



Funded by the
Nordic Council
of Ministers

VALO NSCLC pilot study

Study Report version 1.0

Value from Nordic health data – VALO project

31 October 2025



Document info

0.1 Authors

Author	Partner
Atif Adam, IQVIA Senior Epidemiologist	IQVIA
Collaborator / Buyer Representative	
Markus Kalliola	Finnish Innovation Fund Sitra
Principal Investigator	
Åslaug Helland, Main Principal Investigator	Oslo University Hospital
IQVIA project responsible parties	
Susanna Flaherty, Pilot Project Director	IQVIA
Elina Winblad, Pilot Project Manager	IQVIA
James Brash, Pilot Senior Data Scientist	IQVIA

0.2 Keywords

Keywords	Health data, Nordic, EHDS
----------	---------------------------

Disclaimer

The content of this report represents the views of the author(s) and do not necessarily reflect the views of the experts interviewed.

Copyright Notice

Copyright © 2025 Sitra and the Nordic Council of Ministers. All rights reserved. For more information on the project, please see the [project website](#).



Contents

Executive Summary	4
Tiivistelmä	5
Sammanfattning	6
1 Abbreviations	7
2 Abstract	9
3 Study Background	12
3.1 Disease Burden and Epidemiology	12
3.2 Molecular Landscape and Targeted Therapies.....	12
3.3 Immunotherapy in mNSCLC	12
3.4 Treatment Challenges and Patient Heterogeneity	13
3.5 Real-World Evidence Needs	13
3.6 Study Rationale.....	13
3.6.1 Nordic Healthcare Context and Opportunity	14
3.6.2 The OMOP-Based Federated Analysis Paradigm	14
3.6.3 Pilot Study Design and Learning Objectives.....	15
3.6.4 Pilot Study Framework and Interpretation Guidelines.....	15
4 Research Aim and Objectives	17
4.1 Research Aim.....	17
4.2 Study Objectives	17
4.2.1 Main Objectives.....	17
4.2.2 Exploratory Objectives	17
5 Research Methods	18
5.1 Study Design.....	18
5.1.1 Study Design Summary	18
5.2 Study Population	20
5.3 Data Sources	21
5.3.1 OHDSI Network.....	21
5.3.2 Systematic Use of Databases	21
5.4 Covariates and Variables	22
5.5 Data Analysis	25
5.5.1 Analytical Framework.....	25
5.5.2 Analytical Consideration per Objective	25
5.6 Data Collection Constraints and Site Implementation Variability	27
5.6.1 Selection Bias	27
5.6.2 Data Standardization and Mapping Challenges.....	28
5.6.3 Misclassification	28
5.6.4 Federated Analysis Constraints	28
5.6.5 Temporal and Geographic Limitations	29
5.6.6 Disease Presentation Classification Limitations.....	29
6 Results	30
6.1 Objective 1: Cohort Assembly and Data Completeness	30
6.1.1 Data Completeness and Quality Assessment.....	31
6.2 Objective 1: Baseline Demographic and Clinical Characteristics.....	32
6.2.1 Study Population - OUS.....	32
6.2.2 Study Population - Rigshospitalet/Herlev Hospital.....	36
6.2.3 Study Population - HUS	41
6.2.4 PD-L1 Expression Status	45



6.3 Objective 2a: Treatment Patterns Prior-to and Post Metastatic Progression	46
6.3.1 Treatment Characterization - OUS.....	46
6.3.2 Treatment Characterization – Rigshospitalet/Herlev Hospital.....	49
6.3.3 Treatment Characterization - HUS.....	51
6.4 Objective 2b: Longitudinal Treatment Sequences	53
6.4.1 Treatment Sequence Patterns – OUS	53
6.4.2 Treatment Sequence Patterns – Rigshospitalet/Herlev Hospital	55
6.4.3 Treatment Sequence Patterns – HUS.....	58
6.4.4 Sequential Treatment Patterns and Maintenance Strategies.....	61
6.4.5 Treatment Duration Dynamics	61
6.5 Objective 4: Clinical Outcomes - Overall Survival.....	61
6.5.1 Overall Survival Analysis - OUS.....	61
6.5.2 Overall Survival Analysis - Rigshospitalet/Herlev Hospital	65
6.5.3 Overall Survival Analysis - HUS.....	68
6.6 Incomplete Study Objectives.....	71
7 Discussion	73
7.1 Principal Findings.....	73
7.2 Limitations	76
7.3 Conclusion	80
References.....	82

Executive Summary

This pilot study demonstrates the technical feasibility of conducting federated OMOP-based analytics across Nordic healthcare systems. This proof-of-concept pilot analyzed real-world treatment patterns and outcomes for 755 patients with metastatic non-small cell lung cancer (mNSCLC) receiving first-line immunotherapy in Denmark, Finland, and Norway.

Key results

- The pilot successfully implemented a privacy-preserving federated analysis infrastructure using the OMOP common data model.
- Notable differences were observed in treatment patterns and survival outcomes across sites, with median overall survival ranging from approximately 315 to 542 days, highlighting variability that may stem from data capture and patient selection rather than true clinical differences.

Critical limitations

- Major data completeness gaps were observed: ECOG performance status absent in two sites, smoking history universally missing, and heterogeneous PD-L1 reporting.
- Incomplete radiotherapy documentation and reliance on proxy measures for chemotherapy cycles.
- Selection bias and care fragmentation limit generalizability; observed differences cannot be interpreted as healthcare system or treatment effectiveness.

Implications and next steps

This pilot serves primarily as a methodological milestone rather than a source of actionable clinical evidence. It identifies critical prerequisites for future Nordic collaboration: standardized capture of prognostic covariates, validated algorithms for key classifications (emergent versus de novo disease), harmonized treatment intent and documentation, and prospective data quality monitoring.

The successful execution of this pilot, despite documented data limitations, validates the core premise that Nordic healthcare systems can collaborate effectively through federated OMOP infrastructure. Subsequent studies with prospectively standardized covariate capture, particularly ECOG performance status, smoking history, and PD-L1 expression, will be positioned to generate clinically actionable evidence regarding optimal mNSCLC treatment strategies across Nordic healthcare systems.



Tiivistelmä

Tämä pilottitutkimus osoittaa, että Pohjoismaisessa terveydenhuoltojärjestelmässä on teknisesti mahdollista toteuttaa hajautettua OMOP-pohjaista tutkimusta. Tämä toteutettavuuspilotti analysoi todellisen maailman hoitokäytäntöjä ja tuloksia 755 potilaalta, joilla oli metastaatinen ei-pienisolainen keuhkosityöpä (mNSCLC) ja jotka saivat ensilinjan immunoterapiaa Tanskassa, Suomessa ja Norjassa.

Keskeiset tulokset

- Pilotti toteutti onnistuneesti yksityisyyttä suojaavan hajautetun analyysi-infrastruktuurin OMOP-tietomallia käyttäen.
- Merkittäviä eroja havaittiin hoitokäytännöissä ja elossaolotuloksissa eri toimipaikkojen välillä. Mediaani kokonaiselossaoloaika vaihteli noin 315–542 päivän välillä, mikä korostaa vaihtelua, joka saattaa johtua datan kirjaamisesta ja potilasvalinnasta eikä todellisista kliinisistä eroista.

Keskeiset rajoitteet

- Merkittäviä puutteita havaittiin datan täydellisyydessä: ECOG-toimintakykyluokitus puuttui kahdessa toimipisteessä, tupakointihistoria puuttui kaikilta, ja PD-L1-raportointi oli epäyhtenäistä.
- Säteihoidon dokumentointi oli puutteellista ja kemoterapiakertoja arvioitiin epäsuorien mittareiden avulla.
- Valikoitumisharha ja hoidon pirstoutuminen heikentävät yleistettävyyttä; havaittuja eroja ei voida tulkita terveydenhuoltojärjestelmien tai hoitojen tehokkuuseroiksi.

Johtopäätökset ja seuraavat askeleet

Tämä pilotti toimii ensisijaisesti menetelmällisenä virstanpylväänä eikä kliinisesti hyödynnettävänä näytön lähteenä. Se tunnistaa keskeiset edellytykset tulevalle pohjoismaiselle yhteistyölle: prognostisten muuttujien standardoitu kirjaaminen, validoidut algoritmit keskeisiin luokituksiin (esim. emergentti vs. de novo -tauti), yhdenmukainen hoitotarkoituksen ja dokumentaation kirjaaminen sekä prospektiivinen datan laadun seuranta.

Pilotin onnistunut toteutus, huolimatta dokumentoiduista datan puutteista, vahvistaa keskeisen periaatteen: pohjoismaiset terveydenhuoltojärjestelmät voivat tehdä tehokasta yhteistyötä hajautetun OMOP-infrastruktuurin avulla. Jatkossa tutkimukset, joissa kovariaattien kirjaaminen standardoidaan prospektiivisesti — erityisesti ECOG-toimintakyky, tupakointihistoria ja PD-L1-ilmentyminen — voivat tuottaa kliinisesti hyödynnettävää näyttöä optimaalisten mNSCLC-hoitostrategioiden määrittämiseksi Pohjoismaissa.

Sammanfattning

Denna pilotstudie visar den tekniska genomförbarheten av att utföra federerad OMOP-baserad analys inom nordiska sjukvårdssystem. Denna proof-of-concept-pilot analyserade behandlingsmönster och utfall i verklig klinisk praxis för 755 patienter med metastaserad icke-småcellig lungcancer (mNSCLC) som fick förstalinjens immunterapi i Danmark, Finland och Norge.

Huvudresultat

- Piloten genomförde framgångsrikt en integritetsskyddande federerad analysinfrastruktur baserad på OMOP-modellen.
- Betydande skillnader observerades i behandlingsmönster och överlevnadsutfall mellan sajterna. Den mediana totalöverlevnaden varierade från cirka 315 till 542 dagar, vilket antyder variation som sannolikt beror på datainsamling och patientselektion snarare än verkliga kliniska skillnader.

Viktiga begränsningar

- Stora brister i datakompletthet observerades: ECOG-funktionsstatus saknades i två sajter, rökvanor saknades helt, och PD-L1-rapportering var heterogen.
- Ofullständig dokumentation av strålbehandling samt beroende av proxy-mått för antal cytostatikacykler.
- Selektionbias och fragmenterad vård begränsar generaliserbarheten; observerade skillnader kan inte tolkas som system- eller behandlingseffektivitet.

Implikationer och nästa steg

Denna pilot fungerar framför allt som en metodologisk milstolpe snarare än som en källa till kliniskt handlingsbar evidens. Den identifierar centrala förutsättningar för framtida nordiskt samarbete: standardiserad registrering av prognostiska kovariater, validerade algoritmer för viktiga klassificeringar (t.ex. emergent vs. de novo-sjukdom), harmoniserad dokumentation av behandlingsavsikt, prospektiv övervakning av datakvalitet.

Pilotens framgångsrika genomförande, trots dokumenterade databegränsningar, bekräftar den grundläggande principen att nordiska sjukvårdssystem kan samarbeta effektivt via en federerad OMOP-infrastruktur. Efterföljande studier med prospektivt standardiserad datainsamling — särskilt ECOG-funktionsstatus, rökvanor och PD-L1-uttryck — kommer att kunna generera kliniskt användbar evidens för att identifiera optimala behandlingsstrategier för mNSCLC inom Norden.



1 Abbreviations

Abbreviation	Definition
ALK	Anaplastic lymphoma kinase
AMI	Acute myocardial infraction
ATC	Anatomical Therapeutic Chemical (classification)
BMI	Body mass index
BRAF	B-Raf proto-oncogene
CDM	Common data model
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
COVID-19	Coronavirus disease 2019
DQD	Data Quality Dashboard
DRG	Diagnosis-related group
ECOG PS	Eastern Cooperative Oncology Group Performance Status
EGFR	Epidermal growth factor receptor
EHR	Electronic health records
EMR	Electronic medical records
ERBB2	Erb-B2 receptor tyrosine kinase 2
GDPR	General Data Protection Regulation
HCRU	Healthcare resource utilization
HER2	Human epidermal growth factor receptor 2
HUS	Helsinki University Hospital
ICD	International Classification of Diseases
ICD-10	International Classification of Diseases, 10th Revision
ICI	Immune checkpoint inhibitor
IQR	Interquartile range
KM	Kaplan-Meier
KRAS	Kirsten rat sarcoma virus
MET	Mesenchymal epithelial transition
mNSCLC	Metastatic non-small cell lung cancer
NA	Not available
NSCLC	Non-small cell lung cancer
NTRK	Neurotrophic receptor tyrosine kinase



Abbreviation	Definition
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OS	Overall survival
OUS	Oslo University Hospital
PAD	Peripheral arterial disease
PD-1	Programmed cell death protein 1
PD-L1	Programmed death-ligand 1
R	Statistical software environment
RET	Rearranged during transfection
ROS1	c-ROS oncogene 1
RW	Real-World
SCLC	Small cell lung cancer
SD	Standard deviation
SOPs	Standard operating procedures
TKI	Tyrosine kinase inhibitors
VALO	Value from Nordic Health Data (study acronym)

2 Abstract

Study type / Study phase: Retrospective observational cohort study using federated analysis

Keywords: Metastatic NSCLC, immune checkpoint inhibitors, real-world evidence, Nordic healthcare, treatment patterns, overall survival, OMOP CDM, federated analysis

Rationale and background: Despite advances in immunotherapy for metastatic non-small cell lung cancer (mNSCLC), significant gaps remain in understanding real-world treatment patterns, clinical outcomes, and healthcare utilization across different healthcare systems. The Nordic region, with its universal healthcare coverage and similar treatment guidelines, provides a unique opportunity to examine how healthcare system factors influence patient outcomes independent of drug access.

Federated Infrastructure Validation: As a pilot implementation of Observational Medical Outcomes Partnership (OMOP)-based federated analytics in the Nordic region, this study serves dual purposes: demonstrating technical infrastructure feasibility while generating exploratory clinical insights. The collaborative governance structures, data quality assessment protocols, and analytical workflows established through this pilot provide a foundation for future Nordic real-world evidence research.

Research objectives:

1. To describe baseline demographic and clinical characteristics of mNSCLC patients receiving first-line immune checkpoint inhibitor (ICI) therapy
2. To analyze longitudinal treatment patterns including sequence, duration, and intensity of therapies
3. To evaluate overall survival outcomes stratified by age and country
4. To assess healthcare resource utilization and costs (not completed due to data limitations)

Study design: Multi-national federated analysis utilizing the OMOP common data model (CDM) across three Nordic university hospitals from January 2018 to December 2023, with data extraction completed March 2024. Standardized analysis scripts were executed locally at each site with aggregated results compiled centrally.

Databases:

1. Oslo University Hospital (OUS), Norway
2. Capital Region Rigshospitalet/Herlev Hospital (Rigshospitalet/Herlev Hospital), Denmark
3. Helsinki University Hospital (HUS), Finland

Study population: 755 patients with confirmed mNSCLC initiating ICI as first-line treatment (OUS n=67, Rigshospitalet/Herlev Hospital n=489, HUS n=199), excluding those with epidermal growth factor receptor (EGFR)/anaplastic lymphoma kinase (ALK)/c-ROS oncogene 1 (ROS1) mutations. Age stratification: <75 years (77.5%) and ≥75 years (22.5%).

