*Temporal trends in lung cancer survival: a population-based cohort study*

Running title: Temporal trends in lung cancer survival

Lukas Löfling1, Shahram Bahmanyar1, Helle Kieler1, Mats Lambe2, 3, Gunnar Wagenius4

1Centre for Pharmacoepidemiology, Department of Medicine – Solna, Karolinska Institutet, Solna, Sweden

2 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Solna, Sweden

3 Regional Cancer Centre Uppsala-Örebro, Uppsala, Sweden

4Cancer Theme, Karolinska University Hospital, Solna, Sweden

## Introduction

Lung cancer survival is low, and it remains the number one cancer-related cause of death in Sweden as well as worldwide (1, 2). The overall 5-year relative survival has increased over time, but still varies between countries, with the highest (33%) in Japan and 10 - 20% in most other countries (3).

Survival after a lung cancer diagnosis has been studied in Sweden in recent years, with reported increased survival over time (3-5). However, these studies reported overall results only and lacked estimates for subgroups of important prognostic factor, e.g., sex, histopathology, and cancer stage. Over the past decades, there have been improvements regarding diagnostic procedures for making a more accurate staging (e.g., positron emission tomography in combination with computed tomography [PET-CT] and endobronchial ultrasound [EBUS]) and new treatments (e.g., target therapies and immunotherapies) have been introduced on the market (6-8). These improvements may have affected survival differently for different groups of patients with lung cancer.

To the best of our knowledge, no study has previously estimated temporal trends in survival for Swedish patients diagnosed with lung cancer over the last 20 years, overall and in subgroups by important prognostic factors such as sex, cancer stage, and histopathology.

This study aims to delineate temporal trends in lung cancer survival overall, and by important prognostic factors, by the use of routinely collected health care data in a universal healthcare setting.

## Methods and Materials

### Data Source and study population

A population-based cohort study will be performed with data retrieved from the Swedish National Lung Cancer Register and the Cause of Death Register, and the Total Population Register (9, 10).

We will identify patients diagnosed with lung cancer (International Classification of Disease 7th revision [ICD-7] code: 1621 or ICD Oncology [ICD-O] 3rd revision code: C34) in Sweden between 1995 and 2016, as recorded in the NLCR. Between 1995 and 2001, this was a regional register covering the Uppsala-Örebro healthcare region (approximately two million inhabitants). Patients with adenocarcinoma or squamous cell carcinoma histopathology will be included.

### Variables

Information on the calendar year of diagnosis, sex, age at diagnosis, histopathology, stage, and smoking history will be acquired from the Swedish National Lung Cancer Register and are based on data from the time of diagnosis to the multi disciplinary conference. Year of diagnosis will be between 1995 and 2016. Age (years) at diagnosis will be grouped as <50, 50-59, 60-69, 70-79, ≥80. Histopathology will be categorised as small cell lung cancer, adenocarcinoma, squamous cell carcinoma. Cancer stage (I-IV) at diagnosis is based on the tumour-node-metastasis (TNM) classification system by the American Joint Committee on Cancer, and we will be using the 4th edition from 1995, the 5th edition from 1997, the 6th revision from 2002, and the 7th revision from 2010 (11-14). Smoking history is based on self-reported information from the patient to the physcician and is categorised as never-smokers (never smoked on a regular basis), former smoker (stoped smoking > 1 year before the diagnosis), and current smokers (smoked at the time of diagnosis or stoed smoking ≤ 1 year before diagnosis). Date of death and emigraion will be obtained from the Swedish Population Register (15).

### Statistical Methods and Analysis

Descriptive analysis will be used to describe clinical characteristics and demographics and summarised with counts and per cent, or median and first quartile (q1) and third quartile (q3).

Relative survival, comparing the all-cause mortality in the lung cancer cohort to the all-cause mortality in the general population (16, 17), and 95% confidence intervals (CI) will be estimated over calendar period overall and by subgroups (sex, stage, and histopathology) at 1-, 2- and 5-years post-diagnosis. Survival for patients with lung cancer will be counted from the date of the lung cancer diagnosis until the date of death, emigration, or administrative censoring (31 December 2016), whichever occurs first. The expected survival will be derived from the general population using the life table approach and calculated according to the *Ederer II* estimator (18, 19). The life tables will be stratified by the calendar year, age, and sex. The study population will include patients diagnosed with lung cancer between 1995 and 2016, with follow-up until 31 December 2016. To estimate the 2- and 5-year survival using the cohort approach, all patients must have a potential follow-up of at least two and five years, respectivelly (18). Hence, the cohort approach will be used to estimate the 1-year relative survival for all patients, for estimating the 2-year relative surviva for patients diagnosed ≤2014, and for estimating 5-year relative survival for patients diagnosed ≤2011. In the most recent years, it will not be possible to estimate 2- and 5-year relative survival using the cohort approach. Instead, the period approach will be used to estimate the 2-year survival for patients diagnsosed in 2015 and 2016, and for estimating the 5-year relative survival for patients diagnosed between 2012 and 2016. In the period approach, all observations included in the analysis are left-truncated at the beginning and right-truncated at the end. The relative survival estimates will be age-standardised according to the International Cancer Survival Standards number one (18, 20). Excess mortality will be modelled using Poisson regression models adjusted for age and sex, with the first year of diagnosis (1995-2000) as the reference period (18). The *strs* Stata command will be used to estimate relative survival and to model excess mortality (18).

