STUDY PROTOCOL

Trends in lung cancer incidence by sex, age, tumour stage, and histological subtype in Bavaria, 2006-2023, a registry-based study

Principal Investigator: Jacqueline Müller-Nordhorn, MD, DPH, Director, Bavarian Cancer Registry, Germany

Study Coordinators:
Pia Prechtl, Bavarian Cancer Registry, Germany
Sven Voigtländer, Bavarian Cancer Registry, Germany

Cooperation Partners:
Stefanie J. Klug, Chair of Epidemiology, Technical University of Munich, Germany
Amanda Tufman, Comprehensive Pneumology Center (CPC-M) and LMU University Hospital, Munich, Germany
Sylke R. Zeissig, Institute of Clinical Epidemiology and Biometry (ICE-B), Würzburg, Germany and Bavarian Cancer Registry, Germany

Correspondence
Pia Prechtl
pia.prechtl@lgl.bayern.de

ABSTRACT

Background

Lung cancer is the leading cancer in the world. In this study, we aim to describe recent trends of lung cancer incidence in men and women in Bavaria stratified by age, tumour stage, and histological subtype.

Methods and analysis

The study is an observational, retrospective study based on data from the Bavarian Cancer Registry. Incident lung cancer cases are retrieved for the period from 2006 to 2023 to calculate age-standadised incidence rates by sex, age, tumour stage according to the Union of International Cancer Control (UICC), and histological subtype. The analysis uses Joinpoint regression to estimate short- and long-term trends based on annual percent change (APC) and average annual percent change (AAPC), respectively. Missing data on tumour stage and histological subtype will be addressed by multiple imputation.

Ethics and dissemination

This observational study of routine registry data was deemed not to require formal Ethics Committee approval by the Bavarian State Chamber of Physicians (reference number: 2025-1037). The results of the study will be published in the form of scientific articles in peer-reviewed journals.

BACKGROUND

Lung cancer is the leading cause of cancer-related morbidity and mortality in the world. The World Health Organization reported 2.48 million incident lung cancer cases worldwide in 2022.[1] In Germany, approximately 56,600 individuals were diagnosed with lung cancer and 45,200 died due to this cancer in 2022.[2] Lung cancer is typically diagnosed at an advanced stage of disease with more than 50% of annual cases being diagnosed at stage IV according to the Union of International Cancer Control (UICC).[3] Lung cancer incidence trends differ between men and women. In most countries including Germany, incidence rates in men have been constantly declining over the last decades, while rates in women have been increasing.[2,4]

Changes in lung cancer incidence trends are mainly driven by changes in smoking prevalence in the population. In Germany, smoking rates in men have been declining since the late 1970ies, whereas rates in women were increasing until the early 2000s and started to decline afterwards.[5] Considering a lag time of 20 to 30 years until changes in smoking prevalence are reflected in lung cancer incidence,[6] the increasing incidence trend in women is expected to level off in upcoming years. Changes in lung cancer incidence have also been described with respect to histological subtypes.[7] For Germany, however, there is a lack of recent data on lung cancer incidence trends by histological subtype and there are no population-based studies examining long-term incidence trends by tumour stage.

Objective of the study:

The objective of the present study is to determine recent trends of lung cancer incidence in men and women in Bavaria, Germany, stratified by age, tumour stage, and histological subtype.

METHODS

Study design and setting

This study is a retrospective, population-based observational study on lung cancer incidence. It will be conducted using data from the Bavarian Cancer Registry, covering the entire population of Bavaria in the period from 2006 to 2023. Bavaria is the second most populous federal state of Germany with about 13 million inhabitants. Population data will be obtained from mid-year census estimates.

Case definition

Eligible cases include all incident primary lung cancers among residents of Bavaria diagnosed between 2006 and 2023. Incident lung cancer cases are defined by the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10, codes C33-C34). Death Certificate Only (DCO) cases are included to avoid underestimation of incidence. Cases with unknown sex and cases, for which information on tumour staging is only available after neoadjuvant chemotherapy are excluded from the analysis.

