (2–10) months for ex-smokers and 8 (2–13) months for never-smokers. One or more comorbidities were recorded in 90.7%. $FEV_1$ % predicted values were available in 80.6%.

More ex-smokers (56.1%) than active smokers (51.6%) or never-smokers (32.7%) were male. Ex-smokers and never-smokers were significantly older than active smokers, were more overweight, had less severe airflow obstruction and had a greater number of comorbidities.

In younger patients, active smokers predominated, but ex-smokers predominated among older patients (figure 1): 58.6% of those aged <65 years were active smokers compared to 30.2% in those aged ⩾65 years (p<0.001). This pattern was evident across all severities of airflow obstruction, as defined by the Global Initiative for Chronic Obstructive Lung Disease (figure 2) [19].

Table 2 summarises the outcomes for the cohort as a whole and for each of the smoking categories. During the 3 years, 2101 (12.7%) patients died, with a significantly greater proportion (p<0.001) of ex-smokers, n=1283 (14.3%) and never-smokers, n=50 (16.0%) than active smokers, n=640 (11.1%). A total of 2909 (17.7%) patients had one or more unplanned respiratory-cause hospital admission, with no significant difference in admission rates between the three smoking categories. A total of 1581 (9.6%) patients had one or more respiratory-cause ED attendance, with a significantly higher proportion among active smokers, n=631 (10.9%) than ex-smokers, n=799 (8.9%) (p<0.001) and never-smokers, n=21 (6.7%) (p=0.020). The median time to death, first respiratory admission and first respiratory ED attendance among those experiencing events was 16.7, 14.3 and 16.5 months respectively.

Univariate and adjusted hazard ratios for the effect of smoking status on time to each of the outcomes, using active smokers as comparator, are shown in table 3. In univariate analysis there was an increased

TABLE 1 Baseline demographic and clinical characteristics in total cohort and in each smoking category

| | Total cohort (n=16479) | | Cohort with known smoking status prior to study commencement (n=15040, 91.3%) | | |
| --- | --- | --- | --- | --- | --- |
| | Descriptive statistics | Missing | Active smokers (n=5787, 38.5%) | Ex-smokers (n=8941, 59.4%) | Never-smokers (n=312, 2.1%) |
| Male | 8847 (53.7%) | NA | 2986 (51.6%) | 5018 (56.1%) | 102 (32.7%) |
| Age years | 70.1±11.1 | NA | 65.8±10.8 | 72.8±10.1 | 74.9±13.1 |
| IMD rank decile | 6 (3–8) | 61 (0.4%) | 6 (3–8) | 6 (2–9) | 7 (2–9) |
| BMI kg·m$^{-2}$ | 27.2±6.0 | 1929 (11.7%) | 26.3±6.2 | 27.8±5.8 | 26.5±5.5 |
| FEV$_1$ % pred | 60.7±20.0 | 3186 (19.4%) | 60.4±19.1 | 60.6±20.4 | 65.0±22.5 |
| FEV$_1$ L | 1.51±0.63 | 3256 (19.8%) | 1.57±0.65 | 1.48±0.61 | 1.35±0.59 |
| FEV$_1$/FVC % | 59.4±14.9 | 4586 (27.8%) | 58.9±14.4 | 59.7±15.1 | 64.2±16.1 |
| Comorbidities[#] n | 2.6±1.8 | NA | 2.4±1.7 | 2.8±1.8 | 2.9±1.7 |
| Inhaled LABA | 6386 (38.8%) | NA | 2157 (37.3%) | 3960 (44.3%) | 109 (34.9%) |
| Inhaled LAMA | 4573 (27.8%) | NA | 1621 (28.0%) | 2795 (31.3%) | 44 (14.1%) |
| Inhaled ICS | 7315 (44.4%) | NA | 2501 (43.2%) | 4484 (50.2%) | 150 (48.1%) |
| Months from COPD diagnosis to study start | 61 (26–108) | 337 (2.0%)[¶] | 56 (23–96) | 69 (34–116) | 72 (36–133) |
| Months from last recorded smoking status to study start | 4 (2–9) | 1439 (8.7%) | 3 (1–8) | 5 (2–10) | 8 (2–13) |