Stata version 15.1 (21) and R statistical packages version 3.5.3 (22), or later versions, will be used for all data management and statistical analysis.

#### Sensitivity analyses

Sensitivity analyses will be carried out: to estimate the relative survival when excluding patients in the lung cancer cohort with a history of smoking.

### Tables and figures

Table 1. Baseline characteristics of patients diagnosed with lung cancer, Swedish National Lung Cancer Register, 1995-2016

Table. 2-year excess mortality ratio over calendar period overall and by sex, histopathology and stage for patients diagnosed with lung cancer between 1995 and 2016 as recorded in the Swedish National Lung Cancer Register

Figure 1. Age-standardised 1-, 2- and 5-year relative survival estimates over calendar period overall and by sex, for patients diagnosed with lung cancer between 1995 and 2016 as recorded in the Swedish National Lung Cancer Register

Figure 2. Age-standardised 1-, 2- and 5-year relative survival estimates over calendar period by histopatholgy and stage, for patients diagnosed with lung cancer between 1995 and 2016 as recorded in the Swedish National Lung Cancer Register

Figure 3. Age-standardised 1-, 2- and 5-year relative survival estimates over calendar period by histopathology further subgrouped by cancer stage, for patients diagnosed with lung cancer between 1995 and 2016 as recorded in the Swedish National Lung Cancer Register

#### Supplementary figures

Figure 4. Un-standardised 1-, 2- and 5-year relative survival estimates over calendar period overall and by sex, for patients diagnosed with lung cancer between 1995 and 2016 as recorded in the Swedish National Lung Cancer Register

Figure 5. Age-standardised 1-, 2- and 5-year relative survival estimates over calendar period overall and by sex, for patients diagnosed with lung cancer between 1995 and 2016 as recorded in the Swedish National Lung Cancer Register, excluding patients in the lung cancer cohort with a history of smoking

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018;68(6):394-424.

2. Socialstyrelsen - Swedish National Board of Health and Welfare. Statistics on cancer incidence 2018. Stockholm, Sweden; 2019.

3. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. The Lancet. 2018;391(10125):1023-75.

4. Allemani C, Weir HK, Carreira H, Harewood R, Spika D, Wang XS, et al. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). Lancet (London, England). 2015;385(9972):977-1010.

5. Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. The Lancet. 2011;377(9760):127-38.

6. Regional Cancer Centre - Uppsala Örebro. Lungcancer - Lungcancerrapport för diagnosår 2012-2016. Uppsala, Sweden; 2017.

7. European Medicines Agency. Iressa: EPAR - Product information. London, the United Kingdom; 2014.

8. European Medicines Agency. Tarceva: EPAR - Product information. London, the United Kingdom; 2010.

9. Regional Cancer Centre - Uppsala Örebro. Swedish National Lung Cancer Register 2017 [Available from: <http://www.cancercentrum.se/vast/cancerdiagnoser/lunga-och-lungsack/kvalitetsregister/>.

10. Socialstyrelsen - Swedish National Board of Health and Welfare. The Swedish Cause of Death Register: Socialstyrelsen -Swedish National Bord of Health and Welfare; 2019 [updated 20191018. Available from: <https://www.socialstyrelsen.se/statistik-och-data/register/alla-register/dodsorsaksregistret/>.

11. American Joint Committee on Cancer. AJCC cancer staging manual: 4th edition. 4th ed. Philadelphia, USA: Lippincott; 1992.

12. American Joint Committee on Cancer. AJCC cancer staging manual: 5th edition. 5 ed. Philadelphia, USA: Lippincott-Raven; 1997.

13. American Joint Committee on Cancer. AJCC cancer staging manual: 6th edition. New York, USA: Springer Verlag New York Inc; 2002.

14. American Joint Committee on Cancer. AJCC cancer staging manual: 7th edition. 7th ed. New York, USA: Springer Verlag New York Inc; 2010.

15. The Swedish Tax Agency. The Swedish Population Register 2020 [Available from: <https://www.skatteverket.se/privat/folkbokforing/attvarafolkbokford/folkbokforingsdatabasen.4.3810a01c150939e893f16fe2.html>.

16. Dickman PW, Adami HO. Interpreting trends in cancer patient survival. Journal of internal medicine. 2006;260(2):103-17.

17. Berkson J. The calculation of survival rates. In: Walters W GH, Priestly J,, editor. Carcinoma and other malignant lesions of the stomach Philadelphia, USA: Sanders; 1942. p. 467-84.

18. Dickman PW, Coviello E. Estimating and Modeling Relative Survival. The Stata Journal. 2015;15(1):186-215.

19. Cho H, Howlander N, Mariotto A, Cronin K. Surveillance Research Program, NCI, Technical Report #2011-01 Rockville, MD, USA: National Cancer Institute,, Science DoCCaP; 2011.

20. Association of the Nordic Cancer Registers (NORDCAN). Glossary of statistical terms Lyon, France: International Agency for Research on Cancer (IARC); 2011 [Available from: <http://www-dep.iarc.fr/NORDCAN/english/glossary.htm>.

21. StataCorp. Stata statistical software. 15.1 ed. College Station, Texas, United States of America: StataCorp,; 2017.

22. R Development Core Team. R: A language and enviroment for statistical computing. 3.5.3 ed. Vienna, Austria: R Foundation for statistical computing; 2019.