Received: 3 November 2022 | Revised: 2 May 2023 | Accepted: 4 May 2023

DOI: 10.1002/pds.5637

ORIGINAL ARTICLE

WILEY

Comparison of characteristics of patients with lung cancer in U.K. primary care databases: Clinical Practice Research Datalink Aurum and GOLD

J. L. Gulikers[1,2] | A. J. van Veelen[1,2,3] | J. H. M. Driessen[1,2,3,4] | P. C. Souverein[3] | V. C. G. Tjan-Heijnen[5] | L. E. L. Hendriks[6] | R. M. J. M. van Geel[1,2] | S. Croes[1,2]

[1]Department of Clinical Pharmacy & Toxicology, Maastricht University Medical Centre+, Maastricht, The Netherlands

[2]CARIM School for Cardiovascular Disease, Maastricht University, Maastricht, The Netherlands

[3]Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

[4]NUTRIM, Maastricht University Medical Centre+, Maastricht, The Netherlands

[5]Division Medical Oncology, GROW, Maastricht University Medical Centre+, Maastricht, The Netherlands

[6]Department of Pulmonary Diseases, GROW, Maastricht University Medical Centre+, Maastricht, The Netherlands

Correspondence
S. Croes, Department of Clinical Pharmacy & Toxicology, Maastricht University Medical Centre+, Maastricht, The Netherlands.
Email: s.croes@mumc.nl

Funding information
None

Abstract

Introduction: In recent years, the number of general practices contributing to the Clinical Practice Research Datalink (CPRD) database GOLD is decreasing. Therefore, for research questions addressing for instance novel treatments requiring up-to-date data, sample size will become an important consideration in study feasibility. In recent years, CPRD Aurum, containing information of practices that use EMIS software, has become an additional data source that is being used for CPRD studies. In order to establish whether Aurum is suited to act as data source for future studies in the field of lung cancer research, we aimed to compare characteristics between patients with lung cancer in Aurum and GOLD.

Methods: A retrospective study was performed comparing characteristics and overall survival (OS) of patients with lung cancer in Aurum and GOLD. To further evaluate similarity, hypothetical eligibility of these patients in Aurum and GOLD was compared for 11 randomized clinical trials (RCTs).

Results: Baseline characteristics registered in Aurum and GOLD were largely similar, with some clinically irrelevant differences for previous malignancies, deviant laboratory values and drug use. Median OS was 9.8 and 9.0 months for patients in Aurum and GOLD, respectively. Potential RCT eligibility varied between 49.4% and 79.5% and 49.1% and 78.1% for patients in Aurum and GOLD, respectively. Mortality rates and the comparison of the obtained HRs per hypothetical eligibility cohort per RCT were similar in Aurum and GOLD.

Conclusion: This study showed that data of patients with lung cancer in Aurum and GOLD are largely comparable, suggesting that Aurum is suitable for future epidemiological lung cancer research.

KEYWORDS
CPRD Aurum, CPRD GOLD, database, primary care

Key Points

- A decreasing number of general practices is contributing to Clinical Practice Research Datalink (CPRD) GOLD, while the number contributing practices to the newer CPRD-database Aurum, is increasing.
- In order to use Aurum for future observational lung cancer research instead of GOLD, a mutual comparison is helpful.
- Data of patients with lung cancer from Aurum and GOLD was compared in terms of baseline characteristics, overall survival and hypothetical eligibility to large randomized clinical trials.
- Data from Aurum and GOLD were largely comparable and suggested that Aurum is suitable for future epidemiological research on lung cancer.

1 | INTRODUCTION

The Clinical Practice Research Datalink (CPRD) collects electronic health data from general practitioner (GP) practices around the United Kingdom and is extensively used in observational studies. CPRD GOLD contains information from GP practices located in England, Wales, Scotland, and Northern Ireland from 1987 onwards.[1] It includes primary care data from over 20.8 million patients as of February 2022, with an active patient population of approximately 3.1 million patients (14.9% of the total United Kingdom population).[2] GOLD is considered a well-established database containing data of high quality and is widely used in medical research. However, due to a decreasing number of GPs using Vision software, the number of practices contributing to GOLD is decreasing. In 2013, 674 (8.3%) out of all GP practices in the United Kingdom were contributing to GOLD, but this has declined to 401 GP practices (4.9%) in May 2022.[1,2] Furthermore, the distribution of actively contributing GP practices has also changed over time, as the majority (84.0%) is now located in Scotland and Wales, while only a minority of GP practices (5.7%) is located in England. In 2017, CPRD introduced a new database, CPRD Aurum. Aurum collects data from practices using EMIS software and contains information on GP practices mainly located in England from 1995 onwards. In total, 1358 GP practices are currently contributing to Aurum, which equals 16.6% of all GP practices in the United Kingdom. As of now, Aurum contains records from 40.9 million patients of which 13.4 million patients (32.8%) are currently actively enrolled in a participating practice.[3,4] Data from both Aurum and GOLD can be linked to other databases in order to supplement primary care data with detailed information from hospitals (Hospital Episode Statistics [HES]) or to the National Cancer Registration and Analysis Service (NCRAS) to gain insight in cancer related topics such as tumour diagnosis and anti-cancer treatments.[5–7]

Since a decreasing amount of GP practices is contributing to GOLD, GOLD will become less suitable to use in future observational cohort studies. Clinical research questions addressing novel treatments require up-to-date data and since a decreasing amount of GP practices is contributing to GOLD, the sample size needed for these studies will become an recurring issue. On the other hand, since the number of GP practices contributing to Aurum is increasing, this will be more suited to study novel treatments. However, while there are many years of experience with using CPRD GOLD as a reliable database, with numerous studies reporting on data quality, less is known about the Aurum database. Therefore, evaluating the concordance of data registered in Aurum compared to GOLD in a time period in which GOLD was still in use by many practices will be of added value, before starting to use Aurum as primary study database, or to initiate subsequent lung cancer research with linkage to secondary databases. Since the release of Aurum, a few studies have addressed data similarity between Aurum and GOLD.[8–11] However, a population with a diagnosis of cancer has not been compared yet.

In this light, we evaluated baseline characteristics and overall survival (OS) of patients with lung cancer registered in Aurum and compared them to individuals with lung cancer registered in GOLD. As further attempt to evaluate the level of concordance of both data sets, an earlier performed study with GOLD data, was repeated using Aurum. In this study, the hypothetical eligibility of patients with lung cancer, for 11 selected, previously performed, pivotal randomized clinical trials (RCTs) for systemic therapy (i.e., targeted therapy and immuno-oncology therapy) in lung cancer was assessed.[12] Subsequently, the results of lung cancer populations in Aurum and GOLD were compared, in terms of eligibility percentages and simulated OS of potential eligible patients for those RCTs.

2 | METHODS

2.1 | Data sources

For this study data from both GOLD (release April 2019) and Aurum (release January 2021) was used. GOLD consists of primary care data from GP practices based in the United Kingdom using Vision® software and Aurum consists of primary care data from GP practices based mainly in England using EMIS Web® software. The primary care data include information on demographics, diagnoses, symptoms,

prescriptions, and laboratory tests, among others.[1,3] This study is part of a protocol (#21_000413) approved by the CPRD Independent Scientific Advisory Committee.