Variables:

- Demographics: Age, sex, body mass index (BMI)
- Clinical: Eastern Cooperative Oncology Group (ECOG) performance status, disease presentation (de novo vs emergent), comorbidities (diabetes, cardiovascular disease, chronic obstructive pulmonary disease [COPD])
- Treatment: ICI type, chemotherapy, radiotherapy, surgery, treatment sequences and duration
- Outcomes: Overall survival, treatment line progression

Results:

Objective 1 (Baseline characteristics): The final cohort comprised 755 patients (Rigshospitalet/Herlev Hospital n=489, 64.8%; HUS n=199, 26.4%; OUS n=67, 8.9%). Median age ranged from 67 years (Rigshospitalet/Herlev Hospital) to 70 years (HUS, OUS). OUS's smaller sample could be reflected in documented incomplete patient journey capture, particularly for radiotherapy and later-line treatments delivered at external facilities. Male predominance was consistent across sites (63.3-70.1%). Critical prognostic covariates showed substantial data completeness gaps: ECOG performance status was completely absent (0% capture) in Rigshospitalet/Herlev Hospital and HUS, with 46.3% missing in OUS; smoking history was universally unavailable (100% missing); PD-L1 expression data was heterogeneous and largely unavailable across all sites. Emergent metastatic disease (defined as metastasis >120 days post-diagnosis) ranged from 7.8% (Rigshospitalet/Herlev Hospital) to 14.6% (HUS). Comorbidity prevalence varied substantially between diagnosis-based and medication-based assessment, with cardiovascular medication capture (83.6-99.5%) likely reflecting supportive cancer care rather than true cardiovascular disease burden.

Objective 2a (First-line treatment patterns): Pembrolizumab-based regimens were the predominant first-line therapy, ranging from 51.5% (Rigshospitalet/Herlev Hospital) to 76.1% (OUS). Documented radiotherapy utilization varied from 0% (OUS—representing data capture limitations rather than actual clinical practice) to 75.5% (Rigshospitalet/Herlev Hospital) and 66.3% (HUS). Chemotherapy co-administration with immunotherapy ranged from 56.5% (OUS) to 86.7% (Rigshospitalet/Herlev Hospital).

Objective 2b (Treatment sequences and duration): Median first-line treatment duration ranged from 52 days (HUS) to 100 days (OUS). Second-line therapy was documented in 34.3% (OUS) to 46.7% (HUS) of patients. Median time from first-line end to second-line initiation varied from 52 days (HUS) to 100 days (OUS). Third-line therapy was rare, documented in 9.0-15.4% of patients across sites. Elderly patients (≥75 years) demonstrated 14-18 percentage point reductions in chemotherapy utilization compared to younger patients across all sites.

Objective 3 (Overall survival): Median overall survival ranged from 315 days (95% CI: 220-485, OUS) to 542 days (95% CI: 465-648, Rigshospitalet/Herlev Hospital), with HUS intermediate at 380 days (95% CI: 309-504). Six-month survival probability ranged from 63.5% (OUS) to 80.0% (Rigshospitalet/Herlev Hospital). One-year survival ranged from 35.7% (OUS) to 54.0% (Rigshospitalet/Herlev Hospital). Two-year survival ranged from 20.9% (OUS) to 33.8% (Rigshospitalet/Herlev Hospital). Elderly patients demonstrated consistently lower survival across all sites, with median survival ranging from 197 days (HUS) to 427 days (Rigshospitalet/Herlev Hospital).



Data quality limitations: These survival differences cannot be interpreted as reflecting real-world treatment effectiveness or healthcare system performance due to fundamental data quality limitations. Substantial patient attrition during cohort assembly (from approximately 4,300 initially identified to 755 in final analysis) combined with differential selection patterns across sites creates high risk of selection bias that likely accounts for the observed survival variations. The absence of ECOG performance status in Rigshospitalet/Herlev Hospital and HUS (0% capture), lack of smoking history documentation (100% missing), and heterogeneous PD-L1 reporting precluded adjusted survival analyses accounting for disease severity, patient fitness, or biomarker status. The observed variations may reflect patient population heterogeneity, data capture limitations, or unmeasured confounders. Given the severity of data completeness gaps, these findings cannot support hypothesis generation until fundamental data quality issues are resolved. The variations primarily demonstrate the critical need for standardized covariate capture in federated observational research, rather than reflecting interpretable clinical or healthcare system differences.

Conclusion: This pilot study successfully demonstrates the technical feasibility of federated OMOP-based analytics across Nordic healthcare systems, analyzing 755 patients with metastatic NSCLC receiving first-line immunotherapy. Cross-national variation was observed in median survival (315-542 days), documented radiotherapy utilization (0-75.5%), treatment sequencing approaches, and elderly patient treatment intensity. However, critical data completeness gaps—including complete absence of ECOG performance status in Rigshospitalet/Herlev Hospital and HUS (0% capture), universal lack of smoking history documentation (100% missing), and heterogeneous PD-L1 reporting—prevent meaningful interpretation of these variations. The observed differences cannot be attributed to healthcare system performance, clinical practice patterns, or patient characteristics due to unmeasured confounding. These findings serve primarily as a methodological demonstration of federated infrastructure capabilities and a diagnostic assessment identifying essential data elements for future Nordic RWE research. The observed variations identify priorities for future Nordic collaboration: standardized capture of prognostic covariates to enable adjusted analyses, validated algorithms for key classifications including emergent versus de novo disease, harmonized treatment intent and radiotherapy documentation, and prospective data quality monitoring. This pilot establishes the infrastructure and data quality requirements for future Nordic comparative effectiveness research. Subsequent studies with prospectively standardized covariate capture—particularly ECOG performance status, smoking history, and PD-L1 expression—will be positioned to generate clinically actionable evidence regarding optimal mNSCLC treatment strategies across Nordic healthcare systems.

3 Study Background

3.1 Disease Burden and Epidemiology

Lung cancer is among the most frequently diagnosed cancers and a leading cause of cancer-related deaths globally(1, 2). The primary subtypes of lung cancer are small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC)(3). The most prevalent forms of NSCLC include adenocarcinoma (40%), squamous-cell carcinoma (25–30%), and large-cell carcinoma (5–10%)(4). NSCLC represents 85% of all primary lung cancers worldwide with the global 5-year survival rate being ~5-12% at diagnosis(5). Around 60-70% of NSCLC patients are metastatic (m) at diagnosis, with the annual global number of patients being ~ 1 million(1, 6). For patients with advanced-stage disease according to the TNM 8th edition, 5-year survival rates decline progressively: stage IIIA (36-41%), stage IIIB (24-26%), and stage IIIC (12-13%)(7).

3.2 Molecular Landscape and Targeted Therapies

The therapeutic landscape for mNSCLC has drastically evolved, with numerous treatment options that vary based on factors such as patient performance status, comorbidities, disease stage, histology, and their genomic and programmed death-ligand 1 (PD-L1) profile, thereby necessitating a strong multidisciplinary approach to patient care(8-10).

mNSCLC is characterized by a multitude of genetic alterations that drive oncogenesis. These can act as strong predictive biomarkers and hence are excellent targets of treatment to improve the 5-year survival of patients with mNSCLC(11, 12). European Society for Medical Oncology (ESMO) guidelines for oncogene-addicted metastatic NSCLC recommend comprehensive molecular testing including epidermal growth factor receptor (EGFR) mutations, anaplastic lymphoma kinase (ALK) rearrangements, c-ROS oncogene 1 (ROS1) rearrangements, B-Raf proto-oncogene (BRAF) V600E mutations, rearranged during transfection (RET) rearrangements, mesenchymal epithelial transition (MET) exon 14 skipping mutations and amplifications, neurotrophic receptor tyrosine kinase (NTRK) fusions, erb-B2 receptor tyrosine kinase 2 (ERBB2) (human epidermal growth factor receptor 2 [HER2]) mutations, and Kirsten rat sarcoma virus (KRAS) G12C mutations, in addition to PD-L1 expression(9). Nordic national guidelines generally align with ESMO recommendations for biomarker testing in advanced NSCLC(13, 14). This approach of molecular testing has brought about a radical change in management of mNSCLC patients, resulting in the development of novel, targeted anticancer agents such as tyrosine kinase inhibitors (TKIs)(15, 16).

3.3 Immunotherapy in mNSCLC

Advent of immune checkpoint inhibitors (ICIs: pembrolizumab, atezolizumab, nivolumab, etc.) has further brought about a radical change in the treatment of mNSCLC patients(17, 18). The use of ICIs is implemented in international and national guidelines(8). For instance, in newly diagnosed mNSCLC patients without targetable oncogenic mutations (EGFR, ALK, ROS1), ICIs are used as first-line therapy if PD-L1 expression is 50% or higher(17, 19). For those with PD-L1 expression below 50% or negative, a combination of platinum-based chemotherapy and programmed cell death protein 1 (PD-1)/PD-L1 inhibitors is generally recommended(17, 19).

3.4 Treatment Challenges and Patient Heterogeneity

While ICIs are commonly used, not all eligible individuals will benefit as these treatments are associated with some toxicity and many patients are treated for only a period of time due to progressive disease or toxicity(20). The only currently available validated biomarker with potential to predict response to ICI treatment is PD-L1, with high PD-L1 expression correlating with a higher likelihood of response(20, 21). However, PD-L1 expression has imperfect predictive accuracy: not all patients with high PD-L1 expression respond to ICI therapy, and some patients without PD-L1 expression may achieve long-term survival when treated with ICIs (20, 22). Furthermore, the European population is aging and it is uncertain whether ICIs have the same efficacy among elderly patients as the immune system among old patients is less effective than younger patients, suggesting a lower response rate and shorter duration of treatment(23). Additionally, recent research using Flatiron data from the US indicates that the median survival benefit for patients over 75 years of age is only one month, which is not as clinically significant compared to the 4–5-month benefit seen in patients under 55 years of age(24). Sex disparities in immune response also affect outcomes in patients with advanced cancers treated with ICIs, as shown in multiple clinical trials(24–26). However, the results are mixed, with some studies indicating increased survival benefits for women, while others show higher responses in men or no difference in outcomes .

3.5 Real-World Evidence Needs

Targeted therapies and ICIs are now integrated into routine clinical care for mNSCLC across both first-line and subsequent treatment settings, with substantial impacts on healthcare resource utilization and costs(27, 28). Understanding real-world treatment patterns and clinical outcomes is essential for optimizing patient selection, treatment sequencing, and resource allocation. Real-world data from routine clinical care, claims databases, and electronic health records can provide insights into treatment effectiveness, prescribing patterns, and healthcare resource implications across diverse patient populations.

Multi-institutional studies are essential for generating robust real-world evidence as they enable recruitment of larger, more diverse patient populations and enhance the generalizability of findings across different healthcare settings. Despite the proliferation of real-world evidence studies in oncology, significant methodological challenges persist in conducting multi-institutional observational research (29). Traditional approaches to multi-site studies often require centralized data aggregation, raising substantial privacy, regulatory, and technical barriers. Furthermore, heterogeneity in electronic health record systems, coding practices, and data structures across institutions limits both the feasibility and validity of cross-site comparisons. The absence of standardized data models and federated analytical frameworks represents a critical bottleneck in generating robust, multi-institutional real-world evidence, particularly in the Nordic region where universal healthcare systems provide unique opportunities for comprehensive population-level research.

3.6 Study Rationale

This pilot study was designed with dual, integrated objectives that address both methodological infrastructure and clinical knowledge gaps in Nordic oncology research. The first objective establishes proof-of-concept for federated Observational Medical Outcomes Partnership (OMOP) common data model (CDM)-based analytics across Nordic healthcare systems, demonstrating the technical feasibility, governance frameworks, and collaborative

workflows required for sustainable multi-national real-world evidence generation. The second objective characterizes treatment patterns and outcomes in mNSCLC patients receiving first-line immunotherapy, generating hypothesis-generating observations about real-world clinical practice in the Nordic context.

These objectives are deliberately synergistic: the clinical use case provides sufficient complexity to stress-test federated infrastructure capabilities, while the infrastructure development ensures that observed patterns are generated through reproducible, transparent, and privacy-preserving analytical methods. As a pilot implementation, this study prioritizes establishing methodological foundations and identifying specific data quality enhancement opportunities over definitive clinical inference. The clinical findings should be interpreted as preliminary observations that demonstrate the analytical capabilities of federated OMOP networks while simultaneously revealing the covariate capture standards and documentation harmonization required for future comparative effectiveness research capable of informing clinical practice.

3.6.1 Nordic Healthcare Context and Opportunity

The Nordic countries (Denmark, Finland, Norway, Sweden, and Iceland) share several distinctive healthcare characteristics that create a unique environment for real-world evidence generation. All five nations operate universal, tax-funded healthcare systems with comprehensive population registries and electronic medical records, enabling capture of patient encounters across specialized care settings(30). Treatment guidelines for mNSCLC across Nordic countries are based on international consensus recommendations, primarily the European Society for Medical Oncology (ESMO) clinical practice guidelines for metastatic NSCLC(31), though national implementation shows meaningful variation, and Nordic countries generally provide access to approved immunotherapy agents through tax-funded healthcare systems, though timing and availability vary between countries(32). This general similarity in healthcare access and guideline framework, combined with comparable population demographics and healthcare infrastructure, theoretically should yield similar treatment patterns and clinical outcomes(33). However, systematic comparative analyses leveraging modern data science infrastructure have been limited, and the extent to which real-world practice aligns across these seemingly similar healthcare systems remains empirically underexamined(34, 35).

3.6.2 The OMOP-Based Federated Analysis Paradigm

This study represents a pilot implementation of the OMOP CDM within a Nordic collaborative framework. The OMOP CDM, developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) community(36, 37), provides a standardized vocabulary and data structure that enables consistent analytical approaches across heterogeneous data sources(38-40). By transforming institution-specific electronic medical records into a common format, OMOP facilitates reproducible research while preserving local data governance and privacy protections(41).

The federated analysis approach implemented in this study allows each participating site to maintain full control over patient-level data, with only aggregated results shared for meta-analysis. This methodology addresses critical regulatory constraints under the European General Data Protection Regulation (GDPR) while enabling scientifically rigorous multi-site research. Importantly, the VALO pilot serves as a proof-of-concept for sustainable Nordic

research infrastructure, with participating sites developing institutional capacity in OMOP implementation, federated query execution, and collaborative governance frameworks.

3.6.3 Pilot Study Design and Learning Objectives

This study is explicitly designed as a methodological pilot study with dual objectives: (1) to characterize real-world treatment patterns and outcomes in mNSCLC patients receiving first-line immunotherapy, and (2) to evaluate the feasibility, challenges, and lessons learned from conducting federated OMOP-based research across Nordic institutions. As such, findings should be interpreted within the context of an early-phase collaborative learning exercise rather than a definitive clinical epidemiology study.

Three sites (Oslo University Hospital, Capital Region Rigshospitalet/Herlev Hospital, Helsinki University Hospital) actively participated as data providers, implementing OMOP transformations and executing federated queries. Two sites (Sweden and Iceland) participated as observers, attending workshops and contributing to governance discussions to build capacity for future research participation. This tiered participation model reflects the practical reality of heterogeneous institutional readiness for advanced data science infrastructure, while simultaneously creating a structured pathway for expanding the network in future studies.

3.6.4 Pilot Study Framework and Interpretation Guidelines

As a pilot implementation, this study operates under several acknowledged constraints that shape interpretation of findings. First, OMOP mapping maturity can vary across sites, with some variables incompletely captured or requiring proxy measures. Second, the study period (2018-2023) spans a time of rapid evolution in both immunotherapy treatment practices and institutional data infrastructure, introducing temporal heterogeneity. Third, the relatively small sample sizes—particularly for age-stratified and emergent metastatic disease subgroups—limit statistical power for certain comparisons.

