**Characterizing Non-Small Cell Lung Cancer Patients with Liver metastasis using the OHDSI data network and the OMOP Oncology CDM module**

**Table of Contents**

[**Background and Rationale 4**](#_heading=h.1t3h5sf)

[**Objective(s) 4**](#_heading=h.2s8eyo1)

[**Methods 5**](#_heading=h.hqdlk02ma41g)

[**Study Design 5**](#_heading=h.17dp8vu)

[**Setting 5**](#_heading=h.35nkun2)

[**Study period 6**](#_heading=h.2p2csry)

[**Study identification period 6**](#_heading=h.3o7alnk)

[**Study population: inclusion criteria 6**](#_heading=h.32hioqz)

[**Follow up 7**](#_heading=h.4f1mdlm)

[**Variables 8**](#_heading=h.2u6wntf)

[***Treatment-related information*** 8](#_heading=h.19c6y18)

[***Outcomes*** 8](#_heading=h.28h4qwu)

[***Covariates*** 9](#_heading=h.nmf14n)

[**Study size 9**](#_heading=h.1mrcu09)

[**Data Management 10**](#_heading=h.46r0co2)

[**Analysis 10**](#_heading=h.2lwamvv)

[**Limitations 11**](#_heading=h.111kx3o)

[**Protection of Human Subjects 11**](#_heading=h.3l18frh)

[**Management and reporting of adverse events/adverse reactions 11**](#_heading=h.206ipza)

[**Plans for disseminating and communicating study results 11**](#_heading=h.4k668n3)

[**References 12**](#_heading=h.2zbgiuw)

# **Background and Rationale**

Forty percent of newly diagnosed non-small cell lung cancer (NSCLC) present with de novo metastatic disease1. A number of retrospective studies have indicated that liver metastases are a poor prognostic factor2-4. These studies were retrospective in nature and based on autopsy reports or databases that lack pharmacy data. In the years following publication of these studies, there has been substantial progress with regard to predictive biomarkers and corresponding therapeutics for metastatic NSCLC (mNSCLC). Some recent studies have reported that liver metastases may predict poor outcomes with immune checkpoint inhibitors (ICI)5-7. The MD Anderson immune checkpoint inhibitor (MDA-ICI) score was developed to aid prognostication of patients considering enrollment on phase 1 trials of ICI. This score includes seven clinical factors: age >52 years, ECOG >1, LDH >0.75 × ULN, platelet count >300 × 103 *μ*L−1 ANC >4.9 × 103 *μ*L−1, ALC, <1.8 × 103 *μ*L−1 and liver metastases and outperforms RMH and MDACC scores which had been developed in the pre-ICI era8. However, analysis of 74 heavily pretreated patients with NSCLC enrolled on phase 1 clinical trials of ICI at MD Anderson Cancer Center showed that liver metastases along with histology, mutation status and gender were not predictive of overall survival9. Lung immune prognostic index (LIPI) that uses derived neutrophil-lymphocyte ratio (dNLR) and LDH was associated with ICI but not chemotherapy outcomes10. Subsequent analysis showed that LIPI was associated with overall survival and progression free survival among patients receiving ICI, targeted therapy or chemotherapy11. There are several shortcomings of LIPI due to the heterogenous nature of the dataset. PDL1 expression was unknown in a significant proportion of patients and when available, wasn’t quantified. Many patients received second or subsequent line ICI therapy. Although results of multivariate analysis showed that dNLR and LDH level were independent predictors of overall survival, covariates that were included for univariate analysis are not known. Evaluation of dNLR, LDH level along with other parameters that have demonstrated prognostic significance in literature including liver metastases may help optimize prognostic indices. Therefore, in the current era of ICI-based therapies and selective tyrosine kinase inhibitors for NSCLC, it is necessary to clarify if liver metastases are predictive of therapeutic efficacy, prognostic or both. Defining the significance of liver metastases in the context of modern therapies will help risk stratify patients with metastatic NSCLC for treatment with standard of care regimen as well as clinical trials.