2.2 | Study population

All patients, aged 18 years or older with an incident diagnosis of lung cancer between January 1, 2014 and December 31, 2018 were included. The date of lung cancer diagnosis determined the index date. Diagnoses were based on the first registration of lung cancer using Read codes (GOLD) and SNOMED concept IDs (Aurum) for lung cancer (Tables A1 and A2). All types of lung carcinoma were included, since both SNOMED and Read coding systems do not differentiate between different lung carcinomas in terms of type, stage, molecular status or histology of lung cancer. Information regarding whether the date of diagnosis was systematically based on date of biopsy or on imaging is not included in either of the databases and was therefore impossible to retrieve.

2.3 | Data extraction

Data on comorbidities were extracted using code lists consisting of Read (GOLD) and corresponding SNOMED concept IDs (Aurum). In short, GOLD Read codes were cross mapped to SNOMED concept IDs for Aurum and further supplemented based on string searching of medical terms from the original GOLD code list. These lists were independently reviewed by one other investigator. Depending on the comorbidity, different time windows were used to determine presence of the comorbidities prior to the index date (i.e., 30 days, 90 days, 1 year, 5 years, or ever before the index date [Table A3]).

Only drug prescriptions up to 90 days before the index date were included to assess current use.

Laboratory values in GOLD and Aurum are stored in different ways. For GOLD, entities (numerical codes) are used which are linked to specific (laboratory) terms, and for Aurum laboratory values are stored using medcodeIDs which are also used to store diagnosis of morbidities. The laboratory value closest to the index date was used and only if this was registered within 90 days prior to index date (Table A3). Similar to drug prescriptions, a 90-day period prior to the index date was thought to be still representative of the health status of the patient around the time of diagnosis. Cut-off values for deviant laboratory values are specified in Table A4.

2.4 | Overall survival

Patients were followed from the index date until date of last data collection at the GP practice, transfer out of practice, end of study or date of death, whichever came first. Date of death was determined using the EMIS death date or in absence of an EMIS death date, CPRD death date and was determined for patients registered in Aurum. In GOLD, date of death was determined using CPRD death date.

2.5 | Eligibility for clinical trials

Recently, potential eligibility rates for some previously performed lung cancer RCTs or anticancer targeted- and immunotherapies were evaluated for patients with lung cancer registered in GOLD.[12] We aimed to repeat this potential eligibility assessment with the patient cohort in Aurum. In short, eligibility for RCTs was determined using the inclusion and exclusion criteria of 11 selected pivotal phase III RCTs that were published between January 1, 2014 and December 31, 2018 and evaluated systemic anticancer agents for the treatment of non-small cell lung cancer (NSCLC). These trials were chosen to reflect the new treatments that became available during this period. The included RCTs evaluated the tyrosine kinase inhibitors (TKIs) osimertinib (AURA3 and FLAURA) and alectinib (ALEX and ALUR) and the immune checkpoint inhibitors nivolumab (CheckMate 017 and 057), pembrolizumab (KEYNOTE-024, KEYNOTE-189, and KEYNOTE-407), durvalumab (PACIFIC), and atezolizumab (OAK).[13–23] Exclusion criteria of each clinical trial included presence of certain comorbidities, comedication use that could have an interaction with the drug under evaluation or diminish the function of the immune system and deviant laboratory values. Exclusion criteria per RCT are specified in Table A5. Eligibility criteria did not include molecular gene status or disease stage, since this information is not available in our databases. Patients who met all criteria were classified as potentially eligible. Patients who did not meet all criteria were classified as ineligible. The criteria were not applied sequentially and a patient could be classified as non-eligible based on multiple exclusion criteria. For each RCT the proportion of patients registered in Aurum who were eligible for potential study participation was determined, as was described previously for the patients registered in GOLD.[12] Mortality of hypothetically eligible and ineligible patients was then compared for each RCT followed by a comparison of the mortality rates per RCT for Aurum to GOLD. Additionally, the reasons for ineligibility in RCTs were further specified for patients in Aurum.

2.6 | Data analysis

Baseline characteristics for patients with a diagnosis of lung cancer in Aurum and in GOLD were described. For each RCT the proportion of potential eligible patients in Aurum was estimated and descriptively compared to the proportion of eligible patients in GOLD. Furthermore, the median OS (mOS) in Aurum was estimated and compared to GOLD, using Kaplan Meier analysis. Cox regression analysis was used to estimate the age and sex adjusted risk of mortality in Aurum versus GOLD.

Cox regression analysis was used to compare the risk of mortality between eligible and noneligible patients separately for each RCT. The results were adjusted for age and sex. This was done for both

TABLE 1 Baseline characteristics of patients with lung cancer registered in CPRD Aurum and CPRD GOLD.

	Aurum N = 34 831		GOLD N = 9239	
	N	%	N	%
Index date				
2014	8202	23.6	2426	26.3
2015	6440	18.5	2114	22.9
2016	6602	19.0	1795	19.4
2017	6770	19.4	1510	16.3
2018	6817	19.6	1394	15.1
Sex				
No. of males	18 291	52.5	4710	51.0
Age, mean (SD)	72.5 (10.8)		72.1 (10.5)	
≤50 years	1009	2.9	258	2.8
50–64.9 years	7452	21.4	2055	22.2
65–79.9 years	17 923	51.5	4880	52.8
≥80 years	8447	24.3	2046	22.2
BMI (kg/m^2), mean (SD)	26.0 (5.5)		25.8 (5.6)	
≤18.5	2107	6.1	611	6.6
18.5–25	13 198	37.9	3490	37.8
25–30	10 985	31.5	2836	30.7
30–35	4750	13.6	1243	13.5
>35	2014	5.8	531	5.8
Missing	1777	5.1	528	5.7
Smoking status				
Current	13 927	40.0	3462	37.5
Former	18 250	52.4	5106	55.3
Non-smoker	2229	6.4	582	6.3
Missing	425	1.2	89	1.0
Cancer-related				
Previous malignancies[e]	4713	13.5	939	10.2
Immune-related diseases				
Ankylosing spondylitis[b]	91	0.3	21	0.2
Dermatomyositis[b]	20	0.1	5	0.1
Myasthenia gravis[b]	37	0.1	7	0.1
Multiple sclerosis[b]	102	0.3	26	0.3
Polymyalgia rheumatica[b]	813	2.3	189	2.1
Psoriatic arthritis[b]	135	0.4	36	0.4
Rheumatoid arthritis[b]	2516	7.2	283	3.1
Coeliac disease[b]	125	0.4	48	0.5
Crohn's disease[b]	181	0.5	58	0.6
Ulcerative colitis[b]	417	1.2	114	1.2
Grave's disease[b]	59	0.2	21	0.2
Psoriasis[b]	1996	5.7	558	6.0
Sarcoidosis[b]	75	0.2	12	0.1
SLE[b]	62	0.2	19	0.2
Vasculitis[b]	230	0.7	57	0.6

TABLE 1 (Continued)