Critically, observed differences between sites may reflect a complex interplay of factors including: (1) true differences in clinical practice and healthcare delivery, (2) heterogeneity in data capture and coding practices, (3) patient population differences, and (4) incomplete data extraction during this initial implementation phase. Distinguishing among these explanations requires cautious interpretation support by iterative data quality assessments and ongoing dialogue with clinical teams at each site. Future iterations of this research infrastructure will benefit from the substantial lessons learned during the pilot phase, enabling progressively more refined and clinically interpretable analyses.

The ultimate value of this pilot study extends beyond the specific clinical findings to the establishment of a sustainable, privacy-preserving research infrastructure capable of addressing diverse research questions across multiple disease areas. By demonstrating feasibility and identifying key technical and governance challenges, the VALO pilot lays the foundation for an expanded Nordic real-world evidence network with the potential to inform both clinical practice and health policy throughout the region.

As a methodological pilot study, this work establishes foundational infrastructure for Nordic collaborative research rather than providing definitive clinical evidence. The primary achievement is demonstrating that federated OMOP-based analytics can reliably generate



standardized outcome metrics across heterogeneous Nordic healthcare systems while preserving data sovereignty and regulatory compliance. This technical validation creates the capability for future confirmatory studies designed with prospectively standardized covariate capture.

Several design characteristics inherent to pilot implementation shape the interpretation framework for clinical findings. First, OMOP mapping maturity varies across sites, with some prognostic variables incompletely captured or requiring proxy measures, limiting the ability to perform adjusted comparative analyses. Second, the study period (2018-2023) spans rapid evolution in both immunotherapy treatment practices and institutional data infrastructure, introducing temporal heterogeneity that cannot be fully disentangled from geographic variation. Third, sample sizes, particularly for age-stratified and emergent metastatic disease subgroups, provide adequate power for descriptive characterization but insufficient power for definitive subgroup comparisons.

Critically, observed differences between sites reflect a complex, incompletely measured interplay of factors including: (1) true differences in clinical practice and healthcare delivery models, (2) heterogeneity in data capture completeness and coding practices, (3) patient population differences in disease severity and comorbidity burden, and (4) incomplete variable extraction during initial OMOP implementation. The absence of key prognostic variables, particularly ECOG performance status (0% capture in Rigshospitalet/Herlev Hospital and HUS), smoking history (universally missing), and inconsistent PD-L1 documentation, precludes attribution of outcome differences to specific practice patterns versus unmeasured confounding.

These limitations are tractable rather than insurmountable. The pilot identifies specific enhancement targets including standardized ECOG capture, harmonized radiotherapy documentation, and validated treatment intent classification that subsequent studies can address through prospective data quality protocols. Future Nordic OMOP studies incorporating these enhancements will be positioned to test specific hypotheses about optimal treatment strategies, generating actionable comparative effectiveness evidence for clinical practice and health policy.

4 Research Aim and Objectives

4.1 Research Aim

This pilot study had dual objectives: (1) to demonstrate the feasibility of federated OMOP-based analytics across Nordic institutions, and (2) to describe real-world treatment patterns and outcomes in metastatic NSCLC.

This study aimed to address the following research questions:

1. Investigate the treatment patterns, baseline characteristics and outcomes of patients diagnosed with mNSCLC who received an ICI as first line of treatment.
2. The healthcare and economic burden of treatment with ICIs for mNSCLC, with a focus on variations across the Nordic countries and subgroups defined by demographics.

4.2 Study Objectives

To fulfill the study aims outlined above, the following study objectives were explored and stratified by country. For each objective, subgroup analyses were conducted for patients aged 75 years and under at initiation of ICI treatment.

4.2.1 Main Objectives

1. To describe the baseline demographic and clinical characteristics of patients diagnosed with mNSCLC and receiving first line treatment for metastatic disease with ICI.
2. To analyze longitudinal treatment patterns of patients with mNSCLC:
 - a. To describe the treatment patterns of mNSCLC patients receiving ICI and chemotherapy, either as monotherapy or as a combination, including sequence and proportion of patients receiving each combination.
 - b. To evaluate the duration of treatment between sequential treatment types.
 - c. To characterize the number of cycles of treatment received (if data available).
 - d. To describe the proportion of patients that receive radiotherapy for mNSCLC post index date.
 - i. In addition, describe proportion of patients that receive cranial irradiation for mNSCLC post index date (if data available).
3. To evaluate clinical outcomes of patients with mNSCLC following 1st treatment line:
 - a. OS

4.2.2 Exploratory Objectives

4. To analyze healthcare resource utilization (HCRU) and associated cost among patients diagnosed with mNSCLC and receiving ICI as first line of treatment.
5. To contextualize ICI and chemotherapy treatment patterns (Objective 2) according to clinical guideline defined lines of therapy.

5 Research Methods

This pilot study employs a dual-objective design that integrates clinical epidemiological investigation with methodological infrastructure validation. The clinical component characterizes real-world treatment patterns and outcomes in mNSCLC patients receiving first-line immunotherapy across Nordic healthcare systems. The methodological component evaluates the feasibility, technical performance, and data quality characteristics of federated OMOP CDM-based analytics in the Nordic context.

This integrated design influences methodological choices throughout the study. First, the analytical approach prioritizes descriptive characterization over inferential analyses, reflecting both the pilot nature of the infrastructure and the incomplete capture of key prognostic covariates (ECOG performance status, smoking history, PD-L1 expression) that would be required for adjusted comparative analyses. Second, the federated analysis framework necessitates standardized analytical scripts executable across heterogeneous institutional IT environments, constraining methodological flexibility but ensuring reproducibility. Third, data completeness assessment and quality profiling are treated as primary outcomes rather than ancillary quality control measures, providing empirical evidence about Nordic OMOP network capabilities.

The methods described below address both objectives simultaneously. Cohort definitions, variable specifications, and analytical frameworks serve the clinical characterization objective while simultaneously stress-testing the federated infrastructure. Data completeness reporting fulfills both transparency requirements for clinical interpretation and methodological validation for network assessment. This dual framing is essential for appropriate interpretation of both clinical findings and infrastructure performance.

5.1 Study Design

5.1.1 Study Design Summary

This was a retrospective observational multisite study designed to assess patient characteristics, treatment patterns, and clinical outcomes in patients diagnosed with metastatic non-small cell lung cancer (mNSCLC). The study spanned from January 1, 2018, to December 31, 2023, with patient identification occurring through June 30, 2023, to ensure a minimum follow-up period of 6 months for all participants. The investigation was conducted across three major university hospitals in Denmark, Norway, and Finland, with Sweden and Iceland participating as observers. This research leveraged real-world healthcare data to provide comprehensive insights into the contemporary management of mNSCLC in the Nordic region, with a particular focus on immunotherapy treatment patterns.

The study's methodological framework was built upon a sophisticated federated analysis approach utilizing electronic medical records (EMRs) from participating university hospitals. To ensure standardization and interoperability across sites, all data were mapped to OMOP CDM. This standardization process represented a crucial foundation for maintaining data quality and analytical consistency across the diverse participating institutions.



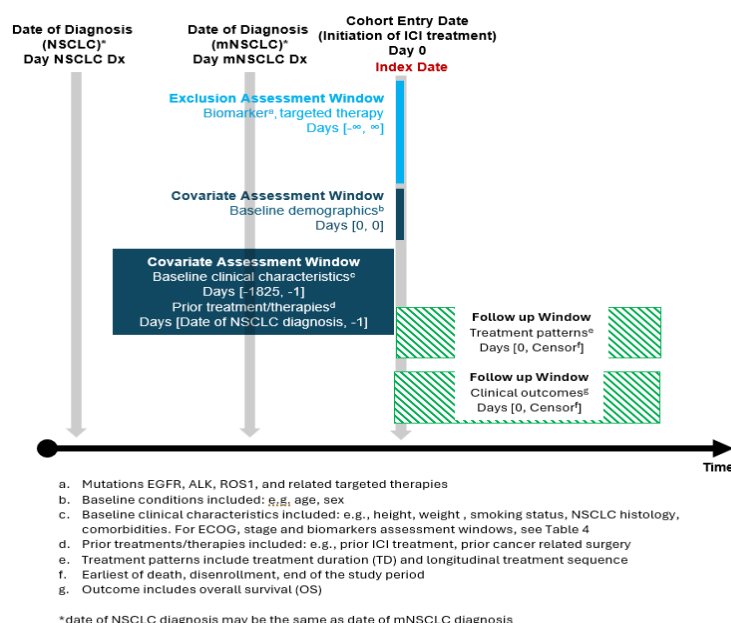
The data infrastructure employed a comprehensive management approach where patient-level information remained securely stored and de-identified at each participating site. This was facilitated through a privacy-preserving federated analytics framework that enabled complex analyses without requiring direct data pooling or transfer between institutions. Each participating site executed standardized analysis scripts locally, with only aggregated results being shared for compilation and interpretation. This methodology ensured compliance with stringent data protection regulations while maintaining the highest standards of analytical rigor.

All centers were required to contribute to Objective 1, focusing on patient baseline demographics and clinical characteristics. This mandatory participation ensured a comprehensive baseline characterization of the study population across all participating centers. However, recognizing the potential heterogeneity in data availability across sites, participation in Objectives 2-3 and the Exploratory Objectives 4-5 was flexible based on each site's ability to provide the necessary data elements. This approach maximized the utility of available data while maintaining scientific integrity.

The federated analysis framework included detailed procedures for addressing between-site heterogeneity in data availability and quality. Results were reported separately for each site. This approach ensured transparent reporting of any systematic differences between sites while maximizing the scientific value of the collected data. Sweden and Iceland's participation as observers facilitated knowledge transfer regarding data harmonization and the implementation of OMOP CDM-based studies across the Nordic region, contributing to the broader development of research infrastructure in these countries.

The study design is summarized in the study design diagram below (Figure 1).

Figure 1: Study Design Setup



5.2 Study Population

The primary study cohort *consisted of* metastatic non-small cell lung cancer (mNSCLC) patients treated with immune checkpoint inhibitors (ICI) as first-line therapy.

Index date definition: The index date *was defined as* the earliest ICI record after metastasis between January 1, 2018, and June 30, 2023. Eligible ICI ingredients *included:*

- Pembrolizumab
- Nivolumab
- Durvalumab
- Cemiplimab
- Avelumab
- Atezolizumab

Inclusion criteria: Patients *were included* if they met all of the following criteria:

- Age ≥ 18 years at index
- A record of metastasis at any time up to and including the index date ($-\infty$, 0 days)
- A record of non-small cell lung cancer (NSCLC) before or within 30 days after the first metastasis record ($-\infty$, 30 days)
- No record of ICI between first metastasis record and index (ensuring index represented start of first-line treatment)
- No record of EGFR, ALK, or ROS mutation at any time
- No targeted therapies at any time
- No non-lung primary cancer before or within 30 days after index (excluding non-melanoma skin cancers)

Follow-up: Patients *were followed* until the earliest of:

- Death
- Loss to follow-up
- End of study period (December 31, 2023)
- Diagnosis of non-lung primary cancer (excluding non-melanoma skin cancers)

5.3 Data Sources

5.3.1 OHDSI Network

The study utilized the framework of the Observational Health Data Sciences and Informatics (OHDSI) as a mechanism to deliver the study objectives. OHDSI has a proven track record in enabling large-scale robust and reproducible research in an Open Science approach, with an international network of thousands of collaborators and hundreds of data partners spanning multiple continents (42).

OHDSI provided an onboarding process for institutions conducting network studies, covering data transformation, quality assurance and training. It has a library of standard analytical methods for characterizing patient populations, estimating intervention effects, and predicting patient outcomes through probabilistic models. Additionally, OHDSI offered tools and applications facilitating network research, such as ATLAS (for cohort development), Data Quality Dashboard, CohortDiagnostics, and ARACHNE (for remote execution).

5.3.2 Systematic Use of Databases

The data sources for this study were standardized to the OMOP CDM, which specified all variables that could be collected throughout the study. This standardization enabled consistent analytics and tools across the network. The OMOP CDM is developed and maintained by OHDSI and is described in detail on the CDM wiki page and in The Book of OHDSI(43).

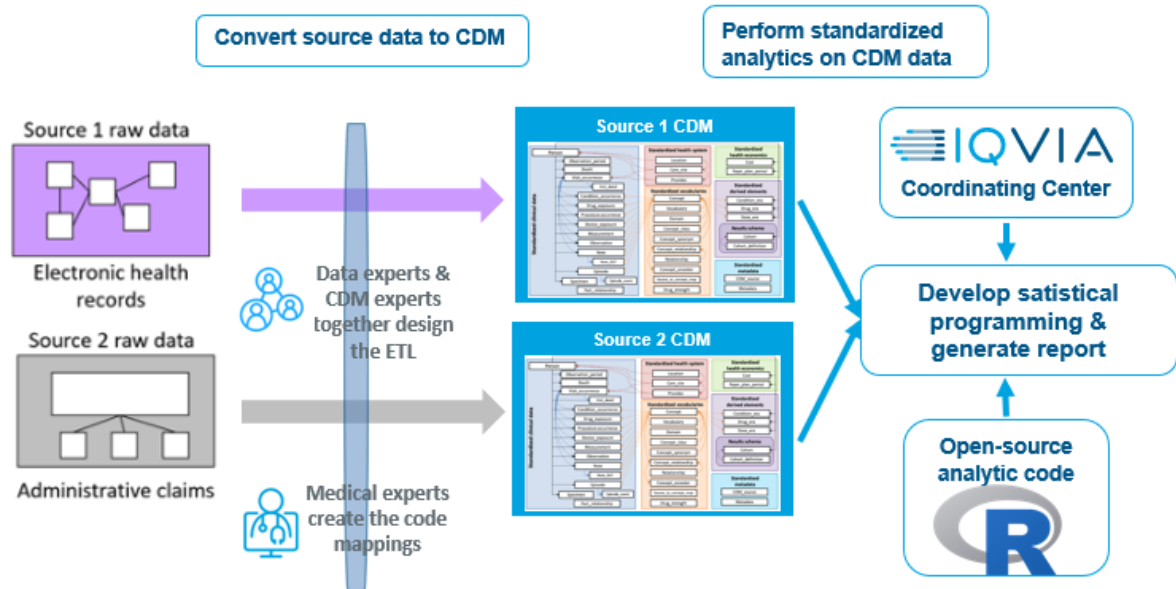
This study implemented a federated analysis approach across the Nordic countries, representing a strategic method for efficient data utilization while maintaining strict data privacy standards. Three countries - Norway, Finland, and Denmark - served as active data providers, while Sweden and Iceland participated as observers. Data were obtained from three hospital-based electronic health record systems across Oslo University Hospital (OUS), Helsinki University Hospital (HUS), and Capital Region Rigshospitalet/Herlev Hospital (Rigshospitalet/Herlev Hospital), with an estimated total sample size of approximately 4,300 patients (Table 1). This arrangement facilitated knowledge transfer and capacity building across the Nordic region, with observer nations gaining valuable insights through structured workshops and knowledge-sharing sessions focused on data harmonization and OMOP CDM implementation.

Each database custodian in OUS, HUS, and Rigshospitalet/Herlev Hospital maintained patient-identifying information securely onsite in accordance with their internal SOPs. Robust security protocols were implemented to safeguard all systems and data, ensuring that access was limited to authorized study staff only. Data storage and archiving followed appropriate procedures, including standard recovery methods for restoring files in the event of hardware or software failure.

This federated approach enabled data to remain at the institutions that collected them while avoiding the legal and technical challenges of traditional data aggregation methods. Only aggregated results were combined in tables and figures for the study report, maintaining patient privacy and information governance standards throughout the process(42).