Liver metastasis at the time of diagnosis is reported among 13.4% of patients with metastatic NSCLC12. Single centre studies are limited in the number of patients and details needed comprehensive patient characterization. A federated network study can provide a more comprehensive picture of these patients particularly.

# **Objective(s)**

The overarching aim of this study is to characterize patients with metastatic NSCLC with and without liver metastasis at the time of diagnosis with metastatic NSCLC.

1. To describe demographics and clinical characteristics of patients with metastatic NSCLC and metastatic NSCLC patients with and without metastasis to liver who received systemic antineoplastic treatment
2. To characterize detailed treatment patterns among patients with metastatic NSCLC, and metastatic NSCLC patients with and without metastasis to liver. Specifically,

* Distribution of treatment regimens, dose, cycle, and scheduling per systemic anti-neoplastic line of treatment LoT (up to two LoTs)
* Treatment flow across 1s and 2nd (treatment pathways)
* Time to treatment discontinuation
* Time to next treatment

1. To estimate overall survival (OS) of patients with metastatic NSCLC, metastatic NSCLC patients with and without metastasis to liver.

# **Methods**

## **Study Design**

This will be a retrospective cohort study of patients diagnosed with metastatic NSCLC who received systemic antineoplastic treatment across a network of academic and commercial electronic health records (EHRs) and national claims data, all standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The OMOP CDM, the OMOP Oncology Module and associated tools will be used to conduct large scale federated analytics within the Observational Health Data Sciences and Informatics (OHDSI) Research Network. The OMOP Oncology Module supports the comprehensive representation of cancer diagnosis, treatments, and treatments and clinically relevant disease and treatment episodes and outcomes13. The feasibility of the OHDSI Research Network, and that of participating cancer centres, to generate detailed epidemiological insights into cancer patients has been successfully tested as a part of an ongoing characterization study of patients with metastatic bladder cancer13. OHDSI Research Network and the OMOP Oncology Module provide a unique opportunity to better characterize NSCLC patients with and without liver metastasis.

## **Setting**

This study will be conducted in a network of data sources. That means data will not be extracted and pooled to conduct the study. Instead, each participating institution will develop and maintain its own data asset.

## **Study period**

The study period starts from the date of database inception and end at latest available date for all data sources.

## **Study identification period**

The identification period will be from the date of database inception to six months prior to the latest available data in each database. This allows for a potential of 6 months of follow-up data for the last person included in the study.

## **Study population: inclusion criteria**

*Metastatic NSCLC (mNSCLC)*

Adult patients (≥18 years at) diagnosed with metastatic NSCLC

*NSCLC Cancer with Metastasis to Liver (NSCLC-Liver Met)*

Adult patients (≥18 years at) diagnosed with metastatic NSCLC who have liver metastasis at the time of diagnosis with metastatic NSCLC

*NSCLC Cancer without Metastasis to Liver (NSCLC-no Liver Met)*

Adult patients (≥18 years at) diagnosed with metastatic NSCLC who have no liver metastasis at the time of diagnosis with metastatic NSCLC

Participants will be identified using pre-specified concept sets that will be developed and reviewed by a core team of epidemiologist, clinicians, vocabulary experts and data scientist with extensive expertise in the use of OMOP CDM and the OHDSI tools.

Since Etoposide is almost exclusively used combination with other anti-neoplastic agents as the first treatment line for small cell lung cancer, patients with exposure to Etoposide after index date were not included in the study.

For this study, three study populations will be defined through condition, tumor attributes and drug concepts.