Data are presented as n (%), mean±SD or median (interquartile range), unless otherwise stated. IMD: index of multiple deprivation (a weighted standardised measure of socioeconomic status); BMI: body mass index; FEV$_1$: forced expiratory volume in 1 s; FVC: forced vital capacity; LABA: long-acting β$_2$-agonist; LAMA: long-acting antimuscarinic bronchodilator; ICS: inhaled corticosteroid; [#]: anxiety/depression, asthma, bronchiectasis, cerebrovascular disease, chronic kidney disease, connective tissue disease, dementia, diabetes mellitus, gastro-oesophageal reflux, heart failure, hyperlipidaemia, hypertension, ischaemic heart disease, lung cancer, obstructive sleep apnoea, osteoporosis, peripheral vascular disease, rhino-sinusitis; [¶]: exact date of chronic obstructive pulmonary disease (COPD) diagnosis unavailable for 2% of cohort.

risk of death in both ex-smokers and never-smokers compared with active smokers. However, this relationship inverts in multivariate analysis when adjusted for confounding factors (age, gender, BMI, IMD, FEV$_1$ % predicted, baseline inhaled medication, presence of comorbidities), with both ex-smokers (HR 0.78, CI 0.70–0.87; p<0.001) and never-smokers (HR 0.72, CI 0.53–0.99; p=0.039) having a reduced risk of death. In both sensitivity analyses, inclusion of additional potential confounders (FEV$_1$/FVC % and time from COPD diagnosis to study start) had minimal effect on hazard ratios, which changed by ⩽0.01 (details in supplementary material).

For hospitalisation data, multivariate analysis showed a significantly reduced risk of respiratory-cause hospital admission (HR 0.82, CI 0.74–0.89; p<0.001) and respiratory-cause ED attendance (HR 0.78, CI 0.70–0.88; p<0.001) in ex-smokers. In never-smokers, there were trends towards reduced hospital admission (HR 0.79, CI 0.59–1.06; p=0.113) and ED attendance (HR 0.71, CI 0.45–1.11, p=0.130) that failed to reach significance at the 5% level (table 3).
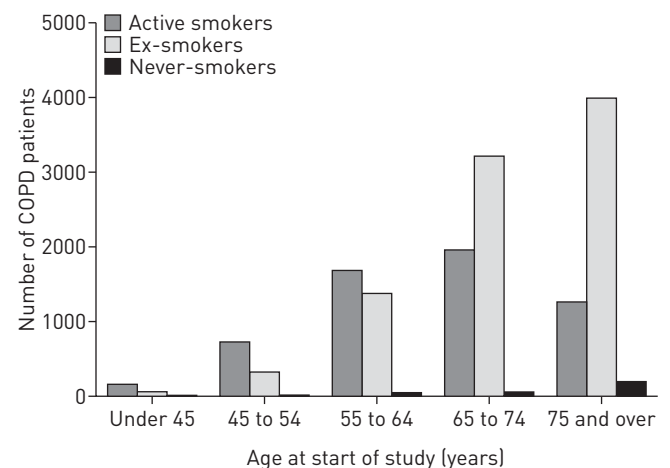