Figure 2: Overview of how OMOP CDM enabled standardized analytics across different databases.



Abbreviations: ETL, Extract-Transform-Load

Table 1: Databases used in the study

Database	Managing organization	Country	Source population	Sample size (estimated)	Data type
Clinical data warehouse	Oslo university hospital	Norway	1 M	1,500	EHR in OMOP format
EPIC	Helsinki University Hospital	Finland	Secondary/tertiary care for 2.2M people (Uusimaa region)	1,700 patients	EHR in OMOP format
EPIC	Capital Region	Denmark	Capital Region	1,100 patients	EHR in OMOP format

5.4 Covariates and Variables

All baseline characteristics (other than staging) were assessed relative to the index date (ICI treatment initiation) (Table 3 and 4). Variables were standardized to OMOP CDM concept IDs, with analysis scripts executed locally at each site.



Table 2 Baseline demographics and clinical characteristics

Variable	Measurement Period	Definition/Categories
Demographics		
Age	At index	Continuous (years): mean, SD, median, IQR Categorical: 18-44, 45-54, 55-64, 65-74, ≥75
Sex	At index	Male, Female
Smoking status	-1825, 0 days	Current, Former*, Non-smoker, Unknown/Missing
Clinical Characteristics		
ECOG performance status	-30, 0 days	0, 1, 2, 3, 4, Unknown/Missing
BMI	-1825, 0 days	Calculated where height and weight available
Cardiovascular disease	-1825, -1 days	AMI, cardiac arrhythmia, coronary arteriosclerosis, heart failure, hypertensive disorder, PAD†
Diabetes mellitus	-1825, -1 days	Diabetes and complications†
COPD	-1825, -1 days	Chronic obstructive pulmonary disease†

Abbreviations: COPD, chronic obstructive pulmonary disease; ECOG, Eastern Cooperative Oncology Group performance status; IQR, interquartile range; PAD, peripheral arterial disease

*Former smoker defined as either recorded 'former' smoker or patient with 'non-smoker' record preceded by 'smoker' record

†Composite variables created combining diagnoses with relevant medication codes (ATC C for cardiovascular, A10 for diabetes, R03 for COPD)

Comorbidity assessment methodology: Comorbidity assessment was conducted using two complementary approaches based on modified Charlson Comorbidity Index components. The diagnosis-based approach identified comorbidities through documented International Classification of Diseases (ICD) codes in the medical records, while the medication-based approach inferred comorbidities from relevant prescription medications using Anatomical Therapeutic Chemical (ATC) classification codes (C for cardiovascular, A10 for diabetes, R03 for COPD).

This dual approach was implemented to maximize comorbidity capture across heterogeneous data sources, recognizing that electronic health records may vary in diagnostic coding completeness while consistently capturing medication prescriptions for reimbursement purposes. The methodology created composite comorbidity variables combining both diagnostic and medication-based indicators, assessed during the period from 1825 days before to 1 day before the index date (see Table 3).



Table 3 Disease characteristics and prior treatment

Variable	Measurement Period	Definition/Categories
Disease Characteristics		
Stage at NSCLC diagnosis	-60, 60 days from NSCLC diagnosis	Stage 1: T1-T2aN0M0 Stage 2: T2N1M0, T2b-T3N0M0 Stage 3: T3N1-3M0, T4NxM0, TxN2-3M0 Stage 4: Metastasis, TxNxM1‡
PD-L1 expression	Prior to index	Continuous or categorical where available
Prior Anti-Cancer Treatment§		
Platinum-based chemotherapy	NSCLC diagnosis, metastasis	Any platinum-containing regimen
Non-platinum chemotherapy	NSCLC diagnosis, metastasis	Non-platinum regimens
Immune checkpoint inhibitors	NSCLC diagnosis, metastasis	Nivolumab, Pembrolizumab, Avelumab, Atezolizumab, Durvalumab, Cemiplimab
Surgery	NSCLC diagnosis, metastasis	Wedge excision, Total pneumonectomy, Lobectomy (excludes biopsy/thoracotomy alone)
Radiotherapy	NSCLC diagnosis, metastasis	Any radiotherapy record
Time to metastasis¶	N/A	Days from NSCLC diagnosis to metastasis Mean (SD), Median (IQR)

Abbreviations: IQR, interquartile range; N/A, not applicable; NSCLC, non-small cell lung cancer; PD-L1, programmed death-ligand 1; SD, standard deviation.

‡If metastasis recorded >120 days after NSCLC diagnosis, it superseded any other stage

§Assessed for patients with emergent metastasis only

¶Calculated for emergent metastasis group only

This structured variable definition facilitated consistent data collection across sites while maintaining clarity for regulatory review. Missing data were documented as a separate category for all variables to ensure transparency in reporting.

Disease presentation classification: Patients were classified as having de novo metastatic disease if metastasis was documented within 120 days of initial NSCLC diagnosis, or emergent metastatic disease if metastasis occurred more than 120 days after diagnosis. This 120-day temporal threshold was established through clinical consultation to provide

operational consistency across sites, rather than through literature-based validation. This classification was used for stratified analyses of treatment patterns and outcomes.

5.5 Data Analysis

The analytical approach prioritizes comprehensive descriptive characterization over inferential statistical comparisons, reflecting the pilot study's dual objectives and current data completeness realities. The descriptive approach adopted here maximizes transparency about observed patterns while avoiding spurious causal inferences that would result from inadequately adjusted comparisons. This conservative analytical strategy aligns with the pilot's objective of identifying specific data quality requirements for future comparative effectiveness research rather than generating preliminary clinical practice recommendations from incompletely characterized populations.

5.5.1 Analytical Framework

All analyses were performed using standardized R scripts executed locally at each participating site through the OMOP CDM framework. Patient-level data remained at each site, with only aggregated results shared for compilation. The analytical approach followed a federated model where centrally developed scripts were distributed to and executed by each site independently.

5.5.2 Analytical Consideration per Objective

5.5.2.1 Objective 1: Baseline Characterization

Baseline demographic and clinical characteristics were analyzed using descriptive statistics:

- Continuous variables: Number of observations, mean, standard deviation, median, interquartile ranges (IQR), minimum, and maximum
- Categorical variables: Frequencies and proportions expressed as percentages

All analyses were stratified by:

- Age group (≥ 75 vs < 75 years at ICI initiation/index date)
- Country (reported separately for each site)

No patient counts below five at each site were reported, with entries marked as '<5' to maintain privacy. Missing data were documented as a separate category with counts and percentages.

5.5.2.2 Objective 2: Treatment Pattern Analysis

Treatment patterns were analyzed using two complementary approaches:

5.5.2.2.1 Longitudinal Treatment Sequences

Sequential treatment pathways were analyzed using the TreatmentPatterns R package with the implemented parameters described in Table 4.

Table 4: Treatment pattern analysis parameters

Parameter	Value	Description
MinEraDuration	1 day	Minimum duration to define a treatment era
CombinationWindow	1 day	Window to identify combination therapy
MinPostCombinationDuration	1 day	Minimum duration after combination start
FilterTreatments	"Changes"	Report only treatment transitions
MaxPathLength	3	Maximum treatment sequences tracked
Persistence Window (ICI)	365 days	Gap allowed before ICI discontinuation
Persistence Window (Chemotherapy)	91 days	Gap allowed before chemotherapy discontinuation

This analysis captured:

- Proportion of patients receiving each treatment type at 1st, 2nd, and 3rd line
- Duration of each treatment line
- Sequential treatment patterns and transitions

5.5.2.2.2 Treatment Utilization Characterization

Separate from the longitudinal sequence analysis, descriptive characterization of treatment utilization was performed to quantify:

- Proportion of patients receiving surgery, chemotherapy, or radiotherapy after metastatic diagnosis
- Proportion of patients receiving each treatment modality prior to metastatic progression (emergent population only)
- Chemotherapy treatment intensity using proxy cycle definitions

5.5.2.2.3 Chemotherapy Treatment Intensity Assessment

Due to the absence of explicit treatment intent documentation in the source databases, proxy definitions were developed as follows.

Treatment Timing Proxies (for emergent metastatic population):

- **Neoadjuvant Proxy:** Chemotherapy was classified as neoadjuvant if the patient had surgery and/or radiotherapy, and chemotherapy was administered before both
- **Adjuvant Proxy:** Chemotherapy was classified as adjuvant if the patient had surgery and/or radiotherapy, and chemotherapy was administered after both
- **Chemotherapy Only:** Chemotherapy without associated surgery or radiotherapy

These proxy definitions were necessary as no participating site had explicit documentation of adjuvant versus neoadjuvant treatment intent.

Treatment Intensity Proxies:

- **28-day gap proxy:** Chemotherapy administrations were considered part of the same cycle if they occurred within 28 days of each other. A gap exceeding 28 days indicated a new cycle
- **14-day gap proxy:** A more sensitive definition using a 14-day threshold to capture dose-dense regimens or maintenance therapy patterns

These proxy measures, developed for this study, represent treatment episodes and temporal relationships rather than documented treatment intent or planned chemotherapy cycles.

5.5.2.3 Objective 3: Survival Analysis

Overall survival was analyzed from index date (ICI initiation) using:

Kaplan-Meier (KM) Analysis

- Survival probabilities were estimated at 6, 12, 18, and 24 months
- Results were presented as:
 - KM curves with 95% confidence intervals
 - Median survival with IQR
 - Number of patients at risk at specified timepoints
 - Number and percentage of events and censored observations

Key analytical parameters:

- Follow-up: Until death or censoring (infinite days specified)
- Censoring: At cohort exit
- Stratification: Elderly flag

5.6 Data Collection Constraints and Site Implementation Variability

The methodological constraints described below reflect the inherent characteristics of pilot-phase federated network research rather than preventable study design flaws. As a proof-of-concept implementation, this study prioritized establishing foundational analytical infrastructure and identifying specific data quality enhancement targets over generating definitive comparative effectiveness evidence. The limitations documented below serve dual purposes: transparently acknowledging constraints for clinical interpretation, and systematically identifying enhancement priorities for future Nordic OMOP network studies. Each limitation represents a tractable improvement opportunity for subsequent research rather than an insurmountable barrier to federated analytics.

5.6.1 Selection Bias

OUS: OUS demonstrated a documented bias toward earlier-stage patients who are preferentially referred for curative treatment. This referral pattern may explain the high proportion of surgical candidates in the emergent population (62.5%) and could contribute to the observed survival patterns. Additionally, OUS acknowledged incomplete capture of the patient journey, as patients may receive portions of their care at other facilities not included in the database. This fragmentation particularly affects the capture of radiotherapy and later-line treatments. Critically, radiotherapy services at OUS are documented in clinical information systems not interfaced with the OMOP CDM infrastructure used for this study, resulting in 0% radiotherapy data availability.

Rigshospitalet/Herlev Hospital: The completeness of patient journey capture and potential selection biases at the Rigshospitalet/Herlev Hospital site require further evaluation. The radiotherapy data availability observed (75.5% of patients with any radiotherapy documentation) suggests these data sources were successfully interfaced with the Rigshospitalet/Herlev Hospital OMOP CDM implementation, though the extent of care fragmentation with non-participating centers remains uncertain.

HUS: Similar to Rigshospitalet/Herlev Hospital, the degree of patient journey completeness and selection bias at HUS requires clarification. The high proportion of emergent disease (14.6%) and prior chemotherapy exposure (79.3% in emergent population) may reflect either complete capture of complex patient journeys or selection of patients with more extensive treatment histories.

These site-specific selection patterns limit the generalizability of findings to community-treated populations and may explain some of the observed differences in treatment patterns and outcomes across countries. The varying completeness of patient journey capture particularly affects the interpretation of treatment sequence complexity and multimodal therapy utilization.

5.6.2 Data Standardization and Mapping Challenges

While OMOP CDM standardization facilitated cross-site analyses, the mapping process from source EHR systems to standardized concepts may have introduced information loss or misclassification. Variations in local coding practices and documentation standards across sites could have affected the comparability of certain variables, particularly for complex concepts like performance status or treatment response reasons.

5.6.3 Misclassification

Disease and Severity Classification: Some conditions may have been poorly recorded in the source databases, potentially leading to misclassification of patients. This was particularly relevant for:

- Stage at diagnosis (if recorded retrospectively)
- Distinction between de novo and emergent metastatic disease
- Biomarker status, especially when testing was not routinely performed

Outcome Ascertainment: Similar to disease classification, misclassification could have occurred in outcome determination, particularly for:

- Exact dates of progression or treatment changes
- Reasons for treatment discontinuation

Censoring Events: Information related to censoring events had varied availability across sites. Clinical trial participation data was particularly heterogeneous, with some sites capturing only intravenous trial therapies while others had complete documentation. This variability was acknowledged but could not be fully standardized across sites.

5.6.4 Federated Analysis Constraints

The federated approach, while preserving privacy, introduced analytical limitations:

- Patient-level pooling was not possible, limiting certain sophisticated analyses

- Cross-site validation of outliers or unusual patterns was restricted
- Heterogeneity assessment relied on aggregated rather than individual data
- Real-time query resolution was limited by the need for script modification and re-execution

5.6.5 Temporal and Geographic Limitations

Treatment patterns observed early in the study period may not have reflected current practice standards. Additionally, country-specific reimbursement policies and treatment guidelines may have influenced prescribing patterns independently of clinical factors.

These methodological limitations were considered in the interpretation of results and were addressed through sensitivity analyses where feasible. The transparent reporting of these limitations aimed to facilitate appropriate interpretation of findings within the context of real-world evidence generation.

5.6.6 Disease Presentation Classification Limitations

The 120-day temporal threshold used to distinguish emergent from de novo metastatic disease, while providing operational consistency, has inherent limitations. This operational definition may not fully capture biological distinctions between disease progression patterns. The threshold systematically biases toward overestimating de novo presentation in referral centers where incomplete longitudinal patient journey capture may misclassify emergent cases as de novo if the original non-metastatic diagnosis occurred outside the captured facility network. Sites lacking complete historical records will only capture patients at their metastatic presentation, misclassifying them as de novo despite potentially having prior non-metastatic NSCLC diagnoses at external facilities.

These methodological constraints, while substantial, do not diminish the pilot's primary achievement: demonstrating that federated OMOP-based analytics can generate standardized, reproducible clinical outcome metrics across heterogeneous Nordic healthcare systems. The systematic identification of specific data gaps (ECOG, smoking history, PD-L1), methodological requirements (treatment intent classification, radiotherapy documentation), and analytical limitations (inability to perform adjusted comparisons) provides an empirical roadmap for network enhancement. Future studies incorporating prospective data quality protocols addressing these identified gaps will be positioned to generate robust comparative effectiveness evidence. The technical infrastructure validated through this pilot enables such enhanced studies without requiring fundamental redesign of the federated analytical framework.

6 Results

The following results demonstrate the federated OMOP network's capability to generate standardized clinical outcome metrics across Nordic healthcare systems while simultaneously characterizing real-world treatment patterns in mNSCLC patients receiving first-line immunotherapy. As a pilot implementation, these findings serve dual purposes: validating the technical infrastructure for federated analytics and identifying specific data quality characteristics that inform future study design. Results are presented descriptively by country and objective, with cross-country comparisons and causal interpretations reserved for the Discussion section. Data completeness metrics are reported transparently throughout, as systematic documentation of variable availability represents a primary methodological outcome of this pilot study. Observed patterns should be interpreted as hypothesis-generating observations that establish feasibility of the analytical approach rather than definitive clinical evidence.