*Metastatic NSCLC (mNSCLC):*

* Have a diagnosis of metastatic NSCLC, defined as having either condition SNOMED 432851 "Secondary malignant neoplastic disease" or any of its descendants, OR tumor attributes 1539031 “lung cancer pM1”, 1538680 “lung cancer pM1a”, 1538946 “lung cancer pM1b”, 1538466 “lung cancer pM1”, 1539339 “lung cancer pM1”, 1538460 “lung cancer pM1a”, 1538152 “lung cancer pM1b”, 1539372, “lung cancer pM1c” of concept class AJCC Category OR 36769180 “metastasis” of concept class “Metastasis” or any of its descendants
* Have a prior diagnosis of NSCLC; condition of SNOMED 258369 "Primary malignant neoplasm of lung" or any of its descendants
* ≥18 years at index date
* No Etoposide (ATC L01CB01) exposure following diagnosis of metastatic NSCLC (index date)
* Anti-neoplastic agents (ATC 21601387) exposure following diagnosis of metastatic NSCLC (index date)

*Metastatic NSCLC with Liver Metastasis (NSCLC-Liver Met):*

* Have a diagnosis of NSCLC with metastasis to liver, defined as having either condition SNOMED 432851 " Secondary malignant neoplasm of liver and intrahepatic bile duct" or any of its descendants, OR tumor attribute 36770544 “Metastasis to the Liver”
* Have a prior diagnosis of NSCLC; condition of SNOMED 258369 "Primary malignant neoplasm of lung" or any of its descendants
* ≥18 years at index date
* No Etoposide (ATC L01CB01) exposure following diagnosis of metastatic NSCLC (index date)
* Anti-neoplastic agents (ATC 21601387) exposure following diagnosis of metastatic NSCLC (index date)

*Metastatic NSCLC without Liver Metastasis (NSCLC-Non-Liver Met):*

* Have a diagnosis of metastatic NSCLC, defined as having either condition SNOMED 432851 "Secondary malignant neoplastic disease" or any of its descendants, OR tumor attributes 1539031 “lung cancer pM1”, 1538680 “lung cancer pM1a”, 1538946 “lung cancer pM1b”, 1538466 “lung cancer pM1”, 1539339 “lung cancer pM1”, 1538460 “lung cancer pM1a”, 1538152 “lung cancer pM1b”, 1539372, “lung cancer pM1c” of concept class AJCC Category OR 36769180 “metastasis” or any of its descendants AND NO diagnosis of NSCLC with metastasis to liver, defined as having either condition SNOMED 432851 " Secondary malignant neoplasm of liver and intrahepatic bile duct" or any of its descendants, OR tumor attribute 36770544 “Metastasis to the Liver”
* Have a prior diagnosis of NSCLC; condition of SNOMED 258369 "Primary malignant neoplasm of lung" or any of its descendants
* ≥18 years at index date
* No Etoposide (ATC L01CB01) exposure following diagnosis of metastatic NSCLC (index date)
* Anti-neoplastic agents (ATC 21601387) exposure following diagnosis of metastatic NSCLC (index date)

## **Follow up**

Index date will be defined as the **date of initiation of first LoT**. Patients will be followed until death, last known activity or end of the study period at 31 March 2021.

## **Variables**

### ***Treatment-related information***

Treatment-related information will be standardized through RxNorm, defining their active ingredient (agent), dose, and form. An algorithm will be developed to parse regimens (defined through Hematology/Oncology (HemOnc))14 from the individual drug exposure records by utilizing the ingredients, their timing and their dosing.

HemOnc treatment regimens will be used to derive LoT. Date of the first drug episode (e.g., first administration or non-cancelled order) for anti-neoplastic agents (ATC 21601387) after diagnosis of metastatic NSCLC or metastatic NSCLC with/without liver metastasis will be considered as the start of the first LoT. For Claims databases, we will require a six-month washout period with no drug exposure to antineoplastic agents (ATC 21601387) prior to the initiation of the treatment regimen. The date patient discontinues first LoT treatment will be considered the end date of the first LoT. Discontinuation will be defined as having a subsequent systemic anti-neoplastic regimen after the first line of treatment; having a gap of more than 120 days with no systemic anti-neoplastic therapy following the last administration; or having a date of death while on the regimen. Patients will be censored at their last known usage within the database or end of follow-up. Receipt of a new anti-neoplastic regimen after the end of the first LoT will initiate a subsequent LoT; and the start and end of subsequent LoTs will be identified using the logic described above.