FIGURE 1 Distribution of smoking status by age at start of study. COPD: chronic obstructive pulmonary disease.
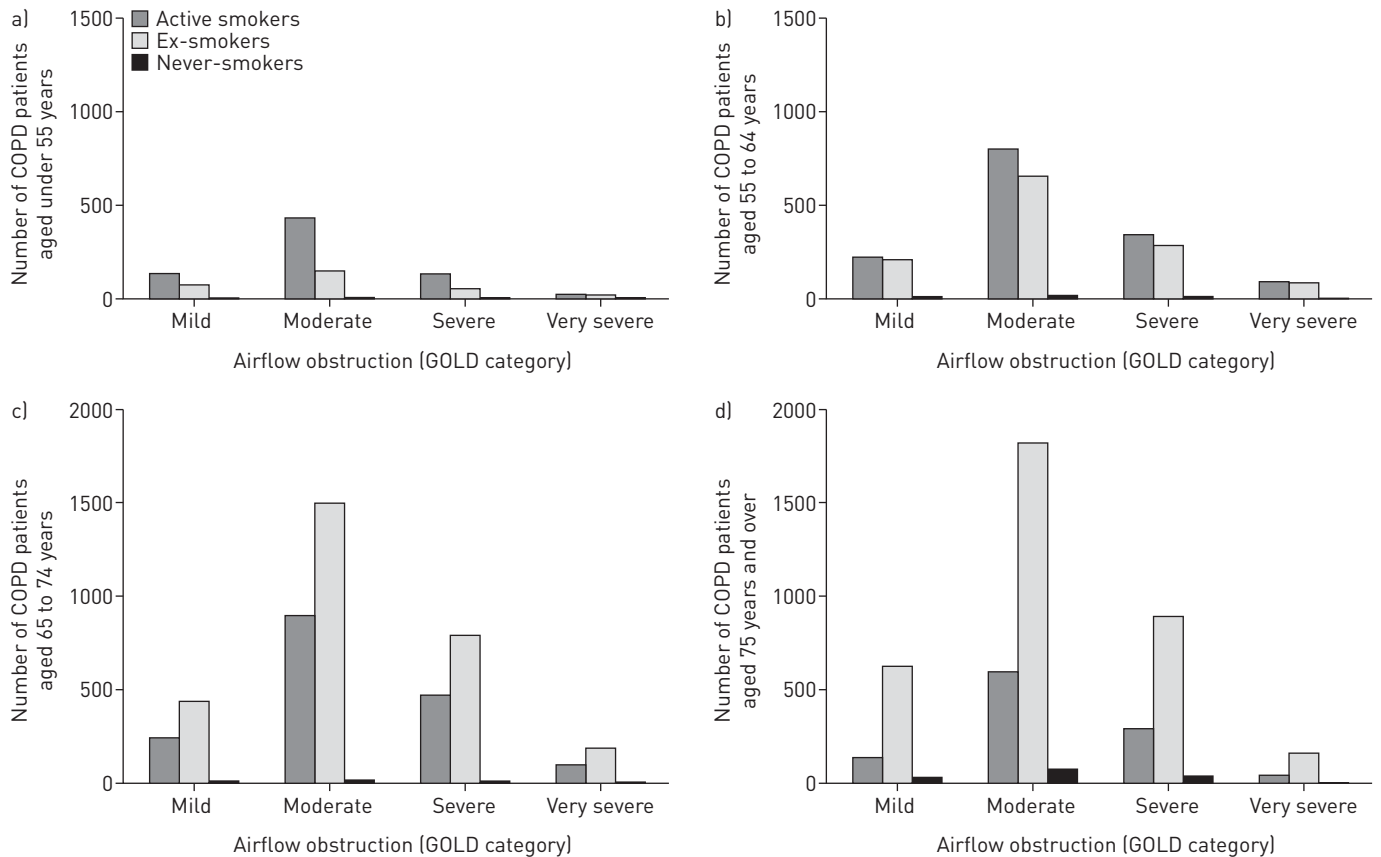
FIGURE 2 Distribution of age-specific smoking status categorised by severity of airflow obstruction: a) under 55 years, b) 55–64 years, c) 65–74 years, and d) 75 years and older. COPD: chronic obstructive pulmonary disease; GOLD: Global Initiative for Chronic Obstructive Lung Disease [19].

Based on PAR fraction estimates, over the 3 years 14.6% of the COPD population (excluding never-smokers) would avoid death, 11.8% would avoid at least one respiratory-cause admission and 14.6% would avoid at least one respiratory-cause ED attendance, if all active smokers became ex-smokers.