6.1 Objective 1: Cohort Assembly and Data Completeness

Sequential exclusion criteria were applied across three participating sites (HUS, Rigshospitalet/Herlev Hospital, OUS) to identify patients with metastatic non-small-cell lung cancer (mNSCLC) receiving first-line immune checkpoint inhibitor (ICI) therapy. Table 6 presents the attrition process.

The initial qualifying population comprised patients with a record of mNSCLC and at least one ICI prescription (pembrolizumab, nivolumab, durvalumab, cemiplimab, avelumab, or atezolizumab) between January 1, 2018, and June 30, 2023. From these qualifying initial records of 11,743 patients (HUS: 8,784; Rigshospitalet/Herlev Hospital: 1,124; OUS: 1,835), the following sequential criteria were applied:

1. First metastasis recorded: Retained 5,672 patients (HUS: 4,334; Rigshospitalet/Herlev Hospital: 716; OUS: 622).
2. Receipt of ICI as first-line therapy: Retained 1,069 patients (HUS: 258; Rigshospitalet/Herlev Hospital: 676; OUS: 135).
3. Age ≥ 18 years at ICI initiation: Retained 1,069 patients (HUS: 258; Rigshospitalet/Herlev Hospital: 676; OUS: 135)—no additional exclusions.
4. No documented EGFR, ALK, or ROS1 mutations: Retained 380 patients (HUS: 258; Rigshospitalet/Herlev Hospital: not reported; OUS: 122).
5. No targeted therapy use: Retained 1,024 patients (HUS: 245; Rigshospitalet/Herlev Hospital: 663; OUS: 116).
6. Cohort records collapsed (duplicate patient records merged): Retained 1,024 patients (HUS: 245; Rigshospitalet/Herlev Hospital: 663; OUS: 116).
7. No non-lung primary cancer (final study cohort): Retained 755 patients (HUS: 199; Rigshospitalet/Herlev Hospital: 489; OUS: 67).
8. Subgroup aged ≥ 75 years: 171 patients (HUS: 46; Rigshospitalet/Herlev Hospital: 111; OUS: 14).

The final analytic cohort comprised 755 patients, representing 6.4% of the initial qualifying records. The most substantial attrition occurred between initial record identification and

confirmation of first metastasis (51.7% excluded), followed by the requirement for ICI receipt as first-line therapy (81.2% of metastatic patients excluded).

Table 4: Sequential exclusion criteria applied across three participating sites

Attrition Metrics	HUS	Rigshospitalet / Herlev hospital	OUS
Qualifying initial records	8784	1124	1835
1st Metastasis	4334	716	622
ICI	258	676	135
ICI + Age => 18	258	676	135
No positive record of EGFR, ALK or ROS1 mutations	258	-	122
No treatment with targeted therapies	245	663	116
Cohort records collapsed	245	663	116
No non-lung primary cancer (study cohort)	199	489	67
>= 75 (subgroup cohort)	46	111	14

6.1.1 Data Completeness and Quality Assessment

Critical prognostic covariates showed heterogeneity in availability across participating sites.

Performance status (ECOG PS) was absent (0.0% capture) at the Rigshospitalet/Herlev Hospital and HUS sites, as confirmed during initial data profiling. Oslo University Hospital (OUS) was the only site with ECOG PS documentation, available for 36 of 67 patients (53.7%), with 31 patients (46.3%) having unknown or missing values.

Smoking status was universally unavailable across all three Nordic sites, with 100.0% missing data for all 755 patients in the final cohort.

PD-L1 expression, the primary biomarker guiding ICI therapy selection, could not be harmonized across sites due to heterogeneous testing platforms, incompatible reporting thresholds, and substantial missing data. OUS captured limited PD-L1 data with small patient numbers (fewer than 10 per category). Rigshospitalet/Herlev Hospital documented PD-L1 $\geq 50\%$ expression in 273 patients (55.8%), but heterogeneous reporting formats (e.g., "0%," "1%," "1-25%," " $\geq 50\%$," "No matching concept") prevented consistent categorization. HUS had no PD-L1 expression data available in the extracted dataset.

Body mass index completeness varied: 73.1% in OUS (49/67 patients), 77.9% in Rigshospitalet/Herlev Hospital (381/489), and data availability was reported but specific percentages were not stated for HUS.

Comorbidity assessment revealed near-universal cardiovascular medication use (83.6% OUS, 97.8% Denmark, 99.5% HUS) contrasted sharply with diagnosis-based cardiovascular disease prevalence (20.9%, 23.9%, 26.9% respectively). COPD demonstrated perfect diagnosis-medication concordance in Rigshospitalet/Herlev Hospital and HUS, with

prevalence of 11.9% (Denmark), 13.4% (OUS), and 18.6% (HUS). Diabetes medication prevalence ranged from 10.4% (OUS) to 23.6% (HUS).

6.2 Objective 1: Baseline Demographic and Clinical Characteristics

6.2.1 Study Population - OUS

A total of 67 patients with mNSCLC receiving first-line ICI treatment were identified at Oslo University Hospital (OUS) during the study period (January 2018 - June 2023) (Table 7). The cohort was stratified by age, with 53 patients (79.1%) aged less than 75 years and 14 patients (20.9%) aged 75 years or older.

6.2.1.1 Demographic Characteristics

Table 5: Baseline demographics – OUS

Characteristic	Overall (N=67)	Age <75 (n=53)	Age ≥75 (n=14)
Age, years			
Mean (SD)	66.8 (8.8)	64.0 (7.7)	77.6 (2.0)
Median IQR	68 59-74	65 58-72	77 76-79
Range	45-81	45-74	75-81
Age groups, n (%)			
45-54	7 (10.4)	7 (13.2)	-
55-64	18 (26.9)	18 (34.0)	-
65-74	28 (41.8)	28 (52.8)	-
≥75	14 (20.9)	-	14 (100.0)
Sex, n (%)			
Male	46 (68.7)	38 (71.7)	8 (57.1)
Female	21 (31.3)	15 (28.3)	6 (42.9)

The overall cohort demonstrated a mean age of 66.8 years (SD 8.8) with a median of 68 years (IQR 59-74), ranging from 45 to 81 years (Table 7). The distribution by age group was: 10.4% aged 45-54, 26.9% aged 55-64, 41.8% aged 65-74, and 20.9% aged 75 or older. The younger cohort (less than 75 years) had a mean age of 64.0 years (SD 7.7), while the elderly cohort (75 years or older) showed a mean of 77.6 years (SD 2.0) with a narrower age range of 75-81 years.

Sex distribution showed male predominance overall (68.7% male, 31.3% female). This male predominance was more pronounced in the younger cohort (71.7% male) compared to the older cohort (57.1% male). Female representation was 28.3% in the less than 75 years group and 42.9% in the 75 years or older group.

6.2.1.2 Clinical Characteristics

Table 6: Clinical characteristics - OUS

Characteristic	Overall (N=67)	Age <75 (n=53)	Age ≥75 (n=14)
ECOG PS, n (%)*			
0	<5†	<5†	0 (0.0)
1	20 (29.9)	14 (26.4)	6 (42.9)
2	11 (16.4)	8 (15.1)	<5†
3	<5†	<5†	0 (0.0)
4	0 (0.0)	0 (0.0)	0 (0.0)
Unknown/Missing	31 (46.3)	26 (49.1)	5 (35.7)
Disease status, n (%)			
De novo metastatic	59 (88.1)	50 (94.3)	9 (64.3)
Emergent metastatic	8 (11.9)	<5†	5 (35.7)
**Stage at diagnosis, n (%)††			
Stage I	0 (0.0)	0 (0.0)	0 (0.0)
Stage II	0 (0.0)	0 (0.0)	0 (0.0)
Stage III	0 (0.0)	0 (0.0)	0 (0.0)
Stage IV	59 (88.1)	50 (94.3)	9 (64.3)
Unknown/Missing	8 (11.9)	<5†	5 (35.7)
Smoking status, n (%)			
Current smoker	0 (0.0)	0 (0.0)	0 (0.0)
Former smoker	0 (0.0)	0 (0.0)	0 (0.0)
Never smoker	0 (0.0)	0 (0.0)	0 (0.0)
Unknown/Missing	67 (100.0)	53 (100.0)	14 (100.0)

*ECOG PS data available only from Oslo University Hospital among participating Nordic sites

**Classification based on temporal threshold of 120 days between initial NSCLC diagnosis and documented metastasis

††Patients with metastasis within 120 days of NSCLC diagnosis were automatically classified as Stage IV

†Values <5 masked for patient privacy

Performance status documentation was available for 36 patients (53.7%), with 31 patients (46.3%) having unknown or missing ECOG PS (Table 8). OUS was the only participating site with available ECOG performance status data. Among patients with documented ECOG PS, 20 patients (29.9%) had ECOG 1 and 11 patients (16.4%) had ECOG 2. Fewer than 5 patients had ECOG 0 or ECOG 3 (values masked for privacy). In the age-stratified analysis,

ECOG 1 was documented in 26.4% of the <75 years group and 42.9% of the ≥75 years group. ECOG 2 was documented in 15.1% of the younger cohort, while the corresponding value was not available for the older cohort. The proportion with unknown/missing ECOG PS was 49.1% in the <75 years group and 35.7% in the ≥75 years group.

Disease presentation patterns showed that 88.1% of patients presented with de novo metastatic disease, while 11.9% had emergent metastatic disease. Classification of metastatic presentation pattern was based on a temporal threshold of 120 days between initial NSCLC diagnosis and documented metastasis. In the younger cohort, 94.3% presented with de novo metastatic disease with fewer than 5 emergent cases (masked for privacy). In contrast, the older cohort showed 64.3% with de novo metastatic disease and 35.7% with emergent metastatic disease.

Stage IV disease at diagnosis was documented in 88.1% of all patients, with staging information missing for 11.9%. The stage assignment algorithm automatically classified any patient with metastasis within 120 days of NSCLC diagnosis as Stage IV, regardless of any other recorded stage information. This distribution exactly mirrored the disease status classification pattern. In the age-stratified analysis, the younger cohort showed 94.3% with Stage IV disease and fewer than 5 patients with missing stage information (masked for privacy), while the elderly cohort had 64.3% with Stage IV disease and 35.7% with missing staging.

Body mass index data was available for 49 patients (73.1% of the cohort). The overall mean BMI was 24.9 kg/m² (SD 4.4) with a median of 25.1 kg/m² (IQR 22.2-27.1). The younger cohort (n=38) showed a mean BMI of 24.7 kg/m² (SD 4.3), while the older cohort (n=11) had a mean BMI of 25.9 kg/m² (SD 4.7).

Smoking status data was unavailable for all 67 patients (100.0%), representing a complete absence of this clinical variable in the dataset.

6.2.1.3 Comorbidity Profile

Table 7: Comorbidity profile - OUS

Comorbidity*, n (%)	Overall (N=67)	Age <75 (n=53)	Age ≥75 (n=14)
Diabetes			
Diagnosis-based	<5†	<5†	0 (0.0)
Medication-based	7 (10.4)	7 (13.2)	0 (0.0)
Cardiovascular disease			
Diagnosis-based	14 (20.9)	8 (15.1)	6 (42.9)
Medication-based	56 (83.6)	45 (84.9)	11 (78.6)
COPD			
Diagnosis-based	9 (13.4)	7 (13.2)	<5†
Medication-based	9 (13.4)	7 (13.2)	<5†

*Comorbidities identified using modified Charlson Comorbidity Index components

†Values <5 masked for patient privacy

Diabetes assessment revealed minimal burden across the cohort (Table 9). Fewer than 5 patients had diagnosis-based diabetes (masked for privacy). Medication-based diabetes was present in 7 patients (10.4%) overall, all within the younger cohort (13.2% of <75 years group), with no diabetes medications documented in the ≥75 years group.

Cardiovascular comorbidity showed high prevalence with notable discrepancy between assessment methods. Diagnosis-based cardiovascular disease was identified in 14 patients (20.9%) overall, with age-stratified rates of 15.1% in the <75 years group and 42.9% in the ≥75 years group. Cardiovascular medication use was documented in 56 patients (83.6%) overall, comprising 84.9% of the younger cohort and 78.6% of the older cohort.

COPD was identified in 9 patients (13.4%) overall, both by diagnosis and medication criteria. In the age-stratified analysis, COPD (both diagnosis and medication-based) was documented in 7 patients (13.2%) in the <75 years group. Fewer than 5 elderly patients had COPD (values masked for privacy).

6.2.1.4 Subgroup Analysis: Emergent Metastatic Disease

6.2.1.4.1 Emergent Metastatic Population Characteristics

Among the 67 patients in the OUS cohort, 8 patients (11.9%) had emergent metastatic disease, defined as progression to metastatic disease following initial non-metastatic NSCLC diagnosis. The remaining 59 patients (88.1%) presented with de novo metastatic disease at initial diagnosis.

6.2.1.4.2 Time to Metastatic Progression Patterns

Table 8: Time to metastatic progression in emergent population - OUS

Characteristic	Overall (n=8)	Age <75 *	Age ≥75 (n=5)
Time to metastasis, days			
Mean (SD)	367.1 (200.1)	-	379.6 (258.8)
Median (IQR)	310 (240-440)	-	244 (229-501)
Range	147-777	-	147-777
Time to metastasis, months			
Mean (SD)	12.1 (6.6)	-	12.5 (8.5)
Median (IQR)	10.2 (7.9-14.5)	-	8.0 (7.5-16.5)

* Values <5 masked for patient privacy

Note: Time to metastasis measured from initial NSCLC diagnosis to documented metastatic disease

The emergent metastatic population demonstrated substantial heterogeneity in time to metastatic progression (Table 10). Overall, patients progressed to metastatic disease at a mean of 367.1 days (SD 200.1) or approximately 12.1 months from initial NSCLC diagnosis.

The median time to metastasis was 310 days (IQR 240-440), with a wide range from 147 to 777 days (4.8 to 25.6 months).

Among the five elderly patients (≥ 75 years) with emergent disease, time to metastatic progression showed considerable variability (mean 379.6 days, SD 258.8). The median time to progression was 244 days, with an interquartile range of 229-501 days and the same overall range of 147 to 777 days.

6.2.1.4.3 Metastatic Presentation Patterns by Age

Table 9: Distribution of metastatic presentation by age group - OUS

Disease Presentation	Overall (N=67) n (%)	Age <75 (n=53) n (%)	Age ≥ 75 (n=14) n (%)
De novo metastatic	59 (88.1)	50 (94.3)	9 (64.3)
Emergent metastatic	8 (11.9)	-	5 (35.7)

The distribution of metastatic presentation showed pronounced age-related differences (Table 11). Among the elderly cohort (≥ 75 years), emergent metastatic disease comprised 35.7% (5/14) of cases, compared to de novo metastatic disease in 64.3% (9/14). The younger cohort (<75 years) showed 94.3% with de novo metastatic disease, with emergent cases masked (<5 patients). Despite the elderly group representing only 20.9% of the overall cohort, this population showed a disproportionately high rate of emergent disease (35.7% vs <5.7% in younger patients).

6.2.2 Study Population - Rigshospitalet/Herlev Hospital

A total of 489 patients with metastatic NSCLC receiving first-line ICI treatment were identified in the Rigshospitalet/Herlev Hospital cohort during the study period (January 2018 - June 2023) (Table 12). Of these, 378 patients (77.3%) were aged less than 75 years, and 111 patients (22.7%) were aged 75 years or older.