### ***Outcomes***

The following outcomes are assessed in the study:

**Diagnosis to Treatment Interval (DT):** the number of days from date of diagnosis to the

initiation of the first LoT.

**Treatment-free Interval (TFI):** Time from discontinuation of one LoT to initiation of the subsequent LoT, or date of death if death occurs prior to start of the subsequent LoT. Patients will be censored at their last activity within the database or end of follow-up.

**Time to Next Treatment (TTNT):** Time from the index date to the date the patient received their next systemic anti-neoplastic treatment regimen or to their date of death if death occurs prior to having another systemic anti-neoplastic treatment regimen. Patients will be censored at their last activity within the database or end of follow-up.

**Time to Treatment Discontinuation (TTD):** Length of time from the initiation of each LoT to the date the patient discontinues the treatment (i.e., the last administration or noncancelled order of a drug contained in the same regimen). TTD will be described for the first two LoTs. *Discontinuation* will be defined as having a subsequent systemic anti-neoplastic therapy regimen after the first LoT; having a gap of more than 120 days with no systemic anti-neoplastic therapy following the last administration; or having a date of death while on the regimen. Patients will be censored at their last known usage within the database or end of follow-up.

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

### ***Covariates***

Baseline covariates will be defined by observations prior to or at the index date.

The following characteristics will be identified **prior to or at index date**.

Age:

age will be categorized into 10-year groups. To ensure patient’s privacy, patients above 85 will be grouped together.

age will also be described as a continuous variable

Biologic sex

Race (where available)

Smoking

Charlson comorbidity index

Diabetes

Hypertension

Cardiovascular disease

Cerebrovascular Disease

Chronic Obstructive Pulmonary Disease

Moderate/Severe Liver Disease

Mild liver disease

Renal disease

Dementia

Autoimmune disease

Other malignancy (except for non-melanoma skin cancer) at or prior top index date

Lab values:

* 1. Serum lactate dehydrogenase (LDH) levels
  2. Hemoglobulin
  3. Creatinine
  4. Albumin level
  5. Lactate dehydrogenase level
  6. Absolute lymphocyte count
  7. Total white blood cell count
  8. Platelet count
  9. Absolute neutrophil count
  10. Derived neutrophil to lymphocyte ratio
  11. Platelet to lymphocyte ratio
  12. PDL-1 expression

## **Study size**

This study will use routinely collected data, all patients meeting the eligibility criteria above will be included. No formal sample size and power calculation will be conducted.

## **Data Management**

Source data at the participating institutions is converted to the OMOP CDM through a process known as ETL (extraction, transformation, loading)15. This process defines the rules for how source codes are mapped to standardized concepts in OMOP while maintaining quality of information. The data is refreshed periodically responding to updates by the data source. The OMOP process maps source codes to a standardized set of concepts depending on their domain. There are five key standardized domains: condition, drug, measurement, procedure and observation. Data sources participating in the study network does not provide ETL information. These data sources are responsible for their own validation to ensure that the OMOP dataset matched the quality of information provided in the original source.

## **Analysis**

All analyses will be performed using code developed for the OHDSI Methods library. A diagnostic package built off the OHDSI Cohort Diagnostics (https://ohdsi.github.io/CohortDiagnostics/) library, is included in the base package as a preliminary step to assess the fitness of use of phenotypes on your database. If a database passes cohort diagnostics, the full study package will be executed. Baseline covariates will be extracted using an optimized SQL extraction script based on principles of the FeatureExtraction package (http://ohdsi.github.io/FeatureExtraction/) to quantify Demographics, Condition Group Eras, and Drug Group Eras. Additional cohort-specific covariates will be constructed using OMOP Standard Vocabulary concepts. At the time of executing Feature Extraction, the package will create a data frame in which individuals’ age will be extracted. Individuals’ medical conditions, procedures, measurements and medications will be summarized 1) over the year prior to their index date, 2) at index date and over the follow-up time.

