Study protocol

Utilizing OMOP-Harmonised Hospital Network-Wide EHR Data to Study Association Between Obesity and Sex-related Survival Difference in Lung Cancer Across a Federated Network of Observational Data - an iCAN SHARE study

Table of Contents

Abbreviations

BMI	Body Mass Index
CCI	Charlson Comorbidity Index
DQD	Data Quality Dashboard
DT	Diagnosis to Treatment Interval
ECOG	Eastern Cooperative Oncology Group
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
ESMO	European Society for Medical Oncology
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IQR	Interquartile Range
ICI	Immune Checkpoint Inhibitors
KPS	Karnofsky Performance Status
LoT	Line of Treatment
mNSCLC	Metastatic Non-Small Cell Lung Cancer
NSCLC	Non-Small Cell Lung Cancer
OHDSI	Observational Health Data Sciences and Informatics
OS	Overall Survival
PD1	Programmed Cell Death Protein 1
PD-L1	Programmed Cell Death Ligand 1
PS	Performance Status
mAb	Monoclonal Antibody
RWD	Real World Data
SD	Standard Deviation
TFI	Treatment-Free Interval
TTNT	Time to Next Treatment
TTD	Time to Treatment Discontinuation
TKI	Tyrosine Kinase Inhibitors
VEGF-A	Vascular Endothelial Growth Factor A

Abstract

Background and significance

In our single hospital analysis we have previously identified that survival discrepancy between male and female patients is smaller among high BMI patients (BMI>25) compared to normal BMI (BMI 18.5-25). We have used Observational Medical Outcomes Partnership (OMOP) Common Data Mode in the analysis. In this study we aim to perform external validation of the results by re/running utilizing federated analysis method.

Study aims

To describe among normal BMl and high BMI patients

- Stage at diagnosis
- Histology
- Smoking status
- Age
- Sex
- First line therapy

To compare overall survival in normal BMI and high BMI patients stratified by sex (and histology if patient numbers are large enough)

To create a regression model to estimate the effect of previously listed features on survival

Study description

This retrospective cohort study uses a federated network of observational healthcare data, standardized to the OMOP CDM.

Population

Adult patients with lung cancer patients (≥18 years) diagnosed from January 1, 2015, to the most recent available data.

Outcomes

Overall survival is the outcome metric.

Design

This study is an observational retrospective cohort study. Findings will inform clinical decision-making, support regulatory and policy development and highlight disparities in global treatment practices. The federated model ensures secure, reproducible and scalable insights, setting a precedent for future collaborative oncology research.

Background and Rationale

Lung cancer is causing 19% of all cancer deaths in the world with a yearly incidence of 2,5 million cases worldwide (1) . Better survival outcomes in female over male lung cancer patients have been noted already in the end of the 1990s (2) . Multiple studies have reported that controlling for clinical covariates such as stage, treatment and age does not remove the association between sex and survival, concluding that sex is an independent prognostic factor in lung cancer (3–6) . However, some studies were not able to support these results stating that controlling for covariates eliminates or significantly reduces the survival difference between male and female patients (7,8,9) . It was further reported that the presence and magnitude of the survival difference between male and female patients depends on the analyzed cohort (for example difference among adenocarcinoma patients being larger than in other histology cohorts (10)). Survival discrepancies may also depend on the geographical location due to different population parameters and treatment practices (11).

In our single hospital analysis we have previously identified that survival discrepancy between male and female patients is smaller among high BMI patients (BMI>25) compared to normal BMI (BMI 18.5-25). We have used Observational Medical Outcomes Partnership (OMOP) Common Data Mode in the analysis. In this study we aim to perform external validation of the results by re/running utilizing federated analysis method.

Research Question and Objectives

The overarching aim of this study is to characterize patients within BMI cohorts and compare the survival of male and female patients within these cohorts

1. To describe among normal BMl and high BMI patients
    a. Stage at diagnosis
    b. Histology
    c. Smoking status
    d. Age
    e. Sex
    f. First line therapy
2. To compare overall survival in normal BMl and high BMI patients stratified by sex (and histology if patient numbers are large enough)
3. To create a regression model to estimate the effect of previously listed features on survival

Research Methods

Study Design

This is a retrospective cohort study of patients diagnosed with lung cancer across a network of observational healthcare databases, all standardized to OMOP CDM.

Databases are owned, managed and standardized locally by the different institutions (Data Partners). The study code is developed for all Data Partners to execute locally and the results are aggregated and analyzed.