6.2.2.1 Demographic Characteristics

Table 10: Baseline demographics – Rigshospitalet/Herlev Hospital

Characteristic	Overall (N=489)	Age <75 (n=378)	Age ≥ 75 (n=111)
Age, years			
Mean (SD)	67.1 (9.2)	63.8 (7.8)	78.2 (3.1)
Median (IQR)	68 (61-74)	65 (59-70)	77 (76-79)
Range	36-90	36-74	75-90
Age groups, n (%)			
18-44	6 (1.2)	6 (1.6)	-
45-54	42 (8.6)	42 (11.1)	-
55-64	129 (26.4)	129 (34.1)	-



Characteristic	Overall (N=489)	Age <75 (n=378)	Age ≥75 (n=111)
65-74	201 (41.1)	201 (53.2)	-
≥75	111 (22.7)	-	111 (100.0)
Sex, n (%)			
Male	220 (45.0)	162 (42.9)	58 (52.3)
Female	269 (55.0)	216 (57.1)	53 (47.7)

The overall Rigshospitalet/Herlev Hospital cohort demonstrated a mean age of 67.1 years (SD 9.2) with a median of 68 years (IQR 61-74), ranging from 36 to 90 years (Table 12). Age distribution analysis revealed that 41.1% of patients were in the 65–74-year age group, representing the largest single age category, followed by 26.4% in the 55–64-year group. The cohort included a small proportion of younger patients, with 1.2% aged 18-44 years and 8.6% aged 45-54 years. The younger cohort (<75 years) had a mean age of 63.8 years (SD 7.8) with a median of 65 years (IQR 59-70). The elderly cohort (≥75 years) showed a mean age of 78.2 years (SD 3.1) with a median of 77 years (IQR 76-79) and a narrower age distribution compared to the younger group.

Sex distribution showed female predominance overall (55.0% female, 45.0% male). This female predominance was more pronounced in the younger cohort (57.1% female, 42.9% male). The elderly cohort demonstrated a more balanced distribution with a slight male predominance (52.3% male, 47.7% female), representing a reversal of the sex distribution pattern compared to the younger age group.

Table 11: Clinical characteristics – Rigshospitalet/Herlev Hospital

Characteristic	Overall (N=489)	Age <75 (n=378)	Age ≥75 (n=111)
ECOG PS, n (%)*			
0	0 (0.0)	0 (0.0)	0 (0.0)
1	0 (0.0)	0 (0.0)	0 (0.0)
2	0 (0.0)	0 (0.0)	0 (0.0)
3	0 (0.0)	0 (0.0)	0 (0.0)
4	0 (0.0)	0 (0.0)	0 (0.0)
Unknown/Missing	489 (100.0)	378 (100.0)	111 (100.0)
Disease status, n (%)			
De novo metastatic	451 (92.2)	350 (92.6)	101 (91.0)
Emergent metastatic	38 (7.8)	28 (7.4)	10 (9.0)
Stage at diagnosis, n (%)†			
Stage I	0 (0.0)	0 (0.0)	0 (0.0)



Characteristic	Overall (N=489)	Age <75 (n=378)	Age ≥75 (n=111)
Stage II	0 (0.0)	0 (0.0)	0 (0.0)
Stage III	0 (0.0)	0 (0.0)	0 (0.0)
Stage IV	451 (92.2)	350 (92.6)	101 (91.0)
Unknown/Missing	38 (7.8)	28 (7.4)	10 (9.0)

*ECOG PS not captured at the Rigshospitalet/Herlev Hospital site per initial data profiling

†Patients with metastasis within 120 days of NSCLC diagnosis were automatically classified as Stage IV

6.2.2.2 Clinical Characteristics

Performance status documentation was completely absent in the Rigshospitalet/Herlev Hospital cohort, with all 489 patients (100.0%) having unknown or missing ECOG PS data (Table 13). This absence was expected as the Rigshospitalet/Herlev Hospital site indicated during initial data profiling that ECOG performance status was not captured in their database.

Disease presentation patterns showed that 451 patients (92.2%) presented with de novo metastatic disease, while 38 patients (7.8%) had emergent metastatic disease. In the younger cohort, 350 patients (92.6%) presented with de novo metastatic disease and 28 patients (7.4%) with emergent disease. The elderly cohort showed similar distribution with 101 patients (91.0%) having de novo metastatic disease and 10 patients (9.0%) with emergent disease. The proportion of emergent disease was slightly higher in the elderly group compared to the younger cohort (9.0% vs 7.4%).

Stage IV disease at diagnosis was documented in 451 patients (92.2%) overall, with staging information missing for 38 patients (7.8%). The stage distribution exactly mirrored the disease status classification, consistent with the algorithmic assignment of Stage IV to patients with metastasis within the defined temporal window. Age-stratified staging showed 350 patients (92.6%) with Stage IV disease in the <75 years group and 101 patients (91.0%) in the ≥75 years group.

Body mass index data was available for 381 patients (77.9% of the cohort). The overall mean BMI was 24.7 kg/m² (SD 6.5) with a median of 23.8 kg/m² (IQR 21.5-26.8). The younger cohort (n=297) showed a mean BMI of 24.6 kg/m² (SD 4.5) with a median of 23.8 kg/m² (IQR 21.5-27.1). The elderly cohort (n=84) had a mean BMI of 25.1 kg/m² (SD 11.1) with a median of 24.2 kg/m² (IQR 21.4-26.3). The standard deviation in the elderly group (11.1) was higher compared to the younger cohort (4.5).

Table 12: Comorbidity profile – Rigshospitalet/Herlev Hospital

Comorbidity*, n (%)	Overall (N=489)	Age <75 (n=378)	Age ≥75 (n=111)
Diabetes			
Diagnosis-based	27 (5.5)	22 (5.8)	5 (4.5)
Medication-based	54 (11.0)	46 (12.2)	8 (7.2)
Cardiovascular disease			



Comorbidity*, n (%)	Overall (N=489)	Age <75 (n=378)	Age ≥75 (n=111)
Diagnosis-based	117 (23.9)	84 (22.2)	33 (29.7)
Medication-based	478 (97.8)	369 (97.6)	109 (98.2)
COPD			
Diagnosis-based	58 (11.9)	42 (11.1)	16 (14.4)
Medication-based	58 (11.9)	42 (11.1)	16 (14.4)

*Comorbidities identified using modified Charlson Comorbidity Index components. Assessment was conducted using two complementary approaches: diagnosis-based (ICD codes) and medication-based

6.2.2.3 Comorbidity Profile

Comorbidity assessment revealed distinct patterns across disease categories in the Rigshospitalet/Herlev Hospital cohort. Diabetes was identified through diagnosis in 27 patients (5.5%) overall, with age-specific rates of 5.8% in the younger cohort and 4.5% in the elderly cohort. Medication-based diabetes assessment identified 54 patients (11.0%) overall, comprising 46 patients (12.2%) in the <75 years group and 8 patients (7.2%) in the ≥75 years group. The discrepancy between diagnosis-based and medication-based assessment was less pronounced for diabetes compared to other comorbidities, with medication-based identification approximately doubling the diagnosis-based rate.

Cardiovascular comorbidity showed marked prevalence with substantial discrepancy between assessment methods. Diagnosis-based cardiovascular disease was identified in 117 patients (23.9%) overall, with age-stratified rates of 22.2% in the <75 years group and 29.7% in the ≥75 years group. Cardiovascular medication use was nearly universal, documented in 478 patients (97.8%) overall, comprising 369 patients (97.6%) of the younger cohort and 109 patients (98.2%) of the elderly cohort.

COPD demonstrated complete concordance between diagnosis-based and medication-based assessment, with 58 patients (11.9%) identified by both methods overall. Age-stratified analysis showed 42 patients (11.1%) with COPD in the younger cohort and 16 patients (14.4%) in the elderly cohort by both assessment methods.

6.2.2.4 Subgroup Analysis: Emergent Metastatic Disease

6.2.2.4.1 Emergent Metastatic Population Characteristics

Among the 489 patients in the Rigshospitalet/Herlev Hospital cohort, 38 patients (7.8%) had emergent metastatic disease, defined as progression to metastatic disease following initial non-metastatic NSCLC diagnosis. The remaining 451 patients (92.2%) presented with de novo metastatic disease at initial diagnosis.

Table 13: Time to Metastatic Progression in Emergent Population - Rigshospitalet/Herlev Hospital

Characteristic	Overall (n=38)	Age <75 (n=28)	Age ≥75 (n=10)
Time to metastasis, days			
Mean (SD)	523.0 (391.9)	526.8 (390.8)	512.2 (415.9)
Median (IQR)	324 (260-813)	338 (281-762)	312 (254-814)
Range	122-1615	122-1615	126-1243

6.2.2.4.2 Time to Metastatic Progression Patterns

The emergent metastatic population in Rigshospitalet/Herlev Hospital demonstrated substantial heterogeneity in time to metastatic progression (Table 15). Overall, patients progressed to metastatic disease at a mean of 523.0 days (SD 391.9) from initial NSCLC diagnosis. The median time to metastasis was 324 days (IQR 260-813), with an extremely wide range from 122 to 1615 days, indicating progression occurring from approximately 4 months to over 4 years after initial diagnosis.

Age-stratified analysis revealed similar progression patterns between age groups. The younger cohort (<75 years) comprising 28 patients showed a mean time to progression of 526.8 days (SD 390.8) with a median of 338 days (IQR 281-762). The range extended from 122 to 1615 days, with the maximum value indicating at least one patient progressing more than 4 years after initial diagnosis. The elderly cohort (≥75 years) of 10 patients demonstrated a mean time to progression of 512.2 days (SD 415.9), slightly shorter than the younger group, with a median of 312 days (IQR 254-814). The range in the elderly group was 126 to 1243 days.

The interquartile ranges showed considerable overlap between age groups, with both demonstrating wide dispersion. The 25th percentile was 281 days in the younger group versus 254 days in the elderly group, while the 75th percentile was 762 days versus 814 days, respectively.

Table 14: Distribution of metastatic presentation by age group - Rigshospitalet/Herlev Hospital

Disease Presentation	Overall (N=489) n (%)	Age <75 (n=378) n (%)	Age ≥75 (n=111) n (%)
De novo metastatic	451 (92.2)	350 (92.6)	101 (91.0)
Emergent metastatic	38 (7.8)	28 (7.4)	10 (9.0)

6.2.2.4.3 Metastatic Presentation Patterns by Age

The distribution of metastatic presentation in the Rigshospitalet/Herlev Hospital cohort showed high predominance of de novo disease across all age groups (Table 16). Overall, 92.2% of patients presented with de novo metastatic disease. The younger cohort showed 92.6% with de novo disease and 7.4% with emergent disease. The elderly cohort

demonstrated a slightly higher proportion of emergent disease at 9.0% compared to 7.4% in younger patients.

Among the 38 emergent cases, 28 (73.7%) occurred in patients aged <75 years and 10 (26.3%) in those ≥75 years. Given that the elderly group represented 22.7% of the overall cohort, the 26.3% representation in emergent cases indicates only a modest over-representation of elderly patients in the emergent population.

6.2.3 Study Population - HUS

A total of 199 patients with metastatic NSCLC receiving first-line ICI treatment were identified at Helsinki University Hospital (HUS) during the study period (January 2018 - June 2023) (Table 17). Of these, 153 patients (76.9%) were aged less than 75 years, and 46 patients (23.1%) were aged 75 years or older.

Table 15: Baseline demographics - HUS

Characteristic	Overall (N=199)	Age <75 (n=153)	Age ≥75 (n=46)
Age, years			
Mean (SD)	67.0 (8.5)	63.8 (7.0)	77.5 (3.0)
Median (IQR)	68 (60-74)	64 (59-69)	76 (75-78)
Range	42-87	42-74	75-87
Age groups, n (%)			
18-44	-	-	-
45-54	10 (5.0)	10 (6.5)	-
55-64	64 (32.2)	64 (41.8)	-
65-74	76 (38.2)	76 (49.7)	-
≥75	46 (23.1)	-	46 (100.0)
Sex, n (%)			
Male	112 (56.3)	91 (59.5)	21 (45.7)
Female	87 (43.7)	62 (40.5)	25 (54.3)

6.2.3.1 Demographic Characteristics

The overall HUS cohort demonstrated a mean age of 67.0 years (SD 8.5) with a median of 68 years (IQR 60-74), ranging from 42 to 87 years. Age distribution analysis revealed that 38.2% of patients were in the 65–74-year age group, representing the largest single age category, followed by 32.2% in the 55–64-year group. The cohort included a small proportion of younger patients, with 5.0% aged 45-54 years and fewer than 5 patients aged 18-44 years (masked for privacy). The younger cohort (<75 years) had a mean age of 63.8 years (SD 7.0) with a median of 64 years (IQR 59-69). The elderly cohort (≥75 years) showed a mean age of 77.5 years (SD 3.0) with a median of 76 years (IQR 75-78), demonstrating a narrow age distribution.



Sex distribution showed male predominance overall (56.3% male, 43.7% female). The younger cohort demonstrated more pronounced male predominance (59.5% male, 40.5% female). The elderly cohort showed a reversal with female predominance (54.3% female, 45.7% male), representing a distinct age-related sex distribution pattern.

Table 16: Clinical characteristics - HUS

Characteristic	Overall (N=199)	Age <75 (n=153)	Age ≥75 (n=46)
ECOG PS, n (%)*			
0	0 (0.0)	0 (0.0)	0 (0.0)
1	0 (0.0)	0 (0.0)	0 (0.0)
2	0 (0.0)	0 (0.0)	0 (0.0)
3	0 (0.0)	0 (0.0)	0 (0.0)
4	0 (0.0)	0 (0.0)	0 (0.0)
Unknown/Missing	199 (100.0)	153 (100.0)	46 (100.0)
Disease status, n (%)			
De novo metastatic	170 (85.4)	130 (85.0)	40 (87.0)
Emergent metastatic	29 (14.6)	23 (15.0)	6 (13.0)
Stage at diagnosis, n (%)†			
Stage I	<5†	<5†	0 (0.0)
Stage II	5 (2.5)	<5†	<5†
Stage III	13 (6.5)	12 (7.8)	<5†
Stage IV	170 (85.4)	130 (85.0)	40 (87.0)
Unknown/Missing	10 (5.0)	8 (5.2)	<5†
Smoking status, n (%)			
Current smoker	0 (0.0)	0 (0.0)	0 (0.0)
Former smoker	0 (0.0)	0 (0.0)	0 (0.0)
Never smoker	0 (0.0)	0 (0.0)	0 (0.0)
Unknown/Missing	199 (100.0)	153 (100.0)	46 (100.0)

*ECOG PS not captured at the HUS site per data availability

†Patients with metastasis within 120 days of NSCLC diagnosis were automatically classified as Stage IV

†Values <5 masked for patient privacy

6.2.3.2 Clinical Characteristics

Performance status documentation was completely absent in the HUS cohort, with all 199 patients (100.0%) having unknown or missing ECOG PS data (Table 17). Disease



presentation patterns showed that 170 patients (85.4%) presented with de novo metastatic disease, while 29 patients (14.6%) had emergent metastatic disease. In the younger cohort, 130 patients (85.0%) presented with de novo metastatic disease and 23 patients (15.0%) with emergent disease. The elderly cohort showed 40 patients (87.0%) with de novo metastatic disease and 6 patients (13.0%) with emergent disease.

Stage distribution revealed a unique pattern compared to other Nordic sites. While Stage IV disease predominated at 85.4% overall, the HUS cohort included fewer than 5 patients with Stage I (masked), 5 patients (2.5%) with Stage II and 13 patients (6.5%) with Stage III disease at diagnosis, suggesting inclusion of some patients who progressed from earlier stages. Age-stratified staging showed 130 patients (85.0%) with Stage IV disease, fewer than 5 with Stage II (masked), 12 patients (7.8%) with Stage III, and 8 patients (5.2%) with unknown staging in the <75 years group. The elderly cohort showed 40 patients (87.0%) with Stage IV disease, with Stage II, Stage III, and unknown staging all masked (<5 patients each).