	Aurum N = 34 831		GOLD N = 9239	
	N	%	N	%
Cardiovascular diseases				
Heart failure[b]	1711	4.9	460	5.0
Heart rhythm disturbances[b]	510	1.5	77	0.8
Myocardial infarction[a]	150	0.4	51	0.6
Poor controlled hypertension[a]	<5	0	<5	0
Unstable angina pectoris[a]	6	0	<5	0
Serious infections				
Meningitis[c]	<5	0	0	0
Pneumonia[c]	248	0.7	118	1.3
Sepsis[c]	35	0.1	18	0.2
Hepatitis[d]	31	0.1	<5	0.0
Psychiatric diseases				
Bipolar disorder[b]	215	0.6	33	0.4
Dementia[b]	1443	4.1	275	3.0
Schizophrenia[b]	331	1.0	64	0.7
Other				
HIV/AIDS[b]	65	0.2	13	0.1
Organ transplant[b]	61	0.2	14	0.1
Substance abuse[e]	11	<0.1	17	0.2
Pregnancy[d]	16	0.1	<5	<0.1
Deviant laboratory values[f]				
Alkaline phosphatase[a]	411	1.2	106	1.2
ALAT[a]	499	1.4	123	1.3
ASAT[a]	68	0.2	34	0.4
eGFR[a]	3255	9.4	969	10.5
Haemoglobin[a]	456	1.3	105	1.1
INR[a]	181	0.5	292	3.0
Neutrophils[a]	57	0.2	14	0.2
Platelets[a]	103	0.3	26	0.3
Total bilirubin[a]	368	1.1	115	1.2
TSH[a]	742	2.1	215	2.3
White blood counts[a]	14	<0.1	4	<0.1
Lymphocyte[a]	159	0.5	35	0.4
Drugs prescriptions				
Systemic corticosteroid[a]	7307	21.0	1903	20.6
Immunosuppressive drugs[a]				
Ciclosporine	15	<0.1	<5	0
Everolimus	0	<0.1	0	0
Sirolimus	0	<0.1	0	0
Tacrolimus	12	<0.1	<5	0
Strong CYP3A4-inhibitors[a]				
Erythromycin	358	1.0	131	1.4
Clarithromycin	2775	8.0	860	9.3
Itraconazole	18	0.1	7	0.1

(Continues)

TABLE 1 (Continued)

	Aurum N = 34 831		GOLD N = 9239	
	N	%	N	%
Ketoconazole	0	0	0	0
Ritonavir	<5	<0.1	0	0
Voriconazole	<5	<0.1	0	0

Abbreviations: %, percentage; AIDS, acquired immunodeficiency syndrome; ALAT, alanine transaminase; ASAT, aspartate transaminase; BMI, body mass index; CYP3A4, cytochrome P450 3A4; HIV/AIDS, human immunodeficiency virus; INR, international normalized ratio; N, number; SD, standard deviation; SLE, systemic lupus erythematosus; TSH, Thyroid stimulating hormone.
[a]Three months prior to index date.
[b]Ever prior to index date.
[c]One month prior to index date.
[d]One year prior to index date.
[e]Five years prior to index date.
[f]As specified in Table A4.

TABLE 2 Eligibility of CPRD Aurum and CPRD GOLD cohort for phase III randomized clinical trials (%).

Name of trial	Drug investigated	Eligible proportion of the Aurum-cohort (%)	Eligible proportion of the GOLD-cohort (%)	Percentage point difference Aurum and GOLD
AURA3	Osimertinib	79.5	78.1	1.4
FLAURA	Osimertinib	71.4	72.4	1.0
ALEX	Alectinib	71.6	73.7	3.1
ALUR	Alectinib	71.5	73.6	2.1
CheckMate 017	Nivolumab	52.4	53.9	1.5
CheckMate 057	Nivolumab	52.7	54.3	1.6
KEYNOTE-024	Pembrolizumab	49.4	49.1	0.3
KEYNOTE-189	Pembrolizumab plus chemotherapy	52.1	52.2	0.1
KEYNOTE-407	Pembrolizumab plus chemotherapy	50.2	50.0	0.2
PACIFIC	Durvalumab	50.7	53.0	2.3
OAK	Atezolizumab	50.9	50.7	0.2

TABLE 3 Proportion of patients in Aurum excluded for each randomized clinical trial by reason of exclusion.

	Reason for exclusion (%)							
Name of trial	A	B	C	D	E	F	G	H
Osimertinib—FLAURA	7.4	–	–	6.0	–	–	8.9	0.3
Osimertinib—AURA3	7.4	10.4	–	6.0	–	–	8.9	0.3
Alectinib—ALEX	5.7	11.7	–	–	–	5.4	8.9	0.5
Alectinib—ALUR	5.7	11.7	–	–	–	5.4	8.9	0.5
Nivolumab—CheckMate 017	6.0	10.4	11.3	–	0.8	5.4	26.2	0.3
Nivolumab—CheckMate 057	5.5	10.4	11.3	–	0.8	5.4	26.2	0.3
Pembrolizumab—KEYNOTE-024	14.7	13.5	11.3	–	0.8	5.4	21.0	0.5
Pembrolizumab—KEYNOTE-189	9.5	13.5	11.3	–	0.8	5.4	21.0	0.5
Pembrolizumab—KEYNOTE-407	13.2	13.5	11.3	–	0.8	5.4	21.0	0.5
Durvalumab—PACIFIC	7.4	13.5	11.3	6.0	0.8	5.4	21.0	0.5
Atezolizumab—OAK	5.4	13.5	11.3	6.4	0.8	–	26.2	0.5

Note: A—Laboratory values; B—Cancer-related; C—Immune-related diseases; D—Cardiovascular diseases; E—Serious infections; F—Psychiatric diseases; G—Concomitant drug use; H—Other.

Aurum and GOLD data, and thereafter results from these databases were compared using a test of interaction.[24] In short, for each RCT a hazard ratio (HR) for mortality was calculated for patients who are hypothetically eligible compared to noneligible patients for a RCT. This HR was calculated for both the Aurum and GOLD cohort. In order to compare the calculated HR for Aurum to the HR for GOLD in for example the AURA3 study, the test of interaction was performed and the relative risk ratio was estimated according to the method described by Altman *et al.*[24] The results are depicted as HR and 95% CI.

2.7 | Sensitivity analysis

Since practices could migrate from Vision to EMIS software during the study period, it is possible that patients are included in both GOLD and Aurum. Therefore, additional Cox regression analysis on the risk of mortality between eligible and noneligible patients for each RCT was performed as sensitivity analysis. In this analysis, patients were excluded from the Aurum dataset if their index date was before the migration date of the practice. In order to select these patients, the data regarding migration of practices was provided by CPRD.

3 | RESULTS

3.1 | Baseline characteristics

Between January 1, 2014 and December 31, 2018, there were 34 831 patients with a diagnosis of lung cancer in Aurum and 9239 patients with a diagnosis of lung cancer in GOLD.

The patients with lung cancer registered in Aurum and in GOLD were largely comparable in terms of demographics, comorbidities and drug use, but some deviations were observed (Table 1). There were more patients with previous malignancies registered in the last 5 years in Aurum (13.5%), compared to GOLD (10.2%). In terms of deviant laboratory values, as specified in Table A4, the percentile difference was largest in deviant international normalized ratio (INR) values. These were more often found in GOLD (3.0%) than in Aurum (0.5%).