Study Setting and Data Sources

The selection of Data Partners for this study is based on their availability of patients with mNSCLC as well as cancer treatments and date of death. All databases are mapped to the OMOP CDM (CDM Specification GitHub link). A description of the participating data sources is provided in Table 1.

Table 1: Databases participating in this study at the moment of protocol submission

Data source	Country	Population size	Data capture process and short description	Data capture timeframe
Helsinki University Hospital	Finland	~2.000.000	Structural data pulled from operational EHR systems. Covers secondary and tertiary care.	2012-present
UZA	Belgium	>2.100.000	Routinely collected structured data of patients visiting UZA tertiary hospital for in- or outpatient care	2006-present (depending on data source)
Oslo University Hospital Comprehensive Cancer Centre	Norway			
Leeds Teaching Hospitals NHS Trust	United Kingdom	85,143	Data has been sourced directly from the hospital's EHR system. The LTHT OMOP database includes data from all patients diagnosed with cancer at LTHT from 2010 onwards.	2010-August 2024

Study Period

The study period starts from 1 January 2015 and ends at the latest available date for all data sources. The identification period is from the date of database inception to six months prior to the latest available data in each database. This six-month buffer allows for a potential of 6 months of follow-up data for the last person included in the study.

Study Population

All patients in a data source meeting cohort inclusion criteria are considered subjects. Broadly, these cohorts consist of adult patients with a diagnosis code for lung cancer who have at least 30 days of prior observation (unless defined otherwise, see below). The specific definitions for cohorts are provided in the following sections.

Follow up

Patients are followed until death or end of study period, whichever occurs first.

Variables

Outcomes

The following outcomes are assessed in the study:

- Overall Survival (OS): Length of time from the index date for each cohort to the date of death, or loss to follow up or end of the study. Patients are censored at last recorded clinical activity within the database or end of follow up.

Covariates

The following characteristics are determined **prior to or at index date**.

Demographics

- Age
  - age groups
  - age described as a continuous variable
- Biological sex
- Smoking status
- Weight, height, body mass index (BMI)
- Histology (based on diagnosis)

Study size

This study uses routinely collected data, all patients meeting the eligibility criteria above are included. No formal sample size and power calculation is conducted.

Data Management

Source data at the participating institutions are converted to the OMOP CDM through an Extract, Transform, Load (ETL) process. It includes source code to standardized concept mapping while maintaining quality and context of information (15). The data are refreshed periodically responding to updates by the data source. There are six key standardized domains: condition, drug, measurement, procedure, visit and observation. Data Partners are responsible for their own validation to ensure that the OMOP database instance matches the information provided in the original source. Each Partner is responsible for following their own local data permits, processing and disclosure publication regulations and standard operating procedures.

This study follows relevant ENCePP guidelines and International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guidelines for data management (ENCePP) (5). Standardization of the databases to the OMOP CDM enables the use of

standardized analytics and tools across the network since the structure of the data and the terminology system is harmonized. The OMOP CDM is developed and maintained by OHDSI and is described in detail on the wiki page of the CDM (CDM Specification GitHub link) and in The Book of OHDSI (16).

Analysis

General

Data partners execute the analytics package against their OMOP CDM instance and review and approve the aggregated results before sharing them with the study team. No patient level data are shared.

A diagnostic package is used to assess fitness of use and evaluate measurement error in the phenotype development and evaluation process. The full study package is only executed on databases that pass data quality assessment and cohort diagnostics.

All analyses are reported by database unless the resulting patient size falls below a given threshold as set by the Data Partner (minimal cell size).

Protection of Human Subjects

This study is conducted in accordance with the International Society of Pharmacoepidemiology (ISPE) Guidelines for Good Pharmacoepidemiology Practices (GPP) and EMA, European Network of Centers for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Guide on Methodological Standards in Pharmacoepidemiology (18, 19). The use of the OMOP CDM and OHDSI tools enable the federated analysis of these different databases without accessing patient-level data outside the participating institutions. The study is conducted in compliance with all applicable data protection, security and privacy laws, rules and regulations with respect to the collection, production, use, processing, storage, transfer, modification, deletion and disclosure of any information related to this study. Each data partner is required to provide a statement about IRB approval or exemption to participate.

Confidentiality of patient records is maintained at all times. Data partners remain in full control of executing the analysis and packaging results. There are no transmission of patient-level data at any time during these analyses. Only aggregate statistics are shared. Study packages contain minimum cell count parameters to obscure any cells which fall below allowable reportable limits.

Management and Reporting of Adverse Events/Reactions

According to the new guidelines for good pharmacovigilance practice (EMA/813938/2011 Rev 3), there is no requirement for expedited reporting of adverse drug reactions from studies with secondary use of data (19).

