

Mortality Outcomes in People with Lung Cancer with and without Type2 Diabetes: A Cohort Study in England

Eseosa Grace Igbinosa^{1,2}, Bodini Dharmasekara¹, Jennifer K Quint¹, Sanjay Popat³, Krishnan Bhaskaran⁴, Daniel Morganstein^{3,5}, Sarah Cook¹

¹School of Public Health, Imperial College London, London, UK; ²School of Computer Science and Mathematics, Keele University, Staffordshire, UK; ³Lung Unit, The Royal Marsden NHS Foundation Trust, London, UK; ⁴Department of Non-Communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK; ⁵Department of Endocrinology, Chelsea and Westminster Hospital NHS Foundation Trust, London, UK

Correspondence: Sarah Cook, School of Public Health, 9th Floor Michael Uren Building, White City Campus, Imperial College London, London, W12 0BZ, UK, Email sarah.cook@lshtm.ac.uk

Introduction: The impact of type 2 diabetes (T2DM) on mortality following lung cancer diagnosis remains unclear, with conflicting evidence across studies. We aimed to assess differences in all-cause and cause-specific mortality between people with lung cancer with and without T2DM within a primary care population in England.

Methods: The study population was 69,674 people with incident lung cancer within the Clinical Practice Research Datalink (CPRD) Aurum primary care database (2010–2022). The study exposure was T2DM at cancer diagnosis, and the outcomes were all-cause and cause-specific mortality (cancer, cardio-vascular, respiratory). Cox models were fitted for each outcome adjusting for age, gender, smoking status, body mass index, calendar year and socioeconomic status (Index of Multiple Deprivation).

Results: After adjusting for age and gender, there was no evidence for a difference in all-cause mortality in people with T2DM compared with people without T2DM (IRR 0.98 95% CI 0.96, 1.01). After fully-adjusting for measured confounders, there was a small positive effect (IRR 1.07 95% CI 1.04, 1.09). After adjusting for age and gender, people with T2DM had lower rates of cancer-specific mortality compared to people without T2DM (IRR 0.96 95% CI 0.94, 0.98). However, after adjustment for all measured confounders there was a small positive association (IRR 1.05 95% CI 1.02, 1.07). In both age and gender adjusted and fully adjusted models people with T2DM had higher cardiovascular (fully adjusted HR 1.30 95% CI 1.15, 1.47) and respiratory disease mortality (fully adjusted HR 1.30 95% CI 1.15, 1.47).

Conclusion: There was robust evidence that people with T2DM had higher cardiovascular and respiratory disease mortality following lung cancer diagnosis. The relationships between T2DM and all-cause and cancer-specific mortality were highly sensitive to adjustment for confounding. Differences in studies on approaches to confounding and levels of missing data may contribute to the mixed findings on this association in the literature.

Plain Language Summary: Lung cancer and type 2 diabetes are both common conditions that have a big impact on health outcomes worldwide. There is some evidence that people with type 2 diabetes have a higher chance of dying if they develop certain cancers than people without type 2 diabetes. It is not clear if having type 2 diabetes affects people's chances of dying if they develop lung cancer, or if any increased risk of death is due to lung cancer or other conditions such as heart disease. We used data from general practice records linked with hospital and death records from England to investigate whether people with diabetes were more or less likely to die following being diagnosed with lung cancer overall and from three specific causes of death: cancer, heart disease and lung disease.

People with type 2 diabetes were more likely to die from heart disease and lung-related illnesses after being diagnosed with lung cancer than people without type 2 diabetes. However, differences in dying overall and of dying from cancer were small and depended on how other factors, like age and smoking, were considered. These findings highlight the importance of managing heart and lung health in people living with type 2 diabetes who are diagnosed with lung cancer.

Keywords: lung cancer, type 2 diabetes, mortality, cardiovascular mortality, respiratory mortality

Introduction

Cancer and Diabetes Mellitus (DM) are common diseases that contribute to high global morbidity and mortality. It is estimated that 20% of people with cancer have DM, mostly Type 2 DM (T2DM).¹ Several epidemiological studies have suggested patients with T2DM may experience higher all-cause mortality following a diagnosis of cancer.^{2–4} It remains unclear if this increased mortality is driven by worse cancer outcomes or by an increased risk of death from other conditions such as cardiovascular disease. Some studies have reported worse cancer-specific mortality, notably in breast and colorectal cancer.⁵ However, there is potential bias in determining the cause of death, with evidence that in people with diabetes, diabetes is more frequently recorded on death certificates if someone dies of cardiovascular disease than if they die of cancer.⁶ Conversely, deaths in people with diabetes have a higher probability of being attributed to cardiovascular causes.⁷ Notably, there are conflicting results from studies assessing outcomes to cancer treatments in diabetes (reviewed in Joharatnam-Hogan et al 2023).⁸

Lung cancer, classified into Non-Small Cell Lung Cancer (NSCLC) and Small-Cell Lung Cancer (SCLC), is the third most diagnosed cancer in the UK accounting for 21% of all cancer-related mortality.⁹ Despite recent advances in diagnosis, staging and treatment, the 5-year survival rate for lung cancer remains unfavourable, at 19.4%.¹⁰

The impact of co-morbid T2DM on lung cancer survival remains unclear as the evidence to date has shown mixed results. Whilst several studies have shown worse overall survival,^{11–18} both following surgical resection,¹¹ and in advanced disease,^{14–17} this is not universal and some studies have shown no difference² or the opposite effect— that prevalent T2DM is associated with better survival from lung cancer.^{19,20} Studies to date have been limited by small sample sizes, often single centre, and have used varying methods to identify DM. In particular, a large proportion of studies have been conducted in East Asian populations, which may differ from Western populations in smoking rates, diet, exercise, insulin resistance levels, and biological lung cancer subtypes,²¹ all risk factors for and potential confounders of the relationship between lung cancer and T2DM and therefore not necessarily generalisable. Although some studies have suggested an increased risk of lung cancer in those with diabetes,²² more recent large studies, including those using Mendelian randomisation do not support an increased risk of lung cancer in those with diabetes.^{23,24} However, they are both common conditions, and numerous studies have shown that up to 20% of people with lung cancer have co-occurring DM.¹ The impact of this is unclear. This study aimed to investigate whether T2DM is associated with differences in all-cause and cause-specific mortality in individuals diagnosed with lung cancer. We utilised a large Nationally representative database of primary care records linked with hospital records and Office for National Statistics (ONS) death registration data in England.