3.2 | Eligibility for phase III clinical trials

The largest difference in potential trial eligibility between GOLD and Aurum for studies investigating TKIs was seen for the ALEX trial, where the percentage point difference was 3.1% (Table 2). For the

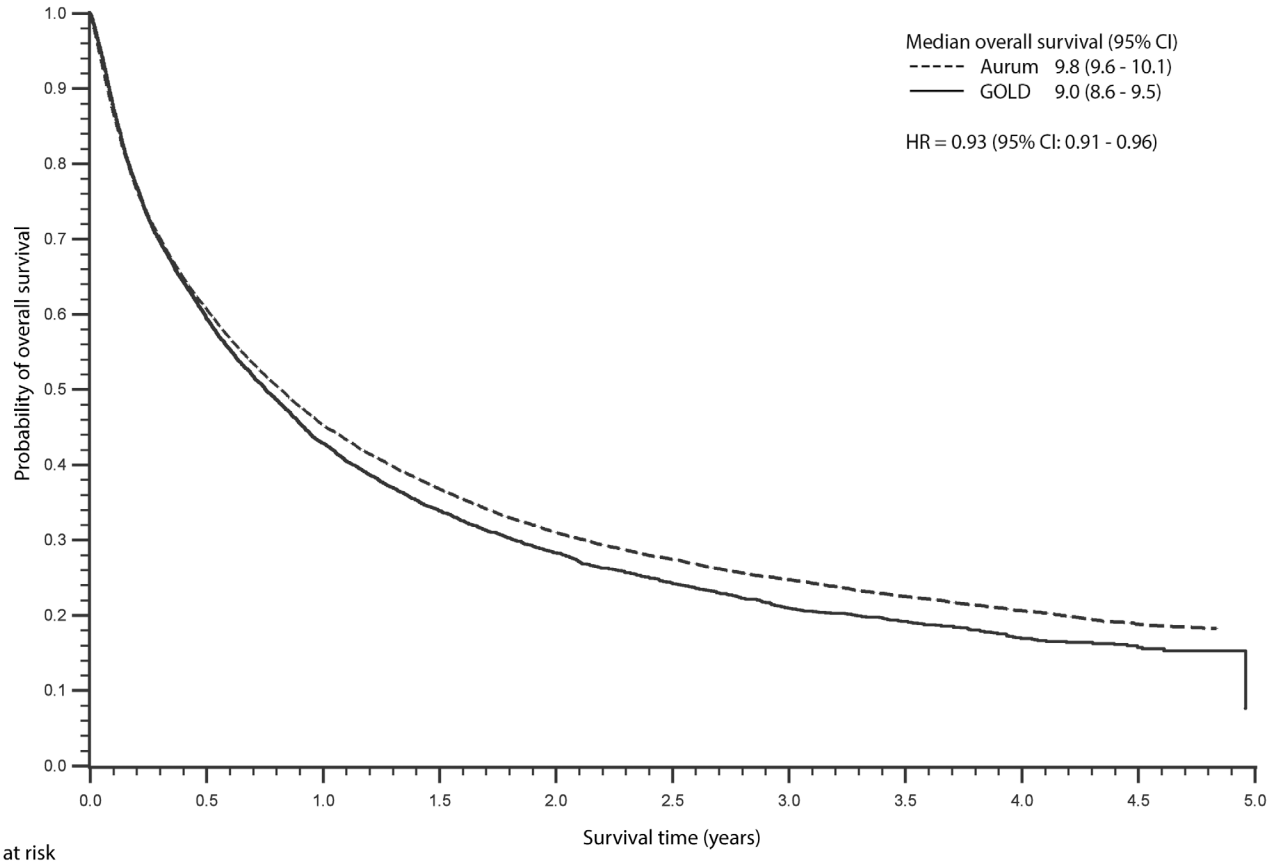

FIGURE 1 Overall survival GOLD cohort (solid line) and Aurum cohort (dotted line).

TABLE 4 Mortality of randomized clinical trial eligible versus ineligible patients and comparison of mortality rates in Aurum to GOLD.

	Aurum				GOLD				
	Incidence rate mortality/1000 PY				Incidence rate mortality/1000 PY				
Name of trial	Eligible	Ineligible	Unadjusted HR (95% CI)	Adjusted HR (95% CI)	Eligible	Ineligible	Unadjusted HR (95% CI)	Adjusted HR (95% CI)	Ratio of mortality rates Aurum vs. GOLD (95% CI)
AURA3	595.2	800.8	0.75 (0.73–0.78)	0.80 (0.78–0.83)	511.4	666.3	0.79 (0.74–0.84)	0.83 (0.78–0.88)	0.97 (0.91–1.04)
FLAURA	593.3	739.8	0.81 (0.79–0.84)	0.85 (0.83–0.87)	516.0	615.3	0.86 (0.81–0.91)	0.89 (0.84–0.94)	0.96 (0.90–1.02)
ALEX	596.3	732.5	0.83 (0.80–0.85)	0.85 (0.82–0.87)	516.1	621.9	0.85 (0.80–0.90)	0.86 (0.81–0.91)	0.98 (0.92–1.05)
ALUR	595.9	733.3	0.82 (0.80–0.85)	0.84 (0.82–0.87)	515.8	622.3	0.85 (0.80–0.90)	0.86 (0.81–0.91)	0.98 (0.92–1.05)
CheckMate 017	544.9	747.3	0.74 (0.72–0.76)	0.75 (0.73–0.77)	478.1	629.5	0.77 (0.73–0.81)	0.77 (0.73–0.81)	0.97 (0.92–1.03)
CheckMate 057	545.6	748.0	0.74 (0.72–0.76)	0.75 (0.73–0.77)	478.2	630.7	0.77 (0.73–0.81)	0.77 (0.73–0.81)	0.97 (0.92–1.03)
KEYNOTE-024	545.8	732.4	0.75 (0.73–0.77)	0.78 (0.76–0.81)	478.7	614.4	0.78 (0.74–0.82)	0.81 (0.77–0.86)	0.97 (0.91–1.02)
KEYNOTE-189	553.4	733.3	0.76 (0.74–0.78)	0.79 (0.77–0.81)	482.3	619.3	0.78 (0.74–0.82)	0.81 (0.76–0.85)	0.98 (0.93–1.04)
KEYNOTE-407	550.2	729.8	0.76 (0.74–0.78)	0.79 (0.77–0.81)	480.6	614.6	0.79 (0.75–0.83)	0.82 (0.78–0.86)	0.97 (0.91–1.03)
PACIFIC	548.4	735.1	0.75 (0.73–0.77)	0.79 (0.77–0.81)	483.6	618.3	0.79 (0.75–0.83)	0.81 (0.77–0.86)	0.97 (0.92–1.03)
OAK	553.8	725.7	0.77 (0.75–0.79)	0.79 (0.77–0.81)	492.0	599.4	0.83 (0.78–0.87)	0.84 (0.80–0.88	0.94 (0.89–0.99)

Abbreviations: CI, confidence interval; HR, hazard ratio; PY, person year.

studies investigating immune checkpoint inhibitors, the largest absolute difference was 2.3% for the PACIFIC trial. In general, a lower proportion of patients was eligible for RCTs investigating immune checkpoint inhibitors (CheckMate, KEYNOTE, PACIFIC, and OAK studies) compared to RCTs with TKIs, but this trend was similar in both databases.

For patients registered in Aurum, the main reasons for ineligibility were previous malignancies and concomitant drug use (Table 3). For RCTs investigating immune checkpoint inhibitors, 21.0%–26.2% of the patients would be ineligible based on concomitant drug use, including corticosteroids.

3.3 | Overall survival Aurum versus GOLD

Median OS of patients with lung cancer registered in Aurum was 9.8 months (95% CI 9.6–10.1) versus 9.0 months (95% CI 8.6–9.5) in GOLD (unadjusted HR = 0.94, 95% CI 0.91–0.97 and adjusted HR = 0.93, 95% CI 0.91–0.96) (Figure 1), meaning that the mOS in Aurum was significantly longer than in GOLD.