Body mass index data was available for 194 patients (97.5% of the cohort), representing the highest BMI data completeness among the three Nordic sites. The overall mean BMI was 25.0 kg/m² (SD 4.4) with a median of 24.1 kg/m² (IQR 22.0-27.8). The younger cohort (n=150) showed a mean BMI of 25.1 kg/m² (SD 4.7) with a median of 24.4 kg/m² (IQR 22.0-27.9). The elderly cohort (n=44) had a slightly lower mean BMI of 24.4 kg/m² (SD 3.5) with a median of 23.7 kg/m² (IQR 22.0-27.2). The HUS elderly cohort demonstrated lower variability (SD 3.5 vs 4.7 in younger patients). The median BMI values across all groups fell within the normal to slightly overweight range.

Table 17: Comorbidity profile - HUS

Comorbidity*, n (%)	Overall (N=199)	Age <75 (n=153)	Age ≥75 (n=46)
Diabetes			
Diagnosis-based	13 (6.5)	8 (5.2)	5 (10.9)
Medication-based	47 (23.6)	32 (20.9)	15 (32.6)
Cardiovascular disease			
Diagnosis-based	65 (32.7)	42 (27.5)	23 (50.0)
Medication-based	198 (99.5)	152 (99.3)	46 (100.0)
COPD			
Diagnosis-based	37 (18.6)	27 (17.6)	10 (21.7)
Medication-based	37 (18.6)	27 (17.6)	10 (21.7)

*Comorbidities identified using modified Charlson Comorbidity Index components using two complementary approaches: diagnosis-based (ICD codes) and medication-based (ATC codes: A10 for diabetes, C for cardiovascular, R03 for COPD)

6.2.3.3 Comorbidity Profile

Diabetes was identified through diagnosis in 13 patients (6.5%) overall, with age-specific rates of 5.2% in the younger cohort and 10.9% in the elderly cohort, showing an expected age-related increase. Medication-based diabetes assessment identified 47 patients (23.6%) overall, comprising 32 patients (20.9%) in the <75 years group and 15 patients (32.6%) in

the ≥ 75 years group. The elderly cohort showed a notably higher diabetes burden by both measures.

Cardiovascular comorbidity demonstrated the familiar pattern of extreme discrepancy between assessment methods. Diagnosis-based cardiovascular disease was identified in 65 patients (32.7%) overall, with marked age-related differences: 27.5% in the < 75 years group versus 50.0% in the ≥ 75 years group. This near doubling of cardiovascular disease prevalence in elderly patients represents the expected age-related cardiovascular risk. Cardiovascular medication use was virtually universal at 198 patients (99.5%) overall, comprising 152 patients (99.3%) of the younger cohort and 46 patients (100.0%) of the elderly cohort.

COPD demonstrated complete concordance between diagnosis-based and medication-based assessment, with 37 patients (18.6%) identified by both methods overall. Age-stratified analysis showed 27 patients (17.6%) with COPD in the younger cohort and 10 patients (21.7%) in the elderly cohort by both assessment methods.

6.2.3.4 Subgroup Analysis: Emergent Metastatic Disease

6.2.3.4.1 Emergent Metastatic Population Characteristics

Among the 199 patients in the HUS cohort, 29 patients (14.6%) had emergent metastatic disease, defined as progression to metastatic disease following initial non-metastatic NSCLC diagnosis using a 120-day temporal threshold. The remaining 170 patients (85.4%) presented with de novo metastatic disease at initial diagnosis.

Table 18: Time to metastatic progression in emergent population - HUS

Characteristic	Overall (n=29)	Age < 75 (n=23)	Age ≥ 75 (n=6)
Time to metastasis, days			
Mean (SD)	463.7 (424.9)	456.4 (459.1)	491.7 (287.1)
Median (IQR)	286 (153-615)	279 (148-498)	513 (254-748)
Range	125-1959	125-1959	139-791

6.2.3.4.2 Time to Metastatic Progression Patterns

The emergent metastatic population in HUS demonstrated substantial heterogeneity in time to metastatic progression (Table 20). Overall, patients progressed to metastatic disease at a mean of 463.7 days (SD 424.9) from initial NSCLC diagnosis. The median time to metastasis was 286 days (IQR 153-615), with an extremely wide range from 125 to 1959 days. The maximum value of 1959 days indicates at least one patient progressing more than 5 years after initial diagnosis, the longest progression time observed among all three Nordic sites.

Age-stratified analysis revealed differential progression patterns. The younger cohort (< 75 years) comprising 23 patients showed a mean time to progression of 456.4 days (SD 459.1) with a median of 279 days (IQR 148-498). The extremely high standard deviation exceeding the mean reflects the presence of the outlier with 1959 days to progression. The elderly cohort (≥ 75 years) of 6 patients demonstrated a mean time to progression of 491.7 days (SD 287.1), slightly longer than the younger group, with a median of 513 days (IQR 254-748). Notably, the elderly cohort showed a substantially higher median (513 vs 279 days), suggesting more indolent disease progression in this small elderly subset.

The range of progression times differed markedly between age groups. While the younger cohort showed the widest range (125-1959 days), the elderly cohort had a more constrained range (139-791 days), with no extreme long progressors. The interquartile ranges showed the elderly group both had a higher 25th percentile (254 vs 148 days) and 75th percentile (748 vs 498 days), indicating consistently longer times to progression in the middle 50% of elderly patients.

Table 19: Distribution of metastatic presentation by age group - HUS

Disease Presentation	Overall (N=199) n (%)	Age <75 (n=153) n (%)	Age ≥75 (n=46) n (%)
De novo metastatic	170 (85.4)	130 (85.0)	40 (87.0)
Emergent metastatic	29 (14.6)	23 (15.0)	6 (13.0)

6.2.3.4.3 *Metastatic Presentation Patterns by Age*

The distribution of metastatic presentation in the HUS cohort showed predominance of de novo disease across all age groups (Table 21), with the highest proportion of emergent disease (14.6%). HUS demonstrated relatively similar proportions of emergent disease between age groups: 15.0% in the younger cohort versus 13.0% in the elderly cohort.

Among the 29 emergent cases, 23 (79.3%) occurred in patients aged <75 years and 6 (20.7%) in those ≥75 years. Given that the elderly group represented 23.1% of the overall cohort, the 20.7% representation in emergent cases indicates a slight under-representation of elderly patients in the emergent population.

The heterogeneity in time to metastatic progression, particularly the presence of very long progressors in the younger cohort (maximum 1959 days). The higher median time to progression in elderly patients (513 vs 279 days), despite the smaller sample size.

6.2.4 PD-L1 Expression Status

PD-L1 expression testing, a critical biomarker for immunotherapy selection in mNSCLC, faced substantial data capture and standardization challenges across the Nordic sites. Despite its established role in treatment decision-making—particularly for identifying patients with ≥50% expression eligible for pembrolizumab monotherapy—harmonized PD-L1 data could not be reliably analyzed across sites.

6.2.4.1 *Site-Specific PD-L1 Data Availability*

OUS: PD-L1 expression data was captured for a limited subset of patients with ranges including <1%, <50%, ≥50%, and ≥75%. However, small patient numbers in each category (all <10, with many masked values) prevented meaningful analysis of PD-L1 distribution or its relationship to treatment selection.

Rigshospitalet and Herlev Hospital: PD-L1 testing data was available for the majority of patients, with 273 patients (55.8%) documented as having ≥50% expression. However, the data revealed significant standardization challenges with multiple overlapping and incompatible range definitions (e.g., "<0", "<1", "≥1 + <25", "≥50", and numerous "No matching concept" entries combined with numerical values). While 86 patients had <1%

expression and various intermediate ranges were captured, the heterogeneous reporting formats prevented consistent categorization into standard clinical groups.

HUS: PD-L1 expression data was not available in the extracted dataset for analysis.

Other biomarker availability: Beyond the exclusion criteria biomarkers (EGFR, ALK, ROS1 mutations), additional molecular markers showed significant variation in availability across sites. OUS had limited additional biomarker availability beyond KRAS. Rigshospitalet and Herlev Hospitals captured biomarkers in measurement plus value format that could have been profiled with extensive effort. HUS utilized OMOP Genomics vocabulary, which requires extensive data profiling and clinical characterization for meaningful analysis. Due to the substantial variance between sites and the extensive profiling required, analysis of additional biomarkers including KRAS G12C, BRAF V600E, MET exon 14 skipping, RET rearrangements, and NTRK fusions was not pursued. This limits understanding of the study population's complete molecular landscape, though it may not impact the primary ICI treatment pattern analyses.

6.3 Objective 2a: Treatment Patterns Prior-to and Post Metastatic Progression

Treatment characterization analyses demonstrate the network's capability to extract longitudinal therapy data from routine clinical records while documenting treatment pattern heterogeneity across sites. This analysis characterizes treatments administered prior to and following metastatic progression. The analysis examined treatment patterns after metastatic diagnosis rather than after ICI initiation (index date) as originally specified in the protocol. This modification was applied consistently across all participating sites.

6.3.1 Treatment Characterization - OUS

6.3.1.1 Emergent Metastatic Population

Among the 8 patients with emergent metastatic disease in OUS, treatment patterns prior to metastatic progression were analyzed (Table 22). The extremely small emergent population (n=8) substantially limits the interpretability of these findings.

Table 20: Treatments received prior to metastatic progression - OUS emergent population

Treatment Modality	Overall (N=8) n (%)	Age <75 (n=<5)* n (%)	Age ≥75 (n=5) n (%)
Any chemotherapy	<5*	<5*	0 (0.0)
Platinum-based	<5*	<5*	0 (0.0)
Non-platinum-based	<5*	<5*	0 (0.0)
Surgery	5 (62.5)	<5*	<5
Radiotherapy	0 (0.0)	0 (0.0)	0 (0.0)
Prior ICI therapy	0 (0.0)	<5*	0 (0.0)



Treatment Modality	Overall (N=8) n (%)	Age <75 (n=<5)* n (%)	Age ≥75 (n=5) n (%)
Chemotherapy timing **			
Adjuvant proxy	<5*	<5*	0 (0.0)
Neoadjuvant proxy	0 (0.0)	<5*	0 (0.0)
Chemotherapy only	0 (0.0)	<5*	0 (0.0)

* Values <5 suppressed for patient privacy

** These proxy classifications apply to treatments administered during the non-metastatic disease phase (before metastatic progression). Neoadjuvant refers to chemotherapy given before surgery/radiotherapy when disease was still localized; adjuvant refers to chemotherapy after surgery/radiotherapy. These designations do not refer to metastatic disease treatment

Radiotherapy data were not captured in the OUS database system used for this analysis (0.0% documentation across all 67 patients). This data limitation does not indicate OUS patients did not receive radiation therapy in clinical practice. Radiotherapy services at OUS are likely documented in separate clinical information systems not interfaced with the OMOP CDM infrastructure or delivered at external facilities whose treatment records were not accessible through the federated query framework. This fundamental data gap prevents any conclusions about actual radiotherapy utilization patterns in the OUS cohort.

6.3.1.2 Treatment Patterns After Metastatic Diagnosis

Post-Metastasis Treatment Characterization

Treatment patterns following metastatic diagnosis were analyzed for the entire OUS cohort (N=67) (Table 23).

Table 21: Treatments received after metastatic diagnosis - OUS

Treatment Modality	Overall (N=67) n (%)	Age <75 (n=53) n (%)	Age ≥75 (n=14) n (%)
Chemotherapy	43 (64.2)	36 (67.9)	7 (50.0)
Radiotherapy	0 (0.0)	0 (0.0)	0 (0.0)
Surgery	<5*	<5*	<5*

* Values <5 suppressed for patient privacy



Table 22: Chemotherapy treatment intensity using proxy cycle definitions - OUS

Chemotherapy Cycles*	Overall (n=43)**	Age <75 (n=36)**	Age ≥75 (n=7)**
28-day gap proxy			
Mean (SD)	2.0 (1.3)	2.0 (1.4)	1.9 (0.4)
Median (IQR)	2 (1-2)	1 (1-2.25)	2 (2-2)
Range	1-6	1-6	1-2
14-day gap proxy			
Mean (SD)	6.6 (6.3)	6.8 (6.8)	5.9 (3.0)
Median (IQR)	4 (2.5-9)	4 (2-9)	4 (4-7.5)
Range	1-32	1-32	3-11

* Proxy definitions based on gaps between chemotherapy administrations; actual cycle counts not recorded in database

** Denominator represents patients who received chemotherapy

Treatment Utilization Patterns

Chemotherapy was administered to 64.2% of patients overall, with 67.9% utilization in patients less than 75 years versus 50.0% in those 75 years or older. Chemotherapy treatment intensity was assessed using gap-based proxy definitions. The 28-day gap proxy showed a median of 2 cycles (IQR 1-2) overall, with a range of 1-6 cycles. The 14-day gap proxy showed a median of 4 cycles (IQR 2.5-9) with a range of 1-32 cycles. Elderly patients showed median values of 2 cycles (28-day proxy) and 4 cycles (14-day proxy).

Chemotherapy treatment intensity was assessed using gap-based proxy definitions. The 28-day gap proxy showed a median of 2 cycles (IQR 1-2) overall, with a range of 1-6 cycles. The 14-day gap proxy showed a median of 4 cycles (IQR 2.5-9) with a range of 1-32 cycles. Elderly patients showed median values of 2 cycles (28-day proxy) and 4 cycles (14-day proxy).

Chemotherapy Treatment Intensity Assessment

The 28-day proxy yielded a median of 2 episodes (IQR 1-2, range 1-6) while the 14-day proxy yielded a median of 4 episodes (IQR 2.5-9, range 1-32) in the OUS cohort. The two-fold difference between proxies reflects the sensitivity of episode counting to the gap threshold selected. Both proxies showed similar patterns regarding age differences, with elderly patients demonstrating lower maximum episode counts (28-day: 1-2 versus 1-6 in younger; 14-day: 3-11 versus 1-32 in younger).

Chemotherapy Treatment Intensity

Using the 28-day gap proxy for conventional chemotherapy cycles, patients received a median of 2 cycles overall (IQR 1-2). The elderly cohort showed a narrow range (1-2 cycles) with minimal variability (SD 0.4), in contrast, younger patients with a maximum of 6 cycles. The 14-day gap proxy revealed a median of 4 treatment episodes overall (IQR 2.5-9). The maximum of 32 episodes in the younger cohort suggests some patients received extended maintenance therapy. Elderly patients showed a higher minimum threshold (3 episodes), but

lower maximum (11 episodes) compared to younger patients, with less variability overall (SD 3.0 vs 6.8).

6.3.2 Treatment Characterization – Rigshospitalet/Herlev Hospital

6.3.2.1 Prior-Treatment Metastatic Population

Among the 38 patients with emergent metastatic disease in Rigshospitalet/Herlev Hospital (Table 25), treatment patterns prior to metastatic progression were analyzed to understand the therapeutic approaches used during the non-metastatic disease phase.