Methods

Data Source

Data came from the Clinical Practise Research Datalink (CPRD) Aurum May 2022 primary care Dataset.²⁵ CPRD Aurum contains anonymised primary care data from more than 60 million patients and can be linked to secondary care and other health-related datasets,²⁶ giving researchers a complete view of data regarding patient care, facilitating the conduct of essential studies to enhance patient care. A limitation for this study was that data on cancer-specific factors (stage, sub-type, treatment) are not included. The dataset was linked with the ONS²⁷ and Index of Multiple Deprivation (IMD) data from 2022.²⁸ ONS data included information on the official date and causes of death (using ICD-10 codes) of each patient, while patient-level IMD, which is a measure of relative deprivation for small areas (Lower Super Output Areas (LSOAs) was used as a proxy for socio-economic position. Linked pseudonymised data was provided for this study by CPRD. Data is linked by NHS Digital, the statutory trusted third party for linking data, using identifiable data held only by NHS Digital. Select general practices (GP) consent to this process at a practice level with individual patients having the right to opt-out.

Study Population

We included people with a lung cancer diagnosis (measured by first mention of lung cancer within their primary care record) between 1st January 2010 and 1st January 2023 using SNOMED-CT codes entered by GPs during consultations.

People were included if they had a code for lung cancer recorded for the first time between 2010 and 2022, they were aged 18 or over at the time of lung cancer diagnosis, were eligible for data linkage with ONS data, were regular (not temporary patients) and had research acceptable data available from CPRD. People with less than 1 year of GP registration at the time of

their first code for lung cancer were not included as this may reflect a historical diagnosis recorded post registration with a new GP Practice. Patients with inconsistent data, for example, date of death prior to date of birth were excluded. The number of people included in the study and reasons for exclusions are shown in [Figure 1](#).

The code lists used for defining the study population, study exposures, outcomes and co-variables are available on Github.²⁹

Exposure Variable

The main exposure in this research work was the prevalence of T2DM at the time of lung cancer diagnosis. Patients having codes categorised as T2DM if they had a code recorded from codelist for prevalent diabetes codes (see codelist available at²⁹) prior to their first code for lung cancer. Diabetes diagnosis after the first lung cancer was not included (whether related to cancer treatment or not).

The codelist for prevalent diabetes was derived by searching the CPRD browser (February 2022) using search terms related to diabetes and remerging with SNOMED Concept IDs to look for additional terms, which may have been missed. All codes included in the case definition were categorised by two coders including one clinician and consensus reached on what to include. Any codes categorised as T2DM, diabetes (type not specified) and diabetes-related (type not specified) and diabetes monitoring were included as prevalent T2DM on the basis that T2DM is the majority of diabetes. Among those classified as T2DM in the study, 83% had at least one code specific to T2DM within their primary care record. Codes which specified diabetes as Type 1 diabetes or something other than type 2, eg gestational diabetes, steroid induced diabetes were not included. As only diabetes diagnosis prior to the cancer diagnosis were considered, steroid-induced diabetes as a result of cancer treatment was not included.

Outcomes

The primary outcome was all-cause mortality following lung cancer diagnosis (as measured by first code within primary care). Survival was calculated from the initial event (date of lung cancer diagnosis defined as first mention in primary care) to the final event (death or censoring (loss to follow-up due to switching to a non-CPRD registered practice or last collection date of the practice)). Patients with an ONS death date were determined as having died, while those without an ONS death date were assumed to be alive.

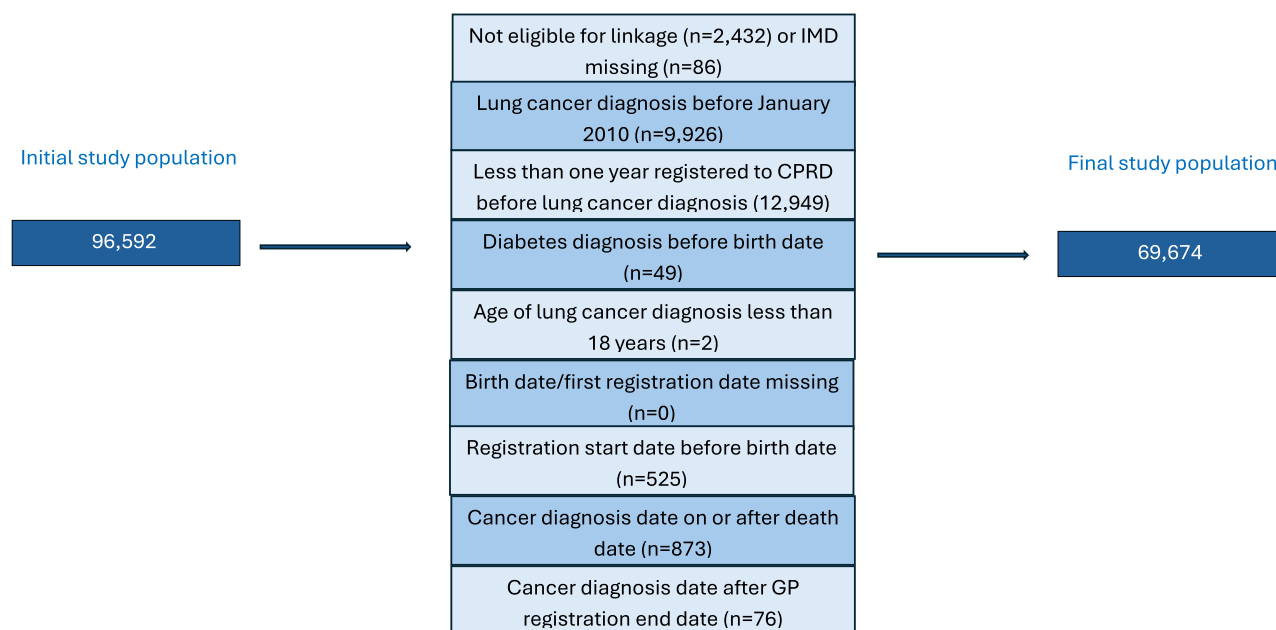


Figure 1 Exclusion criteria for patients included in the analysis, Reasons for ineligibility with the number of participants excluded.