### Discussion

Our objectives were twofold: to quantify smoking status in a large and representative clinically defined UK primary care COPD cohort and to explore the relationship between smoking status and clinical outcomes.

We were able to define smoking status in 91.3% of our cohort and have shown that over one-third continue to smoke, with the highest prevalence of active smokers among younger patients, consistent with

TABLE 2 Variation in outcomes over 3 years (mortality and hospitalisation) for total cohort and for each of the smoking categories

| | Total cohort (n=16 479) | Cohort with known smoking status prior to study commencement (n=15 040) | | | | | |
|---|---|---|---|---|---|---|---|
| | | Active smokers[#] (n=5787) | Ex-smokers[#] (n=8941) | Never-smokers[#] (n=312) | Chi-squared test (overall) | Pairwise test[¶] (active versus ex-smoker) | Pairwise test[¶] (active versus never-smoker) |
| Deaths n (%) | 2101 (12.7) | 640 (11.1) | 1283 (14.3) | 50 (16.0) | p<0.001 | p<0.001 | p=0.007 |
| One or more respiratory cause admissions n (%) | 2909 (17.7) | 1018 (17.6) | 1603 (17.9) | 52 (16.7) | p=0.763 | p=0.601 | p=0.676 |
| One or more respiratory cause ED attendances n (%) | 1581 (9.6) | 631 (10.9) | 799 (8.9) | 21(6.7) | p<0.001 | p<0.001 | p=0.020 |

ED: emergency department. [#]: smoking status at study commencement was unavailable in 1439 subjects (8.7% of total cohort); [¶]: p-values to be compared against a Bonferroni-adjusted target of 0.013 (three possible pairwise tests per outcome).

TABLE 3 Cox regression analysis of time to death, first hospital admission and first emergency department (ED) attendance (hazard ratios with 95% confidence intervals and p-values)

| Outcome | Active smokers[¶] (n=5787) | Ex-smokers (n=8941) | | | Never-smokers (n=312) | | |
|---|---|---|---|---|---|---|---|
| | | Hazard ratio | 95% CI | p-value | Hazard ratio | 95% CI | p-value |
| **Time to death (from any cause)** | | | | | | | |
| Univariate | 1.00 | 1.31 | 1.19–1.43 | <0.001 | 1.48 | 1.11–1.97 | 0.008 |
| Adjusted[#] | 1.00 | 0.78 | 0.70–0.87 | <0.001 | 0.72 | 0.53–0.99 | 0.039 |
| **Time to first respiratory cause admission** | | | | | | | |
| Univariate | 1.00 | 1.03 | 0.95–1.12 | 0.486 | 0.97 | 0.73–1.28 | 0.826 |
| Adjusted[#] | 1.00 | 0.82 | 0.74–0.89 | <0.001 | 0.79 | 0.59–1.06 | 0.113 |
| **Time to first respiratory cause ED attendance** | | | | | | | |
| Univariate | 1.00 | 0.83 | 0.74–0.92 | <0.001 | 0.63 | 0.41–0.98 | 0.038 |
| Adjusted[#] | 1.00 | 0.78 | 0.70–0.88 | <0.001 | 0.71 | 0.45–1.11 | 0.130 |

[#]: the analysis was adjusted for: age, gender, body mass index, index of multiple deprivation rank decile, forced expiratory volume in 1 s % pred, presence of comorbidities (anxiety/depression, asthma, bronchiectasis, cerebrovascular disease, chronic kidney disease, connective tissue disease, dementia, diabetes mellitus, gastro-oesophageal reflux, heart failure, hyperlipidaemia, hypertension, ischaemic heart disease, lung cancer, obstructive sleep apnoea, osteoporosis, peripheral vascular disease, rhino-sinusitis) and inhaled medication at baseline (long-acting $\beta_2$-agonist, long-acting antimuscarinic bronchodilator, inhaled corticosteroid). Multiple imputation was used in the adjusted analysis, as described in the methods section. [¶]: reference category.