3.4 | Mortality of eligible patients versus ineligible patients

Mortality of RCT eligible patients compared to ineligible patients was consistent across all selected RCTs (Table 4). In all investigated RCTs, mortality was lower in the hypothetically eligible patients. For patients in Aurum the age and sex adjusted HR varied between 0.75 (CheckMate 017 and CheckMate 057, 95% CI 0.73–0.77) and 0.85 (FLAURA, 95% CI 0.83–0.87) when comparing mortality in eligible versus ineligible patients and in GOLD this varied between 0.77 (CheckMate 017 and CheckMate 057, 95% CI 0.73–0.81) and 0.89 (FLAURA, 95% CI 0.84–0.94). When comparing the HRs of mortality per RCT from Aurum to GOLD, no differences were found except for the OAK-study. The obtained ratios varied between 0.94 (95% CI 0.89–0.99) and 0.98 (95% CI 0.92–1.05), respectively (Table 4).

3.5 | Sensitivity analysis

In total, 4590 patients were excluded from the Aurum dataset, since they were enrolled in a practice that migrated within the study period. The mortality analysis performed with these patients did show highly similar results to the primary mortality analysis performed per RCT (Table A6).

4 | DISCUSSION

The analysis of baseline characteristics and the eligibility study showed that Aurum and GOLD are largely comparable in terms of demographics, comorbidities and current drug use at the moment of lung cancer diagnosis. Although some differences were found in previous malignancies, psychiatric diseases, and use of (co)medication, these differences were considered not clinically relevant as these percentages were small in general.

The proportion of eligible patients for RCTs was comparable between Aurum and GOLD. In general, a lower proportion of patients was eligible for trials investigating immunotherapy (CheckMate, KEYNOTE, PACIFIC, and OAK), and this finding was similar for both Aurum and GOLD. In these studies, concomitant use of immunosuppressive drugs (including corticosteroids) was prohibited, leading to a larger proportion of ineligible patients compared to RCTs investigating osimertinib and alectinib, in which concurrent use of immunosuppressive drugs was allowed.

The OS in CPRD Aurum was slightly higher compared to GOLD, but when comparing the calculated HRs of mortality per RCT of both databases, no differences were found, indicating large overall concordance between both databases. It is important to mention that comparing mean HRs may have some limitations, since HRs can vary over time and that it may not be collapsible.[25,26] Regarding the latter, the compared HRs were adjusted for age and sex and since these two factors were distributed equally in Aurum and GOLD, collapsibility was considered unlikely to affect the current results.

The minor differences in laboratory values could be due to missing data, since extraction of this information was not registered under a universal number in Aurum, as was the case in GOLD, but had to be done using a manually constructed list. This might have led to an increased amount of hypothetically eligible patients for RCTs, since a patient was only classified as ineligible if a deviant laboratory value was registered and not if this was missing. Furthermore, some laboratory values, such as INR and estimated glomerular filtration rate (eGFR), might be influenced by variations in co-medications and/or differences in daily dosages. For instance, use of coumarin derivatives could not be equally distributed in both datasets. Additionally, an extensive and systematic search is needed to find all registrations linked to one laboratory outcome. Reports describing laboratory values in Aurum are rare and validated methods to extract this data are still missing, as was published previously by Persson *et al.*[27]

To our knowledge, this is the first study to assess data on patients with lung cancer registered in Aurum and to compare this data to patients registered in GOLD. Other comparison studies performed in patients with psoriasis and chronic obstructive pulmonary disease (COPD) and antibiotic use, also did not find substantial differences in the data collected in Aurum in comparison to GOLD.[8–10]

We chose to extract comorbidities and drug prescriptions from Aurum based on Read codes found in GOLD, to ensure that the same terms were used for each extraction. A similar method of converting Read code lists to SNOMED ID code lists has been described before by Gulliford *et al.*[9]

This study also has some limitations. First, the number of potential eligible and noneligible patients could be over- or underestimated due to missing data as was earlier discussed for laboratory values, even though extensive searches were done to minimize this risk. As Trafford *et al.* described, when comparing the two databases,

differences could occur due to differences in the way the databases are built-up and the data are stored. Second, since the eligibility of the patients was tested on the whole lung cancer population registered in GOLD and Aurum, respectively, the reported proportion of hypothetically eligible patients might be different to the actual eligible proportion of patients. We could not differentiate between the major histological subtypes of NSCLC and small cell lung cancer (SCLC), because the subtypes are not registered in these primary care databases. In the United Kingdom, 80%–85% of the patients with lung cancer is diagnosed with NSCLC, therefore we can assume the same percentages are captured in GOLD and Aurum.[28] Third, we noted that approximately 10% of the patients had previous malignancies. In the RCTs investigated, only primary lung cancer cases were eligible for enrolment. With the available information, we were not able to distinguish whether the diagnosed lung cancer was a primary or secondary malignancy. Fourth, we did not have access to information on cancer characteristics such as gene mutation status and stage of the disease. Therefore, patients with other forms of lung cancer could have been wrongfully assigned to either the RCT eligible or to the noneligible group. However, since the above-mentioned information is unavailable in both databases, and the aim of the eligibility substudy was to be an additional uniformity check between lung cancer-related data registered in Aurum and in GOLD, the results from the comparison itself can still be considered valid. Linkage to secondary databases such as the database of NCRAS could prevent misclassification as it contains information on tumour characteristics, tumour stage, and anticancer treatment. Future research is needed to further elaborate on this. Lastly, due to the transition of practices from Vision to EMIS, patients could have been registered in both Aurum and GOLD. However, it was not possible to identify these patients directly, since only data on the practice that migrated was available. We did exclude patients from Aurum that were in a practice that previously used Vision software for GOLD in a sensitivity analysis, but did not find any noticeable differences compared to the results obtained in the complete Aurum lung cancer cohort.

In summary, the uniformity of data, and the completeness of information recorded of patients with lung cancer registered in CPRD Aurum is appropriate and reliable, and similar to the data quality that was retrieved from CPRD GOLD. Therefore, we conclude that the data of patients with lung cancer in Aurum is similar to the data of patients with lung cancer in GOLD. The Aurum database could therefore be considered suitable for future epidemiological research on lung cancer.

CONFLICT OF INTEREST STATEMENT

J. L. Gulikers, A. J. van Veelen, J. H. M. Driessen, P. C. Souverein, R. M. J. M. van Geel, and S. Croes declare no conflict of interest. V. C. G. Tjan-Heijnen: No relationship to disclose in relation to this manuscript. Outside of this manuscript: grants from Novartis, Pfizer, E Lilly, AstraZeneca, Roche, Gilead, Daiichi Sankyo, all outside the submitted manuscript and payed to our institution. L. E. L. Hendriks: No relationship to disclose in relation to this manuscript. Outside of this manuscript: research funding Roche Genentech, Boehringer Ingelheim, AstraZeneca, Takeda, Merck, Pfizer (all institution, Novartis under negotiation); advisory board: BMS, Eli Lilly, Roche Genentech, Pfizer, Takeda, MSD, Boehringer Ingelheim, Amgen, Janssen, Novartis, Merck (all institution, Roche one time self); speaker: MSD, Lilly (institution); travel/conference reimbursement: Roche Genentech (self); mentorship program with key opinion leaders: funded by AstraZeneca; fees for educational webinars: Benecke, Medtalks, VJOncology (self), high5oncology (institution); interview sessions funded by Roche Genentech, Bayer, Lilly (institution); local PI of clinical trials: AstraZeneca, Novartis, BMS, MSD, Merck, GSK, Takeda, Blueprint Medicines, Roche Genentech, Janssen Pharmaceuticals, Mirati.