The secondary outcome was cause-specific mortality divided into cardiovascular, cancer-specific (any cancer) and respiratory causes as the underlying cause of death.

Co-Variates

We accounted for several confounding variables that could potentially influence the relationship between lung cancer mortality and T2DM. These variables were selected based on their established associations with both T2DM and lung cancer outcomes and included: age, sex, IMD, calendar year of lung cancer diagnosis, smoking status, and body mass index (BMI).

Smoking status was categorized into three groups: current smokers, ex-smokers and never-smokers.

BMI was used as a categorical variable, we used the following classification categories from the World Health Organisation: Obese (>30.0), Overweight (25.0 – 29.9), Normal (18.5 – 24.9), and Underweight (<18.5). Sensitivity analysis was conducted comparing results using BMI as categorical with using BMI as a continuous variable. Most recent BMI and smoking status record prior to Lung Cancer diagnosis were defined using previously established codelists and algorithms.³⁰ Since BMI and smoking status can change over time and more recent measurements prior to lung cancer diagnosis may be affected by reverse causality a sensitivity analysis was conducted excluding smoking and BMI records in the previous 12 months prior to lung cancer diagnosis.

Among people with T2DM the most recent HbA1c measurement (as mmol/mol) prior to lung cancer diagnosis was recorded as an indicator of diabetes control using categories (≤ 47.5 ; 47.5 – 69.4 ; 69.4 – 85.8 ; >85.8). HbA1c was only used for describing the study sample and was not included as a co-variate in the survival analysis. Codelists used to define co-variables are available within a Github repository.²⁹

Statistical Analysis

Continuous variables were presented using means and standard deviations, whereas categorical measures were reported using numbers and proportions.

We fitted separate Cox Regression Models for each outcome. Start of follow-up was the date of first code for lung cancer recorded within primary care records. End of follow-up was the earliest of the outcome (death or cause-specific death), death from other causes (for cause-specific mortality outcomes), end of GP registration or end of the study (1st January 2023).

All models were adjusted for age and gender. To understand the impact of adjusting for different confounders models were then adjusted for the remaining confounders (BMI, smoking status, year of diagnosis of lung cancer and IMD) one at a time in turn and then all together. The final models were compared to minimally adjusted models restricted to people with no missing data on any confounding variables (complete case analysis). To investigate the impact of missing data for BMI and smoking, age and gender-adjusted models were also fitted on the full data set and analyses adjusting for smoking and BMI individually were compared to models with complete data for those variables (ie, the model including adjustment for age, gender and smoking was compared to a model adjusted for age and gender among people with no missing data for smoking).

We tested the proportional hazards assumption using log–log plots of survival, Kaplan–Meier and predicted survival plots and Schoenfeld residuals. Where proportional hazards assumption was not met, data were stratified by cumulative follow-up time (<6 months; 1 year; more than 1 year).

All analyses were performed using R³¹ and Stata 18.³² R libraries used were “Survival”, “survminer” and ggplot2.

Ethical Declaration

CPRD has NHS Health Research Authority (HRA) Research Ethics Committee (REC) approval to allow the collection and release of anonymised primary care data for observational research [NHS HRA REC reference number: 05/MRE04/87]. Each year CPRD obtains Section 251 regulatory support through the HRA Confidentiality Advisory Group (CAG), to enable patient identifiers, without accompanying clinical data, to flow from CPRD contributing GP practices in England to NHS Digital, for the purposes of data linkage [CAG reference number: 21/CAG/0008]. The protocol for this research was approved by CPRD’s Research Data Governance (RDG) Process (protocol number: 23–002741) and the approved protocol is available upon request. Linked pseudonymised data was provided for this study by CPRD. Data is

linked by NHS Digital, the statutory trusted third party for linking data, using identifiable data held only by NHS Digital. Select general practices consent to this process at a practice level with individual patients having the right to opt-out.

Results

Study Population

The study population consisted of 69,674 eligible patients with lung cancer after applying specific exclusion criteria (Figure 1). Out of the final study population, 20.9% of people had T2DM.

Table 1 shows the baseline characteristics of the cohort by whether people had T2DM or not. People with T2DM were more likely to be male (T2DM: 58.1% vs non-T2DM: 51.9%) and diagnosed with cancer at an older age compared with people without T2DM (T2DM mean 74.4 (SD 9.2) vs non-T2DM: 71.9 SD (10.9)).

Table 1 Baseline Characteristics of Patients with and without Type 2 Diabetes Mellitus (T2DM) (N = 69,674)

Variable	T2DM (%)	No T2DM (%)
N	14,566 (100)	55,108 (100)
Gender		
Male	8455 (58.1)	28,578 (51.9)
Female	6111 (42.0)	26,530 (48.1)
Mean and SD Age of Lung Cancer Diagnosis	74.4 (9.2)	71.9 (10.9)
Median year of lung cancer diagnosis (IQR)	2017 (2013–2019)	2016 (2013–2019)
IMD quintile		
1 (least deprived)	2064 (14.2)	9193 (16.7)
2	2521 (17.3)	10,450 (19.0)
3	2646 (18.2)	10,470 (19.0)
4	3314 (22.8)	11,387 (20.7)
5 (most deprived)	4021 (27.6)	13,608 (24.7)
Smoking Status		
Never	1343 (9.2)	5374 (9.8)
Ex-smoker	4626 (31.8)	15,212 (27.7)
Current Smoker	8592 (59.0)	34,372 (62.5)
Missing ^a	5	150
BMI (Kg/m ²)		
Underweight (<18.5)	440 (3.0)	3376 (6.4)
Normal (18.5 –25)	4017 (27.8)	22,367 (42.6)
Overweight (25 –30)	5111 (35.3)	17,484 (33.3)
Obese (>30)	4897 (33.9)	9304 (17.7)
Missing ^a	101	2577
HbA1c Level (mmol/mol)		
HbA1c ≤ 47.5	5486 (40.6)	25,048 (97.0)
47.5 <HbA1c ≤ 69.4	6565 (48.6)	759 (2.9)
69.4 < HbA1c ≤ 85.8	1054 (7.8)	12 (0.05)
HbA1c > 85.8	407 (3.0)	11 (0.04)
Missing ^a	1054	29,278

Note: ^aMissing data values were excluded from statistical analysis.