other reports [20]. After adjusting for confounding factors (including age, disease severity and inhaled medication) ex-smokers had a significantly reduced risk of all three adverse outcomes: death, unplanned respiratory-cause hospital admissions and ED attendances, with an estimated reduction in these outcomes of 14.6%, 11.8% and 14.6% respectively if all smokers quit. In never-smokers, mortality risk was reduced, as others have reported [21], but for hospitalisation the adjusted effect estimates were consistent in size and direction with those in ex-smokers, providing weaker evidence of reduced risk. Failure to reach significance at the 5% level is probably because the number of never-smokers was small, but might also be because they represent an especially heterogeneous group.

We acknowledge the limitations and biases inherent in observational studies using routine clinical data that have not been collected to answer specific research questions and where factors motivating data recording vary. We report our findings in accordance with the RECORD statement [15]. Although not all Hampshire practices submit data to HHRA, missing practices are dispersed across the catchment area, with varying rural/urban classification, socioeconomic deprivation and patient composition. We are not aware of any systematic differences between missing practices and those whose data are represented.

There are problems in accurately identifying diseases using electronic health records, since diagnostic codes are used inconsistently in primary care, without quality control at the point of data entry [22]. However, use of electronic templates, common in UK general practice (and widely used in COPD), reduces missing data and encourages correct code use. QUINT et al. [23] have shown that COPD patients can be reliably identified from routine health records using diagnostic codes, with little additional benefit from including spirometry. All our code lists were created by clinicians familiar with the Read code system in routine practice. Our cohort-defining codes included designated codes in the UK "Quality and Outcome Framework" (QoF) (www.hscic.gov.uk/qof) incentive scheme that define COPD for performance-related pay calculations and have been widely used since the scheme began in 2004. COPD is underdiagnosed, but the size of our prevalent cohort (1.2% of HHRA population) is consistent with reported prevalence rates in England [24], so it is likely to be representative of the UK diagnosed COPD population.

Smoking status is now widely recorded in UK primary care [25]. Inconsistent recording may occur, relating to coding, to different interpretations of the terms "current non-smoker", "never", "trivial" and "ex"-smoker in routine practice and because reports of quitting are often exaggerated [26]. We used the last smoking code prior to the study start, to improve the validity of any relationship with outcomes, and have demonstrated that these data were recent. A limitation of our data source is that we were unable to quantify smoking exposure (e.g. pack-years) in most of our patients, or the time since quitting, factors shown to be important in affecting outcomes [27]. Pack-year data were not included in our primary analysis plan, because we expected these to be infrequently and inconsistently recorded in routine primary care. A supplementary "post hoc" analysis supports this view, in that pack-year data were only recorded in

15% of patients and showed no relationship to outcomes (see supplementary material). Reported ex-smoker outcomes also depend on how long subjects are required to have not smoked to be considered ex-smokers [28]. Our data do not allow calculation of time since quitting and estimates of a minimum quit time would be problematic in relation to time-to-event analyses, because estimates would include left-censored covariates. Nevertheless, despite these limitations, we have demonstrated significant and clinically important differences in outcomes between active and ex-smokers.

We may have underestimated the size of the association between smoking status and outcomes because of selective survival bias in our prevalent cohort. Active smokers in the cohort may be healthier than those who died before cohort entry; however, ex-smokers (the main comparator group) might also exhibit this effect, although perhaps to a lesser degree. The extent of survival bias because of differing disease duration may only be modest, since the median time from COPD diagnosis to study start for ex-smokers was only 13 months more than for active smokers. Greater smoking exposure and reduced time since quitting might both increase mortality prior to cohort entry. We were limited in not having smoking cessation date or pack-year history, with which we could have introduced time-varying covariate information to reduce the extent of potential survivorship bias.