Variables and measures

The primary outcome are annual incident lung cancer cases based on ICD-10. Further extracted variables include sex, year of diagnosis, age at diagnosis, TNM classification at time of diagnosis, histological subtype according to the International Classification of Diseases for Oncology, Third Edition (ICD-O-3),[8] and DCO status. Age at diagnosis will be categorised into the age groups "0-49 years", "50-59 years", "60-69 years", "70+ years" for trend analysis. TNM classification includes tumour size, nodal status, presence of distant metastasis, grading, and tumour stage (stages I, II, III, IV) according to the UICC.[9] Histological subtype will be categorised into adenocarcinoma (histological ICD-O-3 codes 8140, 8211, 8230–8231, 8250–8260, 8323, 8480–8490, 8550–8552, 8570–8574, 8576), squamous cell carcinoma (8050–8078, 8083–8084), small-cell carcinoma (8041–8045), large-cell carcinoma (8011–8012, 8014–8031, 8035, 8310), neuroendocrine neoplasms (8013, 8240, 8246, 8249), sarcoma (8800–8811, 8830, 8840-8921, 8890-8991, 9040-9044, 9120-9133, 9150, 9540-9581) and other malignancies (remaining histological codes except 8000-8005, 8010). Not otherwise specified malignant neoplasms (8000-8005) and not otherwise specified carcinoma (8010) will be treated as missing values.

Bias

The study on incidence trends is based on cancer registry data. Since variation in data completeness is a potential source for bias, we limit our study period to calendar years, where completeness of coverage, which is defined as the extent of capturing all incident cancer cases occurring in the registry population, exceeds 90% for lung cancer. Since we also aim to describe incidence trends by tumour stage and histological subtype, missing data may lead to underestimation of incidence, if the missing data mechanism is not "missing completely at random". Missing data on tumour stage and histological subtype has been shown to be associated with higher age and higher comorbidity index.[10] We will address this issue using multiple imputation to avoid exclusion of certain subgroups.

Last, our study period includes years of the COVID-19 pandemic. Studies have shown a significant underreporting of cancer cases during the pandemic.[11] Tumour entities appear to have been impacted to varying degrees, lung cancer incidence seems to have been impacted less compared to other entities such as cancers of the colorectum and the liver.[11] However, in a sensitivity analysis, we will include an additional variable for the COVID-19 pandemic into our imputation model to allow for potential changes of the association between missingness of tumour stage and histological subtype, respectively, and covariates during the pandemic.

Study size

The study size is determined by the total number of lung cancer cases recorded in the cancer registry during the study period. As this is a descriptive population-based study, no formal sample size calculation is required.

Quantitative variables

Age at diagnosis will be used both as quantitative variable for case description and as categorical variable for stratification of incidence rates by age groups. Annual incidence rates will be reported per 100,000 persons.

Statistical analysis

For descriptive analysis, frequencies and percentages of cases will be reported for each categorical variable. For quantitative variables (age at diagnosis), the median and interquartile range will be reported.

Multiple imputation by chained equations[12] will be used to handle missing data on tumour stage and histological subtype. All covariates, which are either associated with the missingness of these variables, or with the variables themselves, will be included into the imputation model. Associations will be assessed using Pearson's X2 test at a significance level of 0.05. The strength of the associations will be furthermore evaluated using Cramer's V statistic. The number of imputed datasets will be determined according to the fraction of missing information.[13] Convergence of the imputation models will be assessed using trace plots. Furthermore, we will compare the distribution before and after imputation to identify and examine any unexpected abnormalities. For sensitivity analysis, we will additionally include a dummy variable for the COVID-19 pandemic into the imputation model.

Age-standardised incidence rates with 95% confidence intervals will be calculated for each calendar year based on the old European Standard Population. Rates will be calculated separately for each imputed dataset and will be pooled according to Rubin's Rule.[14] Incidence rates will be stratified by sex, age group, tumour stage, and histological subtype. Joinpoint regression analysis will be performed to identify any significant joinpoints and to estimate short- and long-term trends (APC and AAPC with respective 95% confidence intervals). Analyses will be carried out using the software package R and Joinpoint Regression Software by the National Cancer Institute.[15]

DISCUSSION

This population-based study will provide recent estimates on lung cancer incidence trends for a study region in Germany, covering more than 13 million inhabitants. It will on the one hand follow up on previous studies regarding incidence trends by sex, and, on the other hand, add valuable information on recent incidence trends by histological subtype and by tumour stage, for which data is particularly scarce. The multiple imputation approach addresses underlying patterns of missingness in the data and helps to avoid underestimation of incidence rates.