Abbreviations: T2DM, Type 2 diabetes mellitus; BMI, Body mass index; SD, standard deviation; IQR, inter-quartile range; IMD, Index of multiple deprivation.

There was a higher proportion of people who were ex-smokers in the T2DM group although this difference was small (T2DM: 31.8% vs non-T2DM: 27.7%). Most people in the study sample, both with and without T2DM, were currently smoking (T2DM: 59.0% vs non-T2DM: 62.5%).

People with T2DM had a higher proportion of people with diabetes classified as BMI>30 (T2DM: 33.9% vs non-T2DM: 17.7%).

During the study period, there were 50,351 deaths (72.2% of the study cohort). The main cause of death was cancer (45,971 deaths; 91.3% of study deaths). The number of deaths due to CVD was 1469 (2.9% of study deaths) and due to respiratory disease was 1446 (2.9% of study deaths).

Follow-up time was 7798.2 person years for people without T2DM and 17,876.6 person years for people with T2DM. Absolute rates of all outcomes were higher in people with T2DM (Table 2).

The hazard ratios for all-cause and cause-specific mortality following Lung cancer diagnosis among people with and without T2DM adjusting for confounders are shown in Table 3.

All-Cause Mortality

After adjusting for age and gender, there was no evidence for a difference in all-cause mortality in people with T2DM compared to people without T2DM (IRR 0.98 95% CI 0.96, 1.01). This fluctuated between remaining null and a small positive effect after adjustment for all confounders and restricting the model to people with no missing data on any adjusted confounders (IRR 1.07 95% CI 1.04, 1.09) (Table 3).

Cancer-Specific Mortality

After adjusting for age and gender, people with T2DM had lower rates of cancer-specific mortality compared to people without T2DM (IRR 0.96 95% CI 0.94, 0.98). This fluctuated between a lower rate, no evidence of a difference and a small positive effect after adjustment for all confounders and restricting the model to people with no missing data on any adjusted confounders (IRR 1.05 95% CI 1.02, 1.07) (Table 3).

Cardiovascular Mortality

After adjusting for age and gender, people with T2DM had higher rates of cardiovascular disease mortality compared with people without T2DM (IRR 1.26 95% CI 1.12, 1.42). This remained consistent after adjusting for different confounders and restricting to people with complete data on all confounders (IRR 1.30 95% CI 1.15, 1.47) (Table 3).

Table 2 Rates of All-Cause and Cause-Specific Mortality Following Lung Cancer Diagnosis in People with and without Known Type 2 Diabetes (T2DM) at Time of Lung Cancer Diagnosis

		Number of Events	Rate (per 1000 Person Years) (95% CI)
All-cause mortality	People with T2DM	10,313	576.9 (565.9, 588.1)
	People without T2DM	40,038	513.4 (508.4, 518.5)
Cancer-specific mortality	People with T2DM	9213	515.4 (505.0, 526.0)
	People without T2DM	36,758	471.4 (466.6, 476.2)
Cardiovascular disease mortality	People with T2DM	376	21.0 (19.0, 23.3)
	People without T2DM	1093	14.0 (13.2, 14.9)
Respiratory disease mortality	People with T2DM	344	19.2 (17.3, 21.4)
	People without T2DM	1102	14.1 (13.3, 15.0)

Abbreviations: T2DM, Type 2 diabetes mellitus; CI, Confidence interval.

Table 3 Hazard Ratios for All-Cause and Cause-Specific Mortality Following Lung Cancer Diagnosis Among People with and without Type 2 Diabetes (T2DM)

Adjusted For:		All- Cause Mortality	Cancer-Specific Mortality	Cardiovascular Mortality	Respiratory Mortality
		IRR (95% CI)	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)
Age + gender		0.98 (0.96, 1.01)	0.96 (0.94, 0.98)	1.26 (1.12, 1.42)	1.19 (1.05, 1.34)
+ calendar year of lung cancer diagnosis		1.04 (1.01, 1.06)	1.01 (0.99, 1.04)	1.33 (1.18, 1.50)	1.25 (1.11, 1.41)
+ IMD		0.98 (0.96, 1.00)	0.95 (0.93, 0.98)	1.24 (1.10, 1.40)	1.16 (1.02, 1.31)
Smoking Status (n=69,519)	Age + gender (no missing data on smoking status)	0.98 (0.96, 1.01)	0.96 (0.94, 0.98)	1.26 (1.12, 1.41)	1.19 (1.05, 1.34)
	+ smoking status	0.99 (0.96, 1.01)	0.96 (0.94, 0.98)	1.26 (1.12, 1.42)	1.19 (1.06, 1.35)
BMI (n=66,996)	Age + gender (no missing data on BMI)	1.00 (0.97, 1.02)	0.97 (0.95, 1.00)	1.26 (1.12, 1.42)	1.17 (1.04, 1.32)
	+BMI	1.03 (1.00, 1.05)	1.00 (0.98, 1.03)	1.25 (1.11, 1.41)	1.27 (1.12, 1.43)
Age + gender (no missing data on all variables) (n=66,968)		1.00 (0.97, 1.02)	0.97 (0.95, 1.00)	1.26 (1.12, 1.42)	1.16 (1.03, 1.32)
+ BMI +smoking status +IMD		1.02 (1.00, 1.04)	0.99 (0.97, 1.02)	1.24 (1.10, 1.40)	1.24 (1.09, 1.40)
+ BMI +smoking status +IMD + calendar year of lung cancer diagnosis		1.07 (1.04, 1.09)	1.05 (1.02, 1.07)	1.30 (1.15, 1.47)	1.30 (1.15, 1.47)

Abbreviations: T2DM, Type 2 diabetes mellitus; BMI, Body mass index; IMD, Index of multiple deprivation; IRR, Incidence rate ratio; CI, confidence interval.