ETHICS STATEMENT

The study protocol was approved by the CPRD Independent Scientific Advisory Committee. Medical ethical approval was not applicable.

ORCID

J. L. Gulikers https://orcid.org/0000-0002-8844-6992

REFERENCES

1. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44(3):827-836.
2. CPRD GOLD February 2022 (Version 2022.02.001). Clinical Practice Research Datalink. 2022.
3. 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.
4. CPRD Aurum February 2022 (Version 2022.02.001). Clinical Practice Research Datalink. 2022.
5. Bright CJ, Lawton S, Benson S, et al. Data resource profile: the systemic anti-cancer therapy (SACT) dataset. *Int J Epidemiol*. 2020;49(1):15-l.
6. Henson KE, Elliss-Brookes L, Coupland VH, et al. Data resource profile: National Cancer Registration Dataset in England. *Int J Epidemiol*. 2020;49(1):16-h.
7. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data resource profile: hospital episode statistics admitted patient care (HES APC). *Int J Epidemiol*. 2017;46(4):1093-i.
8. Trafford AM, Parisi R, Rutter MK, Kontopantelis E, Griffiths CEM, Ashcroft DM. Concordance and timing in recording cancer events in primary care, hospital and mortality records for patients with and without psoriasis: a population-based cohort study. *PLoS One*. 2021;16(7):e0254661.
9. Gulliford MC, Sun X, Anjuman T, Yelland E, Murray-Thomas T. Comparison of antibiotic prescribing records in two UK primary care electronic health record systems: cohort study using CPRD GOLD and CPRD aurum databases. *BMJ Open*. 2020;10(6):e038767.
10. Requena G, Wolf A, Williams R, et al. Feasibility of using Clinical Practice Research Datalink data to identify patients with chronic obstructive pulmonary disease to enrol into real-world trials. *Pharmacoepidemiol Drug Saf*. 2021;30(4):472-481.
11. Mahadevan P, Harley M, Fordyce S, et al. Completeness and representativeness of small area socioeconomic data linked with the UK Clinical Practice Research Datalink (CPRD). *J Epidemiol Community Health*. 2022;76(10):880-886.
12. van Veelen A, Abtahi S, Souverein P, et al. Characteristics of patients with lung cancer in clinical practice and their potential eligibility for clinical trials evaluating tyrosine kinase inhibitors or immune checkpoint inhibitors. *Cancer Epidemiol*. 2022;78:102149.

13. Mok TS, Wu Y-L, Ahn M-J, et al. Osimertinib or platinum–Pemetrexed in EGFR T790M–positive lung cancer. *N Engl J Med*. 2016;376(7):629-640.
14. Soria J-C, Ohe Y, Vansteenkiste J, et al. Osimertinib in untreated EGFR-mutated advanced non–small-cell lung cancer. *N Engl J Med*. 2017;378(2):113-125.
15. Peters S, Camidge DR, Shaw AT, et al. Alectinib versus Crizotinib in untreated ALK-positive non-small-cell lung cancer. *N Engl J Med*. 2017;377(9):829-838.
16. Novello S, Mazières J, Oh IJ, et al. Alectinib versus chemotherapy in crizotinib-pretreated anaplastic lymphoma kinase (ALK)-positive non-small-cell lung cancer: results from the phase III ALUR study. *Ann Oncol*. 2018;29(6):1409-1416.
17. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med*. 2015;373(2):123-135.
18. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med*. 2015;373(17):1627-1639.
19. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med*. 2018;378(22):2078-2092.
20. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*. 2016;375(19):1823-1833.
21. Paz-Ares L, Luft A, Vicente D, et al. Pembrolizumab plus chemotherapy for squamous non-small-cell lung cancer. *N Engl J Med*. 2018;379(21):2040-2051.
22. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after Chemoradiotherapy in stage III non-small-cell lung cancer. *N Engl J Med*. 2017;377(20):1919-1929.
23. Rittmeyer A, Barlesi F, Waterkamp D, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet*. 2017;389(10066):255-265.
24. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ*. 2003;326(7382):219.
25. Hernán MA. The hazards of Hazard ratios. *Epidemiology*. 2010;21:1-15.
26. Daniel R, Zhang J, Farewell D. Making apples from oranges: comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biom J*. 2021;63(3):528-557.
27. Persson R, Vasilakis-Scaramozza C, Hagberg KW, et al. CPRD aurum database: assessment of data quality and completeness of three important comorbidities. *Pharmacoepidemiol Drug Saf*. 2020;29(11):1456-1464.
28. Types of lung cancer: Cancer Research UK. updated January 28, 2020 [cited 2022 Apr 22]. Available from https://www.cancerresearchuk.org/about-cancer/lung-cancer/stages-types-grades/types

How to cite this article: Gulikers JL, van Veelen AJ, Driessen JHM, et al. Comparison of characteristics of patients with lung cancer in U.K. primary care databases: Clinical Practice Research Datalink Aurum and GOLD. *Pharmacoepidemiol Drug Saf*. 2023;32(10):1161-1177. doi:10.1002/pds.5637

APPENDIX A

TABLE A1 List of Read codes and SNOMED IDs of lung cancer diagnosis in CPRD Aurum.

Med code ID	Read code	SNOMED description ID	Term
4026111000006110		510696018	Primary malignant neoplasm of hilus of lung
155287019	B221100	155287019	Malignant neoplasm of hilus of lung
1773111000006110		510792012	Primary malignant neoplasm of lung
733371000006119	B22z.11	3288586014	Lung cancer
4163281000006110		173925017	Overlapping malignant neoplasm of bronchus and lung
723301000006110	B225.00	1219469018	Malignant neoplasm of overlapping lesion of bronchus and lung
288810010	B220100	288810010	Malignant neoplasm of mucosa of trachea
288813012	B221000	288813012	Malignant neoplasm of carina of bronchus
288819011	B222000	288819011	Malignant neoplasm of upper lobe bronchus
288820017	B222100	288820017	Malignant neoplasm of upper lobe of lung
288822013	B223.00	288822013	Malignant neoplasm of middle lobe, bronchus, or lung
880061000006110	B223.99	880061000006110	Ca middle lobe bronchus/lung
288825010	B223z00	288822013	Malignant neoplasm of middle lobe, bronchus, or lung NOS
288823015	B223000	288823015	Malignant neoplasm of middle lobe bronchus
4748061000006110		3443979013	Malignant neoplasm of right middle lobe of lung
288824014	B223100	288824014	Malignant neoplasm of middle lobe of lung
288826011	B224.00	288826011	Malignant neoplasm of lower lobe, bronchus, or lung
880071000006115	B224.99	880071000006115	Ca lower lobe bronchus/lung

(Continues)

TABLE A1 (Continued)