Limitations refer to the possibility of late notifications subsequent to the date of data extraction, and to the generalisability of our results. Although Bavaria covers a substantial part of the population of Germany, results cannot simply be extrapolated to overall Germany due to differences in socioeconomic factors and differences in smoking patterns.

ETHICS AND DISSEMINATION

This observational study of routine registry data was deemed not to require formal Ethics Committee approval by the Bavarian State Chamber of Physicians (reference number: 2025-1037). The results of the study will be published in the form of scientific articles in peer-reviewed journals.

FUNDING

This research does not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DECLARATION OF INTERESTS

Nothing to declare.

REFERENCES

1. Ferlay J EM, Lam F, Laversanne M, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. Global Cancer Observatory: Cancer Today. 2024. https://gco.iarc.who.int/media/globocan/factsheets/cancers/40-all-cancers-excl-non-melanoma-skin-cancer-fact-sheet.pdf (accessed 13 February 2025).
2. Zentrum für Krebsregisterdaten im Robert Koch-Institut: Datenbankabfrage mit Schätzung der Inzidenz, Prävalenz und des Überlebens von Krebs in Deutschland auf Basis der epidemiologischen Landeskrebsregisterdaten. Mortalitätsdaten bereitgestellt vom Statistischen Bundesamt. 05 September 2024. www.krebsdaten.de/abfrage (accessed 27 February 2025).
3. Krebs in Deutschland für 2019/2020. Berlin: Robert Koch-Institut, Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V.; 2023. p. 16, 60–3.doi: 10.25646/11357
4. Zhang Y, Luo G, Etxeberria J, Hao Y. Global Patterns and Trends in Lung Cancer Incidence: A Population-Based Study. *Journal of Thoracic Oncology* 2021; **16**(6): 933–44.doi: https://doi.org/10.1016/j.jtho.2021.01.1626
5. Distribution of the population to groups in terms of smoking habits (in percent). Classification: years, Germany, age, sex, smoking habits. 24. Feb 2025. https://www.gbe-bund.de/gbe/!isgbe.indikatoren_set_orientation?p_uid=gast&p_aid=63113905&p_sprache=E&p_help=2&p_indnr=436&p_ansnr=84268222&p_version=4&p_thema_id=123&p_thema_id2=1&p_thema_id3=&p_thema_id4=&D.000=2&D.002=2&D.003=3&D.469=1.
6. Chen J. A Comparative Analysis of Lung Cancer Incidence and Tobacco Consumption in Canada, Norway and Sweden: A Population-Based Study. *Int J Environ Res Public Health* 2023; **20**(20).doi: 10.3390/ijerph20206930
7. Lortet-Tieulent J, Soerjomataram I, Ferlay J, Rutherford M, Weiderpass E, Bray F. International trends in lung cancer incidence by histological subtype: adenocarcinoma stabilizing in men but still increasing in women. *Lung Cancer* 2014; **84**(1): 13–22.doi: 10.1016/j.lungcan.2014.01.009
8. International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3), 2nd revision (ICD-O-3.2). 2024. https://www.who.int/standards/classifications/other-classifications/international-classification-of-diseases-for-oncology (accessed 21 February 2025).
9. TNM Classification of Malignant Tumours, 8th Edition: Wiley-Blackwell; 2016.
10. Robinson AG, Nguyen P, Goldie CL, Jalink M, Hanna TP. Is cancer stage data missing completely at random? A report from a large population-based cohort of non-small cell lung cancer. *Frontiers in Oncology* 2023; **13**.doi: 10.3389/fonc.2023.1146053
11. Voigtländer S, Hakimhashemi A, Grundmann N, et al. Impact of the COVID-19 pandemic on reported cancer diagnoses in Bavaria, Germany. *J Cancer Res Clin Oncol* 2023; **149**(10): 7493–503.doi: 10.1007/s00432-023-04707-0
12. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2011; **45**(3): 1–67.doi: 10.18637/jss.v045.i03
13. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011; **30**(4): 377–99.doi: 10.1002/sim.4067
14. Rubin DB. Multiple Imputation for Nonresponse in Surveys: Wiley; 1987.
15. Joinpoint Regression Software: National Cancer Institute, 2025.