Missing data are an important limitation of any study using routinely collected health data. We used multiple imputation to impute missing data among the adjustment variables used in our Cox regression analyses. This depends on the assumption that imputed data are "missing at random" (MAR), conditional upon the presence of certain variables. We have attempted to satisfy this condition in our imputation model, but acknowledge that MAR is an essentially untestable assumption. Baseline data used to characterise our cohort reflect the most recent values recorded prior to the study start. Later data were not included but, by the end of the study, it was noted that missing data for lung function and BMI had halved and smoking status was missing in only 2.6%, suggesting better recording within primary care.

The disparity between our univariate and multivariate outcome analyses highlights the fact that observational data need to be interpreted with care, because of potential confounding factors. Unadjusted outcomes in ex-smokers were confounded by factors such as age and severity but, in adjusted analyses, smoking cessation was independently associated with both reduced mortality and hospitalisation. Our findings are consistent with a recent Italian study (using life tables and smoking questionnaire data) which showed that quitting smoking confers mortality benefits at all ages, with a one-third reduction in the risk of dying from COPD in those aged 70–74 years [29]. Similarly, a large Australian study showed reduced hospitalisation because of COPD within 5–14 years of stopping smoking, even at older ages [27]. Reduced hospitalisation has been demonstrated in both early and sustained quitters [11], but other studies report increased admission rates in recent quitters (less than 5 years) but reduced respiratory symptoms in sustained quitters [30] and a reducing risk of exacerbations with longer periods of abstinence [9].

The prevalence of never-smokers in our clinically defined cohort is considerably lower than in population-based cross-sectional surveys of subjects whose spirometry met criteria for COPD [31, 32]. Many factors cause COPD worldwide [33], but clinical experience in the UK suggests that COPD in a non-smoker is uncommon and indicates the need for diagnostic review and consideration of asthma or rare diseases such as $\alpha_1$-antitrypsin deficiency. However, one explanation for low numbers of never-smokers on practice COPD registers is that the diagnosis is less likely to be considered, because of the disease's clear association with smoking. In this study, we re-classified, as ex-smoker, those patients whose latest code was "never-smoker", but whose earlier codes indicated a past history of smoking. We believe this strategy improved the overall reliability of our never-smoker definition, though potentially overcompensating if earlier smoking history was incorrect.

A strength of HHRA is that it links routinely collected patient-anonymised data from primary and secondary care, providing more reliable estimates of secondary care contacts than most UK primary care datasets, in which hospitalisation data rely on coding in primary care records after discharge [34]. We were able to use hospital-coded discharge data for the entire cohort and used "primary" diagnoses to quantify respiratory-cause admissions. A systematic review has shown primary diagnosis coding accuracy of 96.0% (IQR 89.3–96.3), concluding that "routinely collected data are sufficiently robust to support their use for research and managerial decision-making" [35]. In quantifying hospitalisations, we have not made any distinction between patients with one or more episodes, mainly because of the low numbers of patients with more than two. We report all-cause mortality, as we do not have access to certified cause of death.

We believe that our results are generalisable to a wider UK population. HHRA holds data for over 1.4 million patients, representing >75% of Hampshire's population. COPD prevalence, healthcare and outcomes (hospitalisation and mortality) appear to be "average" when Hampshire is compared with the rest of England, although considerable variation has been demonstrated within Hampshire [24]. Hampshire may not be representative of large inner cities with marked ethnic diversity; indeed, 89% of its inhabitants were White

British in the 2011 census [36]. However, there is heterogeneity across the county, with urban areas having greater ethnic diversity, especially Southampton, where the proportion of White British is 77.7%, compared with a national average of 80.5%. Ethnicity was poorly recorded in our dataset but our cohort encompassed a broad socioeconomic spectrum, as reflected in the range of IMD scores (IQR 3–8, median 6) when these weighted standardised measures of socioeconomic status were ranked according to national deciles.