Med code ID	Read code	SNOMED description ID	Term
288829016	B224z00	288826011	Malignant neoplasm of lower lobe, bronchus or lung NOS
288827019	B224000	288827019	Malignant neoplasm of lower lobe bronchus
288828012	B224100	288828012	Malignant neoplasm of lower lobe of lung
403688010	B222.00	403688010	Malignant neoplasm of upper lobe, bronchus or lung
880051000006113	B222.99	880051000006113	Ca upper lobe bronchus/lung
288821018	B222z00	403688010	Malignant neoplasm of upper lobe, bronchus or lung NOS
11925881000006100		482515017	Malignant tumour of lung
11918131000006100		396221000006112	[X]Malignant neoplasm of bronchus or lung, unspecified
6243241000006110		6243241000006110	Malignant tumour of lung
6243261000006110		1228498010	CA—Lung cancer
288832018	B22y.00	482516016	Malignant neoplasm of other sites of bronchus or lung
403689019	B22z.00	482516016	Malignant neoplasm of bronchus or lung NOS
6245791000006110		482663014	Malignant tumour of trachea
6245821000006110		1228559015	CA—Cancer of trachea
6245811000006110		1228558011	Tracheal cancer
6245831000006110		3289017011	Malignant tracheal tumour
6245801000006110		482662016	Malignant tumour of trachea
6245841000006110		3289020015	Malignant tracheal tumour
721391000006116	B220.00	482662016	Malignant neoplasm of trachea
288811014	B220z00	482662016	Malignant neoplasm of trachea NOS
6363661000006110		1218028010	Ca main bronchus
155361017	B221.00	1210643012	Malignant neoplasm of main bronchus
288815017	B221z00	1210643012	Malignant neoplasm of main bronchus NOS
288808013	B22..00	2765453013	Malignant neoplasm of trachea, bronchus, and lung
880031000006118	B22..98	880031000006118	Ca trachea/bronchus/lung NOS
880041000006111	B22..99	880041000006111	Ca trachea/bronchus/lung

TABLE A2 List of Read codes of lung cancer diagnosis in CPRD GOLD.

Med code	Read code	Term
2587	B22z.11	Lung cancer
3903	B22z.00	Malignant neoplasm of bronchus or lung NOS
17391	B221000	Malignant neoplasm of carina of bronchus
33444	B221100	Malignant neoplasm of hilus of lung
18678	B224000	Malignant neoplasm of lower lobe bronchus
12582	B224100	Malignant neoplasm of lower lobe of lung
42566	B224z00	Malignant neoplasm of lower lobe, bronchus, or lung NOS
12870	B221.00	Malignant neoplasm of main bronchus
21698	B221z00	Malignant neoplasm of main bronchus NOS
41523	B223000	Malignant neoplasm of middle lobe bronchus
39923	B223100	Malignant neoplasm of middle lobe of lung
31268	B223.00	Malignant neoplasm of middle lobe, bronchus, or lung
54134	B223z00	Malignant neoplasm of middle lobe, bronchus, or lung NOS
31188	B224.00	Malignant neoplasm of lower lobe, bronchus, or lung
103946	B220100	Malignant neoplasm of mucosa of trachea
38961	B22y.00	Malignant neoplasm of other sites of bronchus or lung
36371	B225.00	Malignant neoplasm of overlapping lesion of bronchus & lung
15221	B220.00	Malignant neoplasm of trachea
37810	B220z00	Malignant neoplasm of trachea NOS
13243	B22..00	Malignant neoplasm of trachea, bronchus and lung
31700	B222000	Malignant neoplasm of upper lobe bronchus
25886	B222100	Malignant neoplasm of upper lobe of lung
10358	B222.00	Malignant neoplasm of upper lobe, bronchus, or lung
44169	B222z00	Malignant neoplasm of upper lobe, bronchus, or lung NOS

TABLE A3 The subdivision of all in- and exclusion criteria in eight different sets and the corresponding time-window of exposure for each criterion.

Criterion	Time-window of exposure
Laboratory values	
AP	Three months prior to index date
ALAT	Three months prior to index date
ASAT	Three months prior to index date
eGFR	Three months prior to index date
Hemoglobin	Three months prior to index date
INR	Three months prior to index date
Lymphocytes	Three months prior to index date
Neutrophils	Three months prior to index date
White blood cells	Three months prior to index date
Platelets	Three months prior to index date
Total bilirubin	Three months prior to index date
Thyroid stimulation hormone	Three months prior to index date
Cancer related	
History of cancer[a]	Two/three/five years prior to index date[a]
Immune related disease	
Vasculitis	Ever before index date
Coeliac disease	Ever before index date
Crohn's disease	Ever before index date
Ulcerative colitis	Ever before index date
Grave's disease	Ever before index date
Multiple sclerosis	Ever before index date
Myasthenia gravis	Ever before index date
Ankylosing spondylitis	Ever before index date
Dermatomyositis	Ever before index date
Polymyalgia rheumatica	Ever before index date
Psoriatic arthritis	Ever before index date
Rheumatoid arthritis	Ever before index date
Psoriasis	Ever before index date
Sarcoidosis	Ever before index date
Systemic lupus erythematous	Ever before index date
Cardiovascular disease	
Heart failure	Ever before index date
Heart rhythm disturbances[b]	Ever before index date
Myocardial infarction	Three months prior to index date
Poor controlled hypertension	Three months prior to index date
Unstable angina pectoris	Three months prior to index date
Serious infections	
Meningitis	One month prior to index date
Pneumonia	One month prior to index date
Sepsis	One month prior to index date

(Continues)

TABLE A3 (Continued)

Criterion	Time-window of exposure
Hepatitis	One year prior to index date
Psychiatric disease	
Bipolar mood disorder	Ever before index date
Dementia	Ever before index date
Schizophrenia	Ever before index date
Drugs	
Systemic corticosteroid treatment[c]	Three months prior to index date
Immunosuppressive drugs[d]	Three months prior to index date
Strong CYP3A4-inhibitors[e]	Three months prior to index date
Other	
HIV/AIDS	Ever before index date
Organ transplant[f]	Ever before index date
Pregnancy	One year before index date
Substance abuse	Five years before index date

Abbreviations: AIDS, acquired immune deficiency syndrome; ALAT, alanine aminotransferase; AP, alkaline phosphatase; ASAT, aspartate aminotransferase; CYP, cytochrome P450; eGFR, estimated glomerular filtration rate; HIV, human immunodeficiency virus; INR, International normalized ratio.

[a]In the 11 clinical trials different requirements were used for the history of other cancer types, and varied between 2, 3, or 5 years before index date. The specific time period used for each study is shown in Table A5.

[b]For heart rhythm disturbances three specific conditions were used: complete left bundle branch block, second degree heart block and third degree heart block.

[c]For systemic corticosteroid treatment six drugs were included: dexamethasone, hydrocortisone, methylprednisolone, prednisolone, prednisone, triamcinolone.

[d]For immunosuppressive drugs four drugs were included: ciclosporin, everolimus, sirolimus and tacrolimus.

[e]For strong CYP3A4-inhibitors six drugs were included: erythromycin, clarithromycin, itraconazole, ketoconazole, ritonavir and voriconazole.

[f]For organ transplant four specific transplantations were used: heart, lung, kidney, liver.

TABLE A4 Criteria deviant laboratory values.

Laboratory value	Normal value	Criteria deviant value
Alkaline phosphatase	<120 U/L	<2.5 × ULN = <300 U/L
ALAT	<45 U/L (men) < 35 U/L (women)	<3.0 × ULN = 135.0 U/L (men) < 3.0 × ULN = 105.0 U/L (women)
ASAT	<35 U/L (men) < 30 U/L (women)	<3.0 × ULN = 105.0 U/L (men) < 3.0 × ULN = 90.0 U/L (women)
INR	1[a]	>1.5 × ULN
Total bilirubin	3–21 umol/L	<1.5 × ULN = 31.5 umol/L
TSH	0.35–5.00 mU/L	Exceeding normal limits = 0.35–5.00 mU/L

Abbreviations: ALAT, alanine aminotransferase; ASAT, aspartate aminotransferase; INR, international normalized ratio; L, liter; m, milli, 10^{-3}; TSH, thyroid stimulating hormone; u, micro, 10^{-6}; U, unit; ULN, upper limit of normal.