Demographic and clinical characteristics will be summarized for all cohorts; mNSCLC, NSCLC-Liver Met and NSCLC-Non-Liver Met. For categorical variables, frequencies and percentages will be presented for each level. Continuous variables will be summarized and mean (standard deviation (SD)), minimum, maximum, median (within each database), and interquartile range (within each database). Standardized mean differences (SMD) will be calculated when comparing characteristics of study cohorts.

Distribution of treatment regimens by LoT (up to two LoTs) will be summarized for each cohort. Pending data availability, distribution of treatment regimens by year will be summarized in each LoT and for each cohort.

For each cohort, DTI will be reported as median (inter quartile range (IQR)).

**Time to event outcomes including TFI, TTD, TTNT and OS will be evaluated descriptively by Kaplan-Meier curves along with two-sided 95% (confidence interval (CI)) for the median time to event estimates.**

The extracted treatment regimens will be used to describe treatment flow across 1L, 2L and 3L treatments (treatment pathways) using Sankey diagrams.

Treatment regimens will be categorized into the following groups:

* EGFR tyrosine kinase inhibitors (TKI) [Erlotinib, Gefitinib, Afatinib, Dacomitinib, Osimertinib]
* Other TKIs [Crizotinib, Ceritinib, Brigatinib, Alectinib, Lorlatinib, Entrectinib,, Capmatinib, Selpercatinib, Pralsetinib, Vandetanib, Cabozantinib, Lenvatinib, Larotrectinib, Dabrafenib+Trametinib]
* Immune checkpoint inhibitors (anti-PD1/L1, anti-CTLA-4 or both)
* Immune checkpoint inhibitors (anti-PD1/L1) and platinum doublet chemotherapy with or without anti-VEGF monoclonal antibody (mAb) or dual immune checkpoint inhibitors (anti-PD1 and anti-CTLA-4) and platinum doublet chemotherapy
* Platinum doublet chemotherapy with or without anti-VEGF mAb
* Single agent chemotherapy with or without anti-VEGF mAb [Pemetrexed with or without Bevacizumab, Docetaxel with or without Ramucirumab]

All the proposed analyses are conducted for the top five treatment groups in each LoT, separately. Analyses are conducted in each database separately using R16 and Redshift SQL.

# **Limitations**

This study will be carried out using data recorded in a collection of HER and Claims systems. Misclassification may occur in identification of patients, outcomes and covariates. As with any healthcare database used for secondary data analysis, the patient records are expected to be incomplete in many respects and may have had erroneous entries. Data regarding diagnosis of metastatic disease, site of metastasis, chemotherapy regimens used or baseline covariates prior to enrollment within the database may not be available. Treatment provided in hospitals or any other setting outside each participating institution is not included. Finally, comorbidities that could influence treatment given in the oncology clinic but that were themselves treated outside of the oncology setting tend to be somewhat under-reported in some databases. The primary purpose of the study is to describe a real-world cohort of patients with NSCLC in the institutions participating in the study and, therefore, the results should not be interpreted as generalizable to the overall population.

# **Protection of Human Subjects**

This study will be conducted in accordance with the International Society of Pharmacoepidemiology (ISPE) Guidelines for Good Pharmacoepidemiology Practices (GPP)17 and EMA, European Network of Centers for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Guide on Methodological Standards in Pharmacoepidemiology18. The use of the OMOP common data model and OHDSI tools enable the federated analysis of these different databases without accessing patient-level data outside the participating institutions. The study will be 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/or disclosure of any information related to this study under the Study Agreement. Each data partner will be required to provide statement about IRB approval or exemption to participate.

# **Management and reporting of adverse events/adverse reactions**

According to the new guidelines for good pharmacovigilance practice (EMA/873138/2011)19 and ISPE17, there is no requirement for expedited reporting of adverse drug reactions from studies with secondary use of data (such as electronic health care databases).

# **Plans for disseminating and communicating study results**

Results of this study will be published following guidelines, including those for authorship, established by the International Committee of Medical Journal Editors20. When reporting results of this study, the appropriate Strengthening the Reporting of Observational Studies in Epidemiology checklist will be followed21.