In conclusion, we used a UK primary–secondary care linked database to define a cohort of >16 000 patients with a clinical diagnosis of COPD in whom we assessed smoking status and its relationship to major clinical outcomes over 3 years. Our findings suggest that 14.6% of deaths and respiratory-cause ED attendances and 11.8% of respiratory-cause admissions could be avoided if all smokers quit. Over one-third of this cohort continues to smoke, as do the majority of younger patients, who have the most to gain by stopping. Our findings could be used to encourage patients to quit. Smoking cessation should be top priority in COPD [37, 38], but worrying evidence suggests that fewer people are seeking specialist services [39], despite advances in the science and delivery of smoking cessation support [40, 41]. While funding cuts have led to reduced provision of these services, our findings emphasise the need for continued investment to improve patient care and reduce pressures on health services.

### References

1   Fletcher C, Peto R. The natural history of chronic airflow obstruction. *BMJ* 1977; 1: 1645–1648.
2   Kohansal R, Martinez-Camblor P, Agustí A, *et al.* The natural history of chronic airflow obstruction revisited. *Am J Respir Crit Care Med* 2009; 180: 3–10.
3   Bartal M. COPD and tobacco smoke. *Monaldi Arch Chest Dis* 2005; 63: 213–225.
4   He JQ, Shumansky K, Connett JE, *et al.* Association of genetic variations in the CSF2 and CSF3 genes with lung function in smoking-induced COPD. *Eur Respir J* 2008; 32: 25–34.
5   Betsuyaku T, Fuke S, Inomata T, *et al.* Bronchiolar epithelial catalase is diminished in smokers with mild COPD. *Eur Respir J* 2013; 42: 42–53.
6   Lange P, Celli B, Agusti A, *et al.* Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med* 2015; 373: 111–122.
7   Anthonisen NR, Connett JE, Murray RP. Smoking and lung function of Lung Health Study participants after 11 years. *Am J Respir Crit Care Med* 2002; 166: 675–679.
8   Anthonisen NR, Skeans MA, Wise RA, *et al.* The effects of a smoking cessation intervention on 14.5-year mortality: a randomized clinical trial. *Ann Intern Med* 2005; 142: 233–239.
9   Au DH, Bryson CL, Chien JW, *et al.* The effects of smoking cessation on the risk of chronic obstructive pulmonary disease exacerbations. *J Gen Intern Med* 2009; 24: 457–463.
10  Dreyse J, Diaz O, Repetto PB, *et al.* Do frequent moderate exacerbations contribute to progression of chronic obstructive pulmonary disease in patients who are ex-smokers? *Int J Chron Obstruct Pulmon Dis* 2015; 10: 525–533.
11  Godtfredsen NS, Vestbo J, Osler M, *et al.* Risk of hospital admission for COPD following smoking cessation and reduction: a Danish population study. *Thorax* 2002; 57: 967–972.
12  Godtfredsen NS, Lam TH, Hansel TT, *et al.* COPD-related morbidity and mortality after smoking cessation: status of the evidence. *Eur Respir J* 2008; 32: 844–853.
13  NIHR CLAHRC Wessex. Wessex CLAHRC: Integrated Respiratory Care (Theme 1) Study protocol. www.clahrc-wessex.nihr.ac.uk/theme/project/42. Date last updated: May 10, 2016. Date last accessed: October 25, 2016.
14  Vandenbroucke JP, Von Elm E, Altman DG, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007; 147: W-163-W-194.
15  Benchimol EI, Smeeth L, Guttmann A, *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med* 2015; 12: e1001885.
16  Crowson CS, Therneau TM, O'fallon WM. Attributable risk estimation in cohort studies. Technical Report #82. Rochester, Mayo Clinic, 2009.
17  Van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007; 16: 219–242.
18  Little R, Rubin D. Analysis with missing data. New York, John Wiley and Sons, 1987.
19  Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD; 2016. www.goldcopd.org/ Date last accessed: October 25, 2016. Date last updated: November 15, 2016.
20  Sanchez-Salcedo P, Divo M, Casanova C, *et al.* Disease progression in young patients with COPD: rethinking the Fletcher and Peto model. *Eur Respir J* 2014; 44: 324–331.
21  Thomsen M, Nordestgaard BG, Vestbo J, *et al.* Characteristics and outcomes of chronic obstructive pulmonary disease in never smokers in Denmark: a prospective population study. *Lancet Respir Med* 2013; 1: 543–550.
22  Manuel DG, Rosella LC, Stukel TA. Importance of accurately identifying disease in studies using electronic health records. *BMJ* 2010; 341: c4226.
23  Quint JK, Müllerova 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.