[a]INR level could not be classified as deviant in case of anticoagulant use.

TABLE A5 List of exclusion criteria per randomized clinical trial.

Criterion	AURA3	FLAURA	ALEX	ALUR	CheckMate 017	CheckMate 057	KEYNOTE-024	KEYNOTE-189	KEYNOTE-407	PACIFIC	OAK
	Used cut-off value										
Age	<18 years	<18 years	<18 years	<18 years	<18 years	<18 years	<18 years	<18 years	<18 years	<18 years	<18 years
ALAT (×ULN)	>2.5	>2.5	>3.0	>3.0	>1.5	>1.5	>1.5	>2.5	>2.5	>2.5	>2.5
ASAT (×ULN)	>2.5	>2.5	>3.0	>3.0	>1.5	>1.5	>1.5	>2.5	>2.5	>2.5	>2.5
AP (×ULN)	–	–	–	–	–	–	>2.5	–	–	–	–
eGFR (mL/min)	<50	<50	<45	<45	<40	<40	<60	<50	<60	<50	<30
Haemoglobin (g/L)	<90	<90	<90	<90	<90	<90	<90	<90	<90	<90	<90
Total bilirubin (×ULN)	>1.5	>1.5	–	–	>1	>1.5	>1	>1.5	>1.5	>1.5	>1
White blood cells ($\times 10^9$/L)	–	–	–	–	<2.0	<2.0	–	–	–	–	<2.5
Lymphocyte ($\times 10^9$/L)	–	–	–	–	–	–	–	–	–	–	<0.5
Neutrophils ($\times 10^9$/L)	<1.5	<1.5	<1.5	<1.5	<1.5	<1.5	<1.5	<1.5	<1.5	<1.5	<1.5
Platelets ($\times 10^9$/L)	<100	<100	<100	<100	<100	<100	<100	<100	<100	<100	<100
INR[a] (×ULN)	–	–	–	–	–	–	>1.5	>1.5	>1.5	–	>1.5
TSH	–	–	–	–	–	–	–	Normal	–	–	–
Previous malignancy	–	<2 years	<3 years	<3 years	<2 years	<2 years	<2 years	<5 years	<5 years	<5 years	<5 years
Active serious infection	+	+	–	–	+	+	+	+	+	+	+
Active auto-immune disease	–	–	–	–	+	+	+	+	+	+	+
Heart failure	+	+	–	–	–	–	–	–	–	+	+
Heart rhythm disturbances	+	+	–	–	–	–	–	–	–	+	+
Hepatitis	–	–	+	+	–	–	–	–	–	–	–
History of organ transplant	–	–	+	+	–	–	+	–	–	+	+
Myocardial infarction	–	–	–	–	–	–	–	–	–	+	+
Psychiatric condition	–	–	–	–	+	+	+	+	+	+	+
Pregnancy	+	+	+	+	+	+	+	+	+	+	+
Substance abuse	–	–	–	–	–	–	+	+	+	–	–
Uncontrolled hypertension	+	+	–	–	–	–	–	–	–	–	–
Unstable angina pectoris	–	–	–	–	–	–	–	–	–	+	+
HIV/AIDS	+	+	+	+	+	+	+	+	+	+	+
Systemic treatment with strong CYP3A4-inhibitors	+	+	+	+	+	+	–	–	–	–	+

(Continues)

TABLE A5 (Continued)

Criterion	AURA3	FLAURA	ALEX	ALUR	CheckMate 017	CheckMate 057	KEYNOTE-024	KEYNOTE-189	KEYNOTE-407	PACIFIC	OAK
Systemic treatment with glucocorticoids	−	−	−	−	+	+	+	+	+	+	+
Systemic treatment with immunosuppressive drugs	−	−	−	−	+	+	+	+	+	+	+

Abbreviations: −, not applicable; +, applicable; AIDS, acquired immunodeficiency syndrome; ALAT, alanine aminotransferase; AP, alkaline phosphatase; ASAT, aspartate aminotransferase; CYP, Cytochrome P450; eGFR, estimated glomerular filtration rate; g, gram; HIV, human immunodeficiency viruses; INR, international normalized ratio; L, liter; min, minute; mL, milliliter; TSH, thyroid-stimulating hormone; ULN, upper limit of normal.

[a]Unless current use of anticoagulation.

TABLE A6 Mortality of randomized clinical trial eligible versus ineligible patients and comparison of mortality rates in a selected Aurum cohort to GOLD.

	Aurum				GOLD				
	Incidence rate mortality/1000 PY				Incidence rate mortality/1000 PY				
Name of trial	Eligible	Ineligible	Unadjusted HR (95% CI)	Adjusted HR (95% CI)	Eligible	Ineligible	Unadjusted HR (95% CI)	Adjusted HR (95% CI)	Ratio of mortality rates Aurum vs. GOLD (95% CI)
AURA3	603.2	807.2	0.76 (0.73–0.79)	0.81 (0.78–0.83)	511.4	666.3	0.79 (0.74–0.84)	0.83 (0.78–0.88)	0.97 (0.91–1.04)
FLAURA	600.4	749.5	0.81 (0.79–0.84)	0.85 (0.82–0.88)	516.0	615.3	0.86 (0.81–0.91)	0.89 (0.84–0.94)	0.96 (0.90–1.02)
ALEX	603.1	743.3	0.82 (0.80–0.85)	0.84 (0.82–0.87)	516.1	621.9	0.85 (0.80–0.90)	0.86 (0.81–0.91)	0.98 (0.92–1.05)
ALUR	602.6	744.2	0.82 (0.80–0.85)	0.84 (0.81–0.87)	515.8	622.3	0.85 (0.80–0.90)	0.86 (0.81–0.91)	0.98 (0.92–1.04)
CheckMate 017	550.3	759.4	0.74 (0.72–0.76)	0.74 (0.72–0.77)	478.1	629.5	0.77 (0.73–0.81)	0.77 (0.73–0.81)	0.96 (0.91–1.02)
CheckMate 057	551.2	759.4	0.74 (0.72–0.76)	0.74 (0.72–0.77)	478.2	630.7	0.77 (0.73–0.81)	0.77 (0.73–0.81)	0.96 (0.91–1.02)
KEYNOTE-024	550.1	742.8	0.75 (0.73–0.77)	0.78 (0.76–0.80)	478.7	614.4	0.78 (0.74–0.82)	0.81 (0.77–0.86)	0.96 (0.91–1.02)
KEYNOTE-189	557.9	744.7	0.76 (0.74–0.78)	0.79 (0.76–0.81)	482.3	619.3	0.78 (0.74–0.82)	0.81 (0.76–0.85)	0.98 (0.92–1.04)
KEYNOTE-407	555.0	744.7	0.76 (0.74–0.78)	0.79 (0.77–0.82)	480.6	614.6	0.79 (0.75–0.83)	0.82 (0.78–0.86)	0.97 (0.91–1.03)
PACIFIC	552.4	747.7	0.75 (0.73–0.77)	0.78 (0.76–0.81)	483.6	618.3	0.79 (0.75–0.83)	0.81 (0.77–0.86)	0.97 (0.91–1.02)
OAK	560.7	734.1	0.77 (0.75–0.80)	0.79 (0.77–0.81)	492.0	599.4	0.83 (0.78–0.87)	0.84 (0.80–0.88	0.94 (0.89–1.00)

Abbreviations: CI, confidence interval; HR, hazard ratio; PY, person year.