Respiratory-Related Mortality

After adjusting for age and gender, people with T2DM had higher rates of respiratory disease mortality compared with people without T2DM (IRR 1.19 95% CI 1.05, 1.34). This remained consistent after adjusting for confounders and restricting to people with complete data on all confounders (1.30 95% CI 1.15, 1.47) (Table 3).

Proportional Hazards Assumption

There was evidence that the proportional hazards assumption was not met for cardiovascular disease mortality (based on Schoenfeld residuals $p = 0.007$). When the analysis was stratified by cumulative follow-up time after adjusting for measured confounders the increased rate in CVD mortality in people with T2DM was not seen in the first 6 months of follow-up (IRR 0.52 (0.25, 1.06) but there was in first year of follow-up (IRR 1.35 95% CI 1.14, 1.59) (Table 4).

Table 4 Hazard Ratios for CVD-Specific Mortality Following Lung Cancer Diagnosis Among People with and without Type 2 Diabetes (T2DM) Stratified by Cumulative Follow-up Time*

N=66,968	CVD Mortality	
Follow Up Time	Age + Gender	All Measured Confounders**
	IRR (95% CI)	IRR (95% CI)
<6 months	0.54 (0.27, 1.10)	0.52 (0.25, 1.06)
<1 year	1.35 (1.14, 1.59)	1.40 (1.18, 1.66)
Total sample	1.26 (1.12, 1.42)	1.30 (1.15, 1.47)

Notes: *Evidence against proportional hazards assumption from Schoenfeld residuals ($p=0.007$). **Age + gender + BMI +smoking status +IMD + calendar year of lung cancer diagnosis.

Abbreviations: T2DM, Type 2 diabetes mellitus; CVD, cardiovascular disease; BMI, Body mass index; IMD, Index of multiple deprivation; IRR, Incidence rate ratio; CI, confidence interval.

There was no evidence that the proportional hazards assumption was not met for the other three outcomes, therefore stratified results are not presented for these outcomes.

Sensitivity Analysis

Sensitivity analyses excluding records for smoking and BMI, which were in the 12 months prior to lung cancer diagnosis ([Supplementary Table 1](#)) and comparing adjustment for BMI as a categorical variable and a continuous variable ([Supplementary Table 2](#)) produced results consistent with the main analysis.

Discussion

Here, we investigated how prevalent T2DM at the time of lung cancer diagnosis was associated with all-cause and cause-specific mortality. There was strong evidence that people with T2DM had 30% higher CVD and respiratory disease mortality following lung cancer diagnosis. The relationship between T2DM and all-cause and cancer-specific mortality following lung cancer diagnosis was smaller and dependent on confounder adjustment. Differences in studies on approaches to confounding and levels of missing data may contribute to the mixed findings on this association in the literature. The increased risk of respiratory and cardiovascular mortality in those with diabetes may reflect known pathological alterations in type 2 diabetes, including vascular disease, increased inflammation and increased risk of infection. Alternatively, it is possible that this reflects altered treatment of underlying co-morbidities in the face of cancer treatment.²⁴

Comparison with Previous Literature

Our findings are not inconsistent with previous literature, which has shown mixed results with regard to the relationship between diabetes and all-cause mortality following lung cancer diagnosis.^{2,11–17,19,20} Our findings with regard to cause-specific mortality are consistent with existing literature that people with diabetes in general have an increased risk of dying from cardiovascular and respiratory causes.³³

A previous meta-analysis has indicated that the relationship between diabetes and lung cancer survival may vary among various cancer subtypes and treatments, with surgically treated non-small cell lung cancer (NSCLC) showing the most pronounced association between diabetes and overall survival (Hazard Ratio: 1.71, 95% Confidence Interval: 0.94–3.08, $p = 0.07$).³⁴ Here we have not stratified by cancer subtype or treatment and are not able to extrapolate our findings by these factors. However, we have utilised a large, multi-ethnic cohort, and used primary care coding of diabetes to determine cases, and national death certification data to determine cause of death, enabling a comparison of cause-specific mortality. This differs from most prior studies that have examined overall mortality, and our finding of increased respiratory and cardiovascular mortality, but not cancer-specific mortality may have implications of strategies to improve outcomes, suggesting a need to focus on cardiovascular health after lung cancer treatment.

Strengths and Limitations

Here, we utilized a large nationally representative database linked to Office for National Statistics Death data. The study included a large sample size of close to 70,000 patients with lung cancer. The outcome was obtained from Office for National Statistics death records using underlying cause of death for cause-specific mortality analyses, however, there still remains the possibility of misclassification. We also did not account for multi-factorial cause of death using the underlying cause of death only. Misclassification for the exposure is a potential concern in our study,³⁵ as we defined T2DM based on the presence of SNOMED CT codes for diabetes within primary care health records including codes for undefined DM diagnoses. This might have impacted the specificity of our study for T2DM as some patients with other types of diabetes may have been misclassified, potentially leading to biased estimates of association. Efforts to refine the coding criteria for diabetes diagnoses could help address this issue and enhance the accuracy of our results.³⁶ Likewise, it is possible that some individuals with lung cancer do not have this coded on the primary care record, and linkage to other data sources such as Hospital Episode Data or National Cancer Registry data would lead to more complete data.