Dissemination and Communication of Study Results

Results of this study will be published following guidelines, including those for authorship, established by the International Committee of Medical Journal Editors (20). When reporting results of this study, the appropriate Strengthening the Reporting of Observational (STROBE) Studies in Epidemiology checklist and ESMO Guidance for Reporting Oncology real-world evidence studies, will be followed (21, 22).

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer J Clin. 2024;74(3):229–63.
2. Ferguson MK, Wang J, Hoffman PC, Haraf DJ, Olak J, Masters GA, et al. Sex-associated differences in survival of patients undergoing resection for lung cancer. Ann Thorac Surg. 2000;69(1):245–9.
3. Mincuzzi A, Carone S, Galluzzo C, Tanzarella M, Lagravinese GM, Bruni A, et al. Gender differences, environmental pressures, tumor characteristics, and death rate in a lung cancer cohort: a seven-years Bayesian survival analysis using cancer registry data from a contaminated area in Italy. Front Public Heal. 2024;11:1278416.
4. Guerreiro T, Forjaz G, Antunes L, Bastos J, Mayer A, Aguiar P, et al. Lung cancer survival and sex-specific patterns in Portugal: A population-based analysis. Pulmonology. 2023;29:S70–9.
5. Visbal AL, Williams BA, Nichols FC, Marks RS, Jett JR, Aubry MC, et al. Gender differences in non–small-cell lung cancer survival: an analysis of 4,618 patients diagnosed between 1997 and 2002. Ann Thorac Surg. 2004;78(1):209–15.
6. Cerfolio RJ, Bryant AS, Scott E, Sharma M, Robert F, Spencer SA, et al. Women With Pathologic Stage I, II, and III Non-small Cell Lung Cancer Have Better Survival Than Men. Chest. 2006;130(6):1796–802.
7. Lærum D, Strand TE, Brustugun OT, Gallefoss F, Falk R, Durheim MT, et al. Evaluation of sex inequity in lung-cancer-specific survival. Acta Oncol. 2024;63:27572.
8. Bugge A, Kongerud J, Brunborg C, Solberg S, Lund MB. Gender-specific survival after surgical resection for early stage non-small cell lung cancer. Acta Oncol. 2017;56(3):448–54.
9. Yu XQ, Yap ML, Cheng ES, Ngo PJ, Vaneckova P, Karikios D, et al. Evaluating Prognostic Factors for Sex Differences in Lung Cancer Survival: Findings From a Large Australian Cohort. J Thorac Oncol. 2022;17(5):688–99.
10. Salmerón D, Chirlaque MD, Izarzugaza MI, Sánchez MJ, Marcos-Gragera R, Ardanaz E, et al. Lung cancer prognosis in Spain: The role of histology, age and sex. Respir Med. 2012;106(9):1301–8.
11. Araghi M, Fidler-Benaoudia M, Arnold M, Rutherford M, Bardot A, Ferlay J, et al. International differences in lung cancer survival by sex, histological type and stage at diagnosis: an ICBP SURVMARK-2 Study. Thorax. 2022;77(4):378–90.
12. Golozar A, Lawrence-Archer L, Zack T, Warner J, Reich C. Introducing ARTEMIS: Advanced Regimen Detection Using an Adapted Smith-Waterman Algorithm. OHDSI 2023 Global Symposium2023.
13. Syed H, Das AK. Temporal Needleman-Wunsch. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)2015. p. 1-9.
14. Warner JL, Dymshyts D, Reich CG, Gurley MJ, Hochheiser H, Moldwin ZH, et al. HemOnc: A new standard vocabulary for chemotherapy regimen representation in the OMOP common data model. J Biomed Inform. 2019;96:103239.
15. Blacketer C, Voss E. Extract Transform Load. The Book of OHDSI 2021.
16. Blacketer C. The Common Data Model. The Book of OHDSI 2021.
17. Schuemie M, Huser V, Blacketer C. Data Quality The Book of OHDSI2021.

18. Guidelines for Good Pharmacoepidemiology Practices (GPP), Revision 3. International Society for Pharmacoepidemiology; 2015.
19. Guideline on Good Pharmacovigilance Practices (GVP) - Module VIII – Post-Authorisation Safety Studies (Rev. 3). European Medicines Agency; 2017. Report No.: EMA/813938/2011 Rev 3.
20. ICMJE. Recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals. . International Committee of Medical Journal Editors; 2024.
21. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344-9.
22. Guide on Methodological Standards in Pharmacoepidemiology (Revision 11). The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP); 2010. Report No.: EMA/95098/2010.