24 National Health Service. NHS Atlas of Variation in Healthcare for People with Respiratory Disease. September 2012: Rightcare. Available from: https://fingertips.phe.org.uk/documents/Atlas_2012%20Respiratory.pdf

25 Taggar JS, Coleman T, Lewis S, et al. The impact of the Quality and Outcomes Framework (QOF) on the recording of smoking targets in primary care medical records: cross-sectional analyses from The Health Improvement Network (THIN) database. BMC Public Health 2012; 12: 329.

26 Stelmach R, Fernandes FL, Carvalho-Pinto RM, et al. Comparison between objective measures of smoking and self-reported smoking status in patients with asthma or COPD: are our patients telling us the truth? J Bras Pneumol 2015; 41: 124–132.

27 Tran B, Falster MO, Douglas K, et al. Smoking and potentially preventable hospitalisation: The benefit of smoking cessation in older ages. Drug Alcohol Depend 2015; 150: 85–91.

28 Leffondré K, Abrahamowicz M, Siemiatycki J, et al. Modeling smoking history: a comparison of different approaches. Am J Epidemiol 2002; 156: 813–823.

29 Carreras G, Pistelli F, Falcone F, et al. Reduction of risk of dying from tobacco-related diseases after quitting smoking in Italy. Tumori 2015; 101: 657–663.

30 Abu Hassan H, Abd Aziz N, Hassan Y, et al. Does the duration of smoking cessation have an impact on hospital admission and health-related quality of life amongst COPD patients? Int J Chron Obstruct Pulmon Dis 2014; 9: 493–498.

31 Lamprecht B, McBurnie MA, Vollmer WM, et al. COPD in never smokers: results from the population-based burden of obstructive lung disease study. Chest 2011; 139: 752–763.

32 Tan W, Sin D, Bourbeau J, et al. Characteristics of COPD in never-smokers and ever-smokers in the general population: results from the CanCOLD study. Thorax 2015; 70: 822–829.

33 Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. Lancet 2009; 374: 733–743.

34 Millett ER, Quint JK, De Stavola BL, et al. Improved incidence estimates from linked vs. stand-alone electronic health records. J Clin Epidemiol 2016; 75: 66–69.

35 Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. J Public Health (Oxf) 2012; 34: 138–148.

36 Hampshire County Council. Ethnicity statistics for Hampshire; 2016. www3.hants.gov.uk/factsandfigures/population-statistics/ethnicity.htm Date last accessed: October 20, 2016. Date last updated: February 11, 2014.

37 Tønnesen P, Carrozzi L, Fagerström K, et al. Smoking cessation in patients with respiratory diseases: a high priority, integral component of therapy. Eur Respir J 2007; 29: 390–417.

38 Tønnesen P. Smoking cessation and COPD. Eur Respir Rev 2013; 22: 37–43.

39 Kmietowicz Z. Action is needed to boost uptake of stop smoking services, say campaigners. BMJ 2015; 351: h4578.

40 Quaak M, van Schayck CP, Knaapen AM, et al. Genetic variation as a predictor of smoking cessation success. A promising preventive and intervention tool for chronic respiratory diseases? Eur Respir J 2009; 33: 468–480.

41 Zwar NA, Mendelsohn CP, Richmond RL. Supporting smoking cessation. BMJ 2014; 348: f7535.