Our study leveraged the CPRD Aurum database, which utilizes routinely collected health record data not originally collected for research purposes. As a result, measurement and recording errors, as well as clinician coding variability,

may have influenced our findings. For example, confounding factors such as smoking and body mass index were taken from the last recording within primary care records and therefore may not have accurately reflected status at the time of lung cancer diagnosis in all people. Furthermore, only smoking status was available and not data on the amount smoked such as pack-years. Importantly, certain crucial details such as cancer stage, subtype, and details of cancer treatment are not available within primary care data. We also were not able to adjust for all potential confounders, for example, data on air pollution and occupational exposure and environmental exposures were not available, and there may be residual confounding by unmeasured factors. We used the first lung cancer code as a surrogate for the date of cancer diagnosis and the accuracy of the initial diagnosis date might not be uniform across all cases. Primary care records are also not completely sensitive for capturing lung cancer diagnoses with a previous study using CPRD GOLD dating estimating sensitivity at approximately 65%.³⁷ This may have led to a bias if the probability of being recorded within primary care is related to diabetes and mortality, although we are not aware of a reason why this should be the case. Notably, lung cancer is renowned for being diagnosed at an advanced stage,³⁸ which introduces limitations in using the follow-up time from cancer diagnosis as a reliable indicator of cancer stage. It is also important to consider people with T2DM may have a higher probability of having their cancer detected and therefore to be diagnosed at an earlier stage given they are in more frequent contact with health care services,^{39,40} especially for smokers. It is important to acknowledge without accurate data on cancer stage, variations in mortality rates according to cancer stage and subtype could not have been accounted for and residual or unmeasured confounding may be important factors for explaining our results. Further research linking primary care data to other data sources such as cancer registry data, which include data on cancer staging and sub-type would strengthen the evidence base for this question.

As the study conditioned on lung cancer status, collider bias could have impacted our results. However, whilst a number of cancers have been shown to occur at increased frequency in those with diabetes, suggesting an aetiological link, this has not clearly been shown with lung cancer^{41–44} although it is a possibility. In the absence of a clear aetiological link it is therefore less likely that there is collider bias of the association with mortality following lung cancer diagnosis.

Future Work

While this study has shed light on the intricate relationship between diabetes and cause-specific survival outcomes in patients with lung cancer, there remain several avenues for further research.

The current study primarily focused on mortality in all types of lung cancer. Future research could delve into more specific subtypes and stage of lung cancer, as the impact of diabetes on survival might vary depending on these factors. Stratifying the analysis by cancer stage and subtype could uncover nuanced associations that might not be apparent in overall survival analyses. Further research is needed to investigate whether there are disparities in outcomes based on sociodemographic variables such as ethnicity, geographic location, and healthcare access.

Given recent suggestions of metformin's potential impact on survival in diabetic patients with non-small cell lung cancer, further research could explore the effects of specific diabetes medications on survival outcomes and its interaction with biological factors. Investigating potential synergies between diabetes treatments and lung cancer therapies could offer novel therapeutic avenues. Further investigation of diabetes drugs associated with reduced cardiovascular events, such as SGLT-2 inhibitors and GLP-1 analogues will also be of interest.

This study lays the groundwork for future investigations aimed at unravelling the complex relationship between diabetes and cause-specific survival outcomes in patients with lung cancer. In particular, our study findings highlight the need for strategies to improve survival for people with diabetes following cancer diagnosis to include a focus on addressing CVD risk factors given the increased risk of CVD mortality in this group.

Conclusion

Among people with lung cancer identified through their primary care records, cancer-specific mortality was lower in people with prevalent T2DM at the time of their lung cancer diagnosis compared to people without T2DM but this was explained by adjusting for measured confounders. Cardiovascular and respiratory mortality among people with lung cancer was higher in people with prevalent T2DM.

Abbreviations

BMI, Body Mass Index; CAG, Confidentiality Advisory Group; CI, Confidence Interval; CPRD, Clinical Practice Research Datalink; CVD, Cardiovascular Disease; DM, Diabetes Mellitus; GP, General Practice; HES, Hospital Episode Statistics; HRA, Health Research Authority; ICD-10, International Classification of Diseases –10; IMD, Index of multiple deprivation; IQR, Inter-quartile range; IRR, Incidence rate ratio; LSOA, Lower super output areas; NHS, National Health Service; NSCLC, Non-Small Cell Lung Cancer; ONS, Office for National Statistics; OR, Odds ratio; REC, Research Ethics Committee; SCLC, Small cell lung cancer; SD, Standard deviation; T2DM, Type 2 Diabetes.

Data Sharing Statement

Data are available on request from the CPRD. Their provision requires the purchase of a license, and this license does not permit the authors to make them publicly available to all. This work used data from the version collected in May 2022 and have clearly specified the data selected within each Methods section. To allow identical data to be obtained by others, via the purchase of a license, the code lists will be provided upon request. Licenses are available from the CPRD (<http://www.cprd.com>): The Clinical Practice Research Datalink Group, The Medicines and Healthcare products Regulatory Agency, 10 South Colonnade, Canary Wharf, London E14 4PU.

Ethical Approval

CPRD has NHS Health Research Authority (HRA) Research Ethics Committee (REC) approval to allow the collection and release of anonymised primary care data for observational research [NHS HRA REC reference number: 05/MRE04/87]. Each year CPRD obtains Section 251 regulatory support through the HRA Confidentiality Advisory Group (CAG), to enable patient identifiers, without accompanying clinical data, to flow from CPRD contributing GP practices in England to NHS Digital, for the purposes of data linkage [CAG reference number: 21/CAG/0008]. The protocol for this research was approved by CPRD's Research Data Governance (RDG) Process (protocol number: 23_002741) and the approved protocol is available upon request. Linked pseudonymised data was provided for this study by CPRD. Data is linked by NHS Digital, the statutory trusted third party for linking data, using identifiable data held only by NHS Digital. Select general practices consent to this process at a practice level with individual patients having the right to opt-out.

Acknowledgments

This study is based in part on data from the Clinical Practice Research Datalink obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. The data is provided by patients and collected by the NHS as part of their care and support. The interpretation and conclusions contained in this study are those of the author/s alone.