# **References**

1. Morgensztern D, Waqar S, Subramanian J, et al: Improving Survival for Stage IV Non-small Cell Lung Cancer: A Surveillance, Epidemiology, and End Results Survey from 1990 to 2005. Journal of Thoracic Oncology 4:1524-1529, 2009

2. Ren Y, Dai C, Zheng H, et al: Prognostic effect of liver metastasis in lung cancer patients with distant metastasis. Oncotarget 7:53245-53253, 2016

3. Riihimäki M, Hemminki A, Fallah M, et al: Metastatic sites and survival in lung cancer. Lung Cancer 86:78-84, 2014

4. Tamura T, Kurishima K, Nakazawa K, et al: Specific organ metastases and survival in metastatic non-small-cell lung cancer. Molecular and clinical oncology 3:217-221, 2015

5. Schmid S, Diem S, Li Q, et al: Organ-specific response to nivolumab in patients with non-small cell lung cancer (NSCLC). Cancer Immunology, Immunotherapy 67:1825-1832, 2018

6. Tamiya M, Tamiya A, Suzuki H, et al: The relationship between metastatic sites and progression free survival of nivolumab in non-small cell lung cancer patients. Journal of Clinical Oncology 35:e20679-e20679, 2017

7. Addepalli S, Stroud CRG, Cherukuri SD, et al: Outcomes with immune checkpoint inhibitor use in lung cancer patients with hepatic metastases. Journal of Clinical Oncology 35:38-38, 2017

8. Sen S, Hess K, Hong DS, et al: Development of a prognostic scoring system for patients with advanced cancer enrolled in immune checkpoint inhibitor phase 1 clinical trials. British journal of cancer 118:763-769, 2018

9. Maymani H, Hess K, Groisberg R, et al: Predicting outcomes in patients with advanced non-small cell lung cancer enrolled in early phase immunotherapy trials. Lung Cancer 120:137-141, 2018

10. Mezquita L, Auclin E, Ferrara R, et al: Association of the Lung Immune Prognostic Index With Immune Checkpoint Inhibitor Outcomes in Patients With Advanced Non–Small Cell Lung Cancer. JAMA Oncology 4:351-357, 2018

11. Kazandjian D, Gong Y, Keegan P, et al: Prognostic Value of the Lung Immune Prognostic Index for Patients Treated for Metastatic Non–Small Cell Lung Cancer. JAMA Oncology 5:1481-1485, 2019

12. Tamura T, Kurishima K, Nakazawa K, et al. Specific organ metastases and survival in metastatic non-small-cell lung cancer. Mol Clin Oncol. 2015;3(1):217-221. doi:10.3892/mco.2014.410

13. Belenkaya R, Gurley MJ, Golozar A, et al. Extending the OMOP Common Data Model and Standardized Vocabularies to Support Observational Cancer Research. JCO Clin Cancer Inform. 2021;5:12-20. doi:10.1200/CCI.20.00079

14. Warner JL, Dymshyts D, Reich CG, et al. HemOnc: A new standard vocabulary for chemotherapy regimen representation in the OMOP common data model. J Biomed Inform. 2019;96: 103239

15. OHDSI. The Book of OHDSI: OHDSI, 2019:458.

16. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

17. International Society for Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practices (GPP), Revision 3; June 2015. Available at: https://www.pharmacoepi.org/resources/policies/guidelines-08027/. Accessed 6 May 2021.

18. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). EMA/95098/2010. Available at http://www.encepp.eu/standards\_and\_guidances. /. Accessed 6 May 2021.

19. European Medicines Agency. Guidelines on good pharmacovigilance practices (GVP): Module VIII – Post-authorisation safety studies. \* EMA/813938/2011 Rev 3\*. 9 October 2017. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-viii-post authorisation-safety-studies-rev-3\_en.pdf. Accessed 6 May 2021

20. ICMJE. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. International Committee of Medical Journal Editors; December 2019. Available at: http://www.icmje.org/urm\_main.html. Accessed 6 May 2021.

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 Apr;61(4):344-9.