Funding

This research was supported by the NIHR Imperial Biomedical Research Centre (BRC). EI received funding from the HDR UK Health Data Science Black Internship Programme. KB is supported by a Wellcome Senior Research Fellowship (220283/Z/20/Z).

Disclosure

Jennifer Quint reports grants from HDR UK, during the conduct of the study; Sanjay Popat reports personal fees from Anheart Therapeutics, Amgen, AstraZeneca, Bayer, Arcus Biosciences, BMS, Boehringer Ingelheim, Ellipses, EQRx, Daiichi Sankyo, Gilead, GSK, Guardant Health, IO Biotech, Janssen, Lilly, Merck Serono, Mirati, MSD, Novocure, Novartis, Pharmamar, Roche, Takeda, Pfizer, Pierre Fabre, Turning Point Therapeutics, Regeneron, outside the submitted work; Krishnan Bhaskaran reports grants from Wellcome, during the conduct of the study; Daniel Morganstein reports personal fees from BMS and Pfizer, outside the submitted work; Sarah Cook reports grants from NIHR, outside the submitted work. The other authors report no conflicts of interest in this work.

References

- Joharatnam-Hogan N, Chambers P, Dhataria K, Board R. A guideline for the outpatient management of glycaemic control in people with cancer. *Diabet Med.* 2022;39(1):e14636. doi:10.1111/dme.14636
- Björnsdóttir HH, Rawshani A, Rawshani A, et al. A national observation study of cancer incidence and mortality risks in type 2 diabetes compared to the background population over time. *Sci Rep.* 2020;10(1):17376. doi:10.1038/s41598-020-73668-y
- Barone BB, Yeh HC, Snyder CF, et al. Long-term all-cause mortality in cancer patients with preexisting diabetes mellitus: a systematic review and meta-analysis. *JAMA.* 2008;300(23):2754–2764. doi:10.1001/jama.2008.824
- Barua R, Templeton AJ, Seruga B, Ocana A, Amir E, Ethier JL. Hyperglycaemia and survival in solid tumours: a systematic review and meta-analysis. *Clin Oncol.* 2018;30(4):215–224. doi:10.1016/j.clon.2018.01.003
- Tao H, O'Neil A, Choi Y, et al. Pre- and post-diagnosis diabetes as a risk factor for all-cause and cancer-specific mortality in breast, prostate, and colorectal cancer survivors: a prospective cohort study. *Front Endocrinol.* 2020;11:60. doi:10.3389/fendo.2020.00060
- McEwen LN, Kim C, Haan M, et al. Diabetes reporting as a cause of death: results from the translating research into action for diabetes (TRIAD) study. *Diabetes Care.* 2006;29(2):247–253. doi:10.2337/diacare.29.02.06.dc05-0998
- Murray CJ, Dias RH, Kulkarni SC, Lozano R, Stevens GA, Ezzati M. Improving the comparability of diabetes mortality statistics in the U.S. and Mexico. *Diabetes Care.* 2008;31(3):451–458. doi:10.2337/dc07-1370
- Joharatnam-Hogan N, Morganstein DL. Diabetes and cancer: optimising glycaemic control. *J Hum Nutr Diet.* 2023;36(2):504–513. doi:10.1111/jhn.13051
- Kaseda K, Hishida T, Masai K, Asakura K, Hayashi Y, Asamura H. Clinicopathological and prognostic features of operable non-small cell lung cancer patients with diabetes mellitus. *J Surg Oncol.* 2021;123(1):332–341. doi:10.1002/jso.26243
- Nirantharakumar K, Marshall T, Kennedy A, Narendran P, Hemming K, Coleman JJ. Hypoglycaemia is associated with increased length of stay and mortality in people with diabetes who are hospitalized. *Diabet Med.* 2012;29(12):e445–8. doi:10.1111/dme.12002
- Shewale JB, Corsini EM, Correa AM, et al. Time trends and predictors of survival in surgically resected early-stage non-small cell lung cancer patients. *J Surg Oncol.* 2020;122(3):495–505. doi:10.1002/jso.25966
- Jeon HW, Kim KS, Kim YD, et al. Lymphatic vessel invasion in pathologic stage I and II non-small cell lung tumors. *Surg Today.* 2015;45(8):1018–1024. doi:10.1007/s00595-014-1082-2
- Li X, Fang H, Zhang D, et al. Long-term survival analysis of patients with stage IIIB-IV non-small cell lung cancer complicated by type 2 diabetes mellitus: a retrospective propensity score matching analysis. *Thorac Cancer.* 2022;13(23):3268–3273. doi:10.1111/1759-7714.14676
- Zeng X, Xu C, Cheng J, et al. Poor glycemic control might compromise the efficacy of chemotherapy in non-small cell lung cancer patients with diabetes mellitus. *Cancer Med.* 2020;9(3):902–911. doi:10.1002/cam4.2750
- Danila E, Linkevičiūtė-Ulinskienė D, Zablockis R, Gruslys V, Cicėnas S, Smalytė G. A cohort study of exposure to antihyperglycemic therapy and survival in patients with lung cancer. *Int J Environ Res Public Health.* 2020;17(5):1747. doi:10.3390/ijerph17051747
- Wang G, Li X, Xiong R, Wu H, Xu M, Xie M. Long-term survival analysis of patients with non-small cell lung cancer complicated with type 2 diabetes mellitus. *Thorac Cancer.* 2020;11(5):1309–1318. doi:10.1111/1759-7714.13398
- Qian J, Wang W, Wang L, et al. The survival benefit for optimal glycemic control in advanced non-small cell lung cancer patients with preexisting diabetes mellitus. *Front Oncol.* 2021;11:745150. doi:10.3389/fonc.2021.745150
- Tseng C-H. Higher risk of mortality from lung cancer in Taiwanese people with diabetes. *Diabetes Res Clin Pract.* 2013;102(3):193–201. doi:10.1016/j.diabres.2013.10.019
- Hatlen P, Grønberg BH, Langhammer A, Carlsen SM, Amundsen T. Prolonged survival in patients with lung cancer with diabetes mellitus. *J Thorac Oncol.* 2011;6(11):1810–1817. doi:10.1097/JTO.0b013e31822a75be
- De Giorgio R, Barbara G, Ceconi A, Corinaldesi R, Mancini AM. Diabetes Is Associated With Longer Survival Rates in Patients With Malignant Tumors. *Arch Int Med.* 2000;160(14):2217.
- Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. *Ann N Y Acad Sci.* 2013;1281(1):64–91. doi:10.1111/nyas.12098
- Tseng CH. Diabetes but not insulin increases the risk of lung cancer: a Taiwanese population-based study. *PLoS One.* 2014;9(1932–6203 (Electronic)):e0101553. doi:10.1371/journal.pone.0101553
- Li Y, Tian J, Hou T, et al. association between age at diabetes diagnosis and subsequent incidence of cancer: a longitudinal population-based cohort. *Diabetes Care.* 2024;47(3):353–361. doi:10.2337/dc23-0386
- Pearson-Stuttard J, Bennett J, Cheng YJ, et al. Trends in predominant causes of death in individuals with and without diabetes in England from 2001 to 2018: an epidemiological analysis of linked primary care records. (2213-8595 (Electronic)).
- Clinical Practice Research Datalink. CPRD Aurum May 2022 (Version 2022.05.001). Clinical Practice Research Datalink. doi:10.48329/t89s-kf12
- Wolf A, Dedman D, Campbell J, et al. Data resource profile: clinical practice research datalink (CPRD) Aurum. *Int J Epidemiol.* 2019;48(6):1740–g. doi:10.1093/ije/dyz034
- Clinical Practice Research Datalink. CPRD Aurum ONS Deaths January 2022 (Version 2022.01.001). Clinical Practice Research Datalink. doi:10.48329/Q34F-F505
- Clinical Practice Research Datalink. CPRD Aurum Small Area data (Patient) January 2022 (Version 2022.01.001). Clinical Practice Research Datalink. doi:10.48329/AYTT-H222
- GitHub repository for “Mortality outcomes in lung cancer patients with and without type 2 diabetes: a cohort study in England” in England. 2025. Available from: <https://github.com/NHLI-Respiratory-Epi/Lung-cancer-survival-in-people-with-diabetes>.
- Hatam S, Scully ST, Cook S, et al. A Harmonised Approach to Curating Research-Ready Datasets for Asthma, Chronic Obstructive Pulmonary Disease (COPD) and Interstitial Lung Disease (ILD) in England, Wales and Scotland Using Clinical Practice Research Datalink (CPRD), Secure Anonymised Information Linkage (SAIL) Databank and DataLoch. *Clin Epidemiol.* 2024;16:235–247. doi:10.2147/CLEP.S437937
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria; R Foundation for Statistical Computing; 2021.
- StataCorp. *Stata Statistical Software: Release 18*. College Station, TX: StataCorp LLC; 2023.
- Campbell PT, Newton CC, Patel AV, Jacobs EJ, Gapstur SM. Diabetes and cause-specific mortality in a prospective cohort of one million U.S. adults. *Diabetes Care.* 2012;35(9):1835–1844. doi:10.2337/dc12-0002

34. Zhu L, Cao H, Zhang T, et al. The effect of diabetes mellitus on lung cancer prognosis: a PRISMA-compliant meta-analysis of cohort studies. *Medicine*. 2016;95(17):e3528. doi:10.1097/MD.00000000000003528
35. de Lusignan S, Sadek N, Mulnier H, Tahir A, Russell-Jones D, Khunti K. Miscoding, misclassification and misdiagnosis of diabetes in primary care. *Diabetic Med*. 2012;29(2):181–189. doi:10.1111/j.1464-5491.2011.03419.x
36. Harding JL, Shaw JE, Peeters A, Cartensen B, Magliano DJ. Cancer risk among people with type 1 and type 2 diabetes: disentangling true associations, detection bias, and reverse causation. *Diabetes Care*. 2015;38(2):264–270. doi:10.2337/dc14-1996
37. Strongman H, Williams R, Bhaskaran K. What are the implications of using individual and combined sources of routinely collected data to identify and characterise incident site-specific cancers? A concordance and validation study using linked English electronic health records data. *BMJ Open*. 2020;10(8):e037719. doi:10.1136/bmjopen-2020-037719
38. Ellis PM, Vandermeer R. Delays in the diagnosis of lung cancer. *J Thorac Dis*. 2011;3(3):183–188.
39. Hu Y, Zhang X, Ma Y, et al. Incident type 2 diabetes duration and cancer risk: a prospective study in two US cohorts. *JNCI J National Cancer Inst*. 2020;113(4):381–389. doi:10.1093/jnci/djaa141
40. Johnson JA, Bowker SL, Richardson K, Marra CA. Time-varying incidence of cancer after the onset of type 2 diabetes: evidence of potential detection bias. *Diabetologia*. 2011;54(9):2263–2271. doi:10.1007/s00125-011-2242-1
41. Ling S, Brown K, Miksza JK, et al. Association of type 2 diabetes with cancer: a meta-analysis with bias analysis for unmeasured confounding in 151 cohorts comprising 32 million people. *Diabetes Care*. 2020;43(9):2313–2322. doi:10.2337/dc20-0204
42. Hall GC, Roberts CM, Boulis M, Mo J, MacRae KD. Diabetes and the risk of lung cancer. *Diabetes Care*. 2005;28(3):590–594. doi:10.2337/diacare.28.3.590
43. Yi ZH, Luther Y, Xiong GH, et al. Association between diabetes mellitus and lung cancer: meta-analysis. *Eur J Clin Invest*. 2020;50(10):e13332. doi:10.1111/eci.13332
44. Leiter A, Charokopos A, Bailey S, et al. Assessing the association of diabetes with lung cancer risk. *Transl Lung Cancer Res*. 2021;10(11):4200–4208. doi:10.21037/tlcr-21-601

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

Dovepress
Taylor & Francis Group