Table 23: Treatments received prior to metastatic progression - Rigshospitalet/Herlev Hospital emergent population

Treatment Modality	Overall (N=38) n (%)	Age <75 (n=28) n (%)	Age ≥75 (n=10) n (%)
Any chemotherapy	11 (28.9)	8 (28.6)	<5*
Platinum-based	10 (26.3)	7 (25.0)	<5*
Non-platinum-based	10 (26.3)	7 (25.0)	<5*
Surgery	19 (50.0)	15 (53.6)	<5*
Radiotherapy	15 (39.5)	10 (35.7)	5 (50.0)
Prior ICI therapy	<5*	<5*	<5*
Chemotherapy timing**			
Adjuvant proxy	7 (18.4)	<5*	<5*
Neoadjuvant proxy	<5*	<5*	0 (0.0)
Chemotherapy only	<5*	<5*	0 (0.0)

* Values <5 suppressed for patient privacy

** These proxy classifications apply to treatments administered during the non-metastatic disease phase (before metastatic progression). Neoadjuvant refers to chemotherapy given before surgery/radiotherapy when disease was still localized; adjuvant refers to chemotherapy after surgery/radiotherapy. These designations do NOT refer to metastatic disease treatment.

Half of the emergent metastatic population (19/38, 50.0%) underwent surgical resection during their non-metastatic disease phase, suggesting initial presentation with potentially resectable disease. This proportion was slightly higher in the younger cohort (53.6%) compared to data available for the elderly cohort. Surgical intervention as part of early-stage management aligns with standard care pathways for resectable NSCLC.

Systemic chemotherapy was administered to approximately one-third of patients (28.9%) prior to metastatic progression. The distribution between platinum-based and non-platinum-based regimens was identical (both 26.3%), with some patients likely receiving both types sequentially or in combination.

Radiotherapy was documented in 39.5% of patients overall, with notable age-related differences where available data showed 50.0% documentation in the elderly cohort versus



35.7% in younger patients. Prior ICI exposure before metastatic progression was rare, with fewer than 5 patients receiving immunotherapy during the non-metastatic phase.

Analysis of chemotherapy timing using proxy definitions revealed that 18.4% of patients received chemotherapy in an adjuvant setting (after surgery or radiotherapy), while neoadjuvant chemotherapy and chemotherapy-only approaches were uncommon (<5 patients each). These proxy definitions, while imperfect due to lack of explicit treatment intent documentation in the database, provide insight into multimodal treatment sequencing.

6.3.2.2 Post-Treatment Metastatic Population

Treatment patterns following metastatic diagnosis were analyzed for the entire Rigshospitalet/Herlev Hospital cohort (N=489) to understand therapeutic interventions administered after progression to metastatic disease, including the first-line ICI therapy and subsequent treatments (Table 26 and 27).

Table 24: Treatments received after metastatic diagnosis - Rigshospitalet/Herlev Hospital

Treatment Modality	Overall (N=489) n (%)	Age <75 (n=378) n (%)	Age ≥75 (n=111) n (%)
Chemotherapy	322 (65.8)	265 (70.1)	57 (51.4)
Radiotherapy	369 (75.5)	292 (77.2)	77 (69.4)
Surgery	34 (7.0)	29 (7.7)	5 (4.5)

Table 25: Chemotherapy treatment intensity using proxy cycle definitions Rigshospitalet/Herlev Hospital

Chemotherapy Cycles*	Overall (n=322)**	Age <75 (n=265)**	Age ≥75 (n=57)**
28-day gap proxy			
Mean (SD)	2.0 (1.3)	2.0 (1.3)	1.8 (1.0)
Median (IQR)	2 (1-3)	2 (1-3)	1 (1-2)
Range	1-9	1-9	1-5
14-day gap proxy			
Mean (SD)	6.9 (6.5)	7.0 (6.7)	6.3 (5.0)
Median (IQR)	5 (3-9)	5 (2-9)	4 (3-9)
Range	1-41	1-41	1-22

* Proxy definitions based on gaps between chemotherapy administrations; actual cycle counts not recorded in database

** Denominator represents patients who received chemotherapy

Treatment Utilization patterns

Following metastatic diagnosis, the majority of patients received systemic and local therapies in addition to first-line ICI. Chemotherapy was administered to 65.8% of the overall cohort,



with marked age-related differences: 70.1% in patients <75 years versus 51.4% in those ≥75 years. Radiotherapy emerged as the most commonly utilized treatment modality, administered to 75.5% of all patients. Age-stratified analysis showed 77.2% utilization in younger patients and 69.4% in elderly patients. Surgical intervention after metastatic diagnosis was uncommon, performed in only 7.0% of patients overall. The younger cohort had a slightly higher surgical rate (7.7%) compared to elderly patients (4.5%), though both rates were low.

Using the 28-day gap proxy, patients received a median of 2 cycles (IQR 1-3), with a maximum of 9 cycles observed. The elderly cohort showed lower treatment intensity with a median of 1 cycle (IQR 1-2) compared to 2 cycles in younger patients. The 14-day gap proxy showed a median of 5 treatment episodes (IQR 3-9) overall. The maximum of 41 episodes suggests some patients received extended maintenance therapy. Elderly patients again showed lower intensity with a median of 4 episodes compared to 5 in younger patients, and a narrower range (1-22 vs 1-41).

6.3.3 Treatment Characterization - HUS

6.3.3.1 Prior-Treatment Metastatic Population

Among the 29 patients with emergent metastatic disease in HUS, treatment patterns prior to metastatic progression were analyzed (Table 28).

Table 26: Treatments received prior to metastatic progression - HUS emergent population (HUS)

Treatment Modality	Overall (N=29) n (%)	Age <75 (n=23) n (%)	Age ≥75 (n=6) n (%)
Any chemotherapy	23 (79.3)	18 (78.3)	5 (83.3)
Platinum-based	22 (75.9)	17 (73.9)	5 (83.3)
Non-platinum-based	23 (79.3)	18 (78.3)	5 (83.3)
Surgery	7 (24.1)	6 (26.1)	<5*
Radiotherapy	12 (41.4)	9 (39.1)	<5*
Prior ICI therapy	5 (17.2)	5 (21.7)	0 (0.0)
Chemotherapy timing**			
Adjuvant proxy	5 (17.2)	<5*	<5*
Neoadjuvant proxy	<5*	<5*	0 (0.0)
Chemotherapy only	9 (31.0)	7 (30.4)	<5*

* Values <5 suppressed for patient privacy

** These proxy classifications apply to treatments administered during the non-metastatic disease phase (before metastatic progression). Neoadjuvant refers to chemotherapy given before surgery/radiotherapy when disease was still localized; adjuvant refers to chemotherapy after surgery/radiotherapy. These designations do NOT refer to metastatic disease treatment.

The HUS emergent metastatic population demonstrated markedly more intensive treatment prior to progression. Chemotherapy was administered to 79.3% of patients. Both platinum-



based (75.9%) and non-platinum-based (79.3%) chemotherapy were extensively utilized, with the similar percentages suggesting many patients received both types, likely as combination therapy or sequential regimens. The elderly cohort showed higher chemotherapy rates (83.3%) than younger patients (78.3%). Surgical resection was performed in 24.1% of patients overall and radiotherapy was utilized in 41.4% of patients. The combination of high chemotherapy and moderate radiotherapy use indicates multimodal treatment approaches, possibly including chemoradiation for locally advanced disease. Notably, 5 patients (17.2%) received ICI therapy during the non-metastatic phase, all in the younger cohort. Analysis of chemotherapy timing revealed that 31.0% received chemotherapy alone without associated local therapies, suggesting treatment of unresectable disease. Adjuvant chemotherapy (proxy definition) was documented in 17.2%, while neoadjuvant use was rare (<5 patients).

6.3.3.2 Post-Treatment Metastatic Population

Treatment patterns following metastatic diagnosis were analyzed for the entire HUS cohort (N=199) to understand therapeutic interventions administered after progression to metastatic disease, including the first-line ICI therapy and subsequent treatments (Table 29 and 30).

Table 27: Treatments received after metastatic diagnosis - HUS

Treatment Modality	Overall (N=199) n (%)	Age <75 (n=153) n (%)	Age ≥75 (n=46) n (%)
Chemotherapy	143 (71.9)	115 (75.2)	28 (60.9)
Radiotherapy	132 (66.3)	107 (69.9)	25 (54.3)
Surgery	10 (5.0)	8 (5.2)	<5*

* Values <5 suppressed for patient privacy

Table 28: Chemotherapy treatment intensity using proxy cycle definitions - HUS

Chemotherapy Cycles*	Overall (n=143)**	Age <75 (n=115)**	Age ≥75 (n=28)**
28-day gap proxy			
Mean (SD)	2.0 (1.3)	2.1 (1.4)	1.5 (0.8)
Median (IQR)	2 (1-3)	2 (1-3)	1 (1-2)
Range	1-6	1-6	1-4
14-day gap proxy			
Mean (SD)	5.1 (4.1)	5.3 (4.2)	4.3 (3.5)
Median (IQR)	4 (3-6)	4 (3-7)	3 (3-5)
Range	1-24	1-24	1-18

* Proxy definitions based on gaps between chemotherapy administrations; actual cycle counts not recorded in database

** Denominator represents patients who received chemotherapy



Treatment Utilization Patterns

The HUS cohort demonstrated high utilization of both systemic and local therapies following metastatic diagnosis. Chemotherapy was administered to 71.9% of patients overall. Age-stratified analysis showed 75.2% utilization in younger patients versus 60.9% in elderly patients, a 14.3 percentage point difference reflecting age-related variation in documented radiotherapy. The substantial documented radiotherapy in HUS (66.3%) and Rigshospitalet/Herlev Hospital (75.5%) contrasts with OUS's 0% radiotherapy data capture, reflecting differences in data infrastructure rather than clinical practice patterns. The available HUS radiotherapy data suggests active use of radiation for both palliative and consolidative purposes, though incomplete capture across all sites limits cross-country comparisons. Surgical intervention remained uncommon at 5.0% overall and consistent with expected limited surgical options in metastatic disease. The younger cohort rate of 5.2% was comparable to the overall rate, with elderly surgical data masked due to small numbers.

Using the 28-day gap proxy, patients received a median of 2 cycles overall (IQR 1-3). However, elderly patients showed notably lower treatment intensity with a median of 1 cycle (IQR 1-2) and maximum of 4 cycles, compared to younger patients with a median of 2 cycles and maximum of 6 cycles. The 14-day gap proxy revealed a median of 4 treatment episodes overall (IQR 3-6). The maximum of 24 episodes indicates some patients received extended therapy. Elderly patients showed consistently lower intensity with a median of 3 episodes (IQR 3-5) versus 4 episodes in younger patients, and a lower maximum (18 vs 24).

6.4 Objective 2b: Longitudinal Treatment Sequences

Sequential treatment pattern analyses demonstrate the network's capability to reconstruct complex therapeutic trajectories while identifying proxy measure requirements necessitated by incomplete treatment intent documentation.

6.4.1 Treatment Sequence Patterns – OUS

Longitudinal treatment sequences were analyzed for the OUS cohort (N=67) to characterize treatment pathways following first-line immune checkpoint inhibitor initiation in metastatic NSCLC.

Table 29: Most common treatment sequences by age group - OUS

Treatment Sequence	Overall (N=67) n (%)	Age <75 (n=53) n (%)	Age ≥75 (n=14) n (%)
Single Line Therapy			
Pembrolizumab monotherapy	22 (32.8)	17 (32.1)	5 (35.7)
Chemotherapy + Pembrolizumab	12 (17.9)	10 (18.9)	<5 ^a
Atezolizumab monotherapy	5 (7.5)	<5*	<5*
Durvalumab monotherapy	<5*	<5*	-
Cemiplimab monotherapy	<5*	-	<5*
Two-Line Sequences			
Chemo + Pembrolizumab → Chemotherapy	12 (17.9)	11 (20.8)	<5*
Pembrolizumab → Chemotherapy	<5*	<5*	-



Treatment Sequence	Overall (N=67) n (%)	Age <75 (n=53) n (%)	Age ≥75 (n=14) n (%)
Chemo + Pembrolizumab → Pembrolizumab	<5*	<5*	<5*
Atezolizumab → Chemotherapy	<5*	<5*	-
Three-Line Sequences			
All three-line sequences	<5*	<5*	<5*

* Values <5 masked for patient privacy

Table 30: Duration of treatment by line of therapy - OUS

Treatment Line	Overall Median days [IQR] (n)	Age <75 Median days [IQR] (n)	Age ≥75 Median days [IQR] (n)
Line 1 (All treatments)	100 [42-182] (67)	84 [28-169] (53)	127 [63-186] (14)
Pembrolizumab monotherapy	126 [37-401] (28)	126 [37-557] (20)	107 [87-147] (5)
Chemo + Pembrolizumab	84 [50-156] (29)	72 [45-139] (23)	NA* (<5)
Atezolizumab monotherapy	69 [24-203] (10)	28 [22-74] (5)	NA* (<5)
Line 2 (All treatments)	63 [43-117] (23)	68 [43-116] (20)	NA* (<5)
Chemotherapy	63 [43-114] (17)	62 [43-114] (15)	NA* (<5)
Line 3 (All treatments)	NA* (<5)	NA* (<5)	NA* (<5)

* Data not available or masked for groups <5 patients

First-Line Treatment Patterns and Limited Therapeutic Diversity

Single-line pembrolizumab monotherapy was the most common treatment approach, administered to 22 patients (32.8%) overall, with 17 patients (32.1%) in the younger cohort and 5 patients (35.7%) in the elderly cohort. Chemotherapy plus pembrolizumab as single-line therapy was utilized in 12 patients (17.9%) overall, with 10 patients (18.9%) in the younger group and fewer than 5 in the elderly group.

The median duration of first-line pembrolizumab monotherapy (n=28 total receiving this treatment) was 126 days (IQR 37-401). Among the 20 younger patients, the median duration was 126 days (IQR 37-557), showing extreme heterogeneity. The 5 elderly patients demonstrated a median of 107 days (IQR 87-147) with a narrower range.

First-line chemotherapy plus pembrolizumab (n=29 total) showed a median duration of 84 days (IQR 50-156). The 23 younger patients had a median duration of 72 days (IQR 45-139), while elderly patient data was not available due to small numbers.



Atezolizumab monotherapy was administered to 5 patients (7.5%) as single-line therapy overall. Among 10 total patients receiving atezolizumab across all lines, the median duration was 69 days (IQR 24-203). The 5 younger patients showed a notably brief median of 28 days (IQR 22-74). Durvalumab and cemiplimab were each used in fewer than 5 patients.

Sequential Treatment Strategies

Twenty-three patients (34.3%) received characterized second-line therapy. Among the younger cohort, 20 of 53 patients (37.7%) progressed to second-line, while fewer than 5 elderly patients received second-line treatment.

The dominant two-line sequence was chemotherapy plus pembrolizumab followed by chemotherapy alone, observed in 12 patients (17.9% of total cohort), representing 52.2% of all patients receiving second-line therapy (12/23). This pattern included 11 patients (20.8%) in the younger cohort and fewer than 5 elderly patients. All other two-line sequences (pembrolizumab → chemotherapy, chemotherapy plus pembrolizumab → pembrolizumab, atezolizumab → chemotherapy) occurred in fewer than 5 patients each. Among younger patients, 20 of 53 (37.7%) received second-line therapy, while fewer than 5 elderly patients progressed to second-line, preventing meaningful age-stratified analysis.

Third-line therapy was essentially absent, with fewer than 5 patients across the entire cohort receiving characterized third-line treatment or the impact of radiotherapy (data not captured) as a bridging option.

Treatment Duration Dynamics

First-line treatment duration showed a paradoxical age relationship, with elderly patients maintaining longer median treatment duration (127 vs 84 days).

First-line treatment duration for all therapies combined showed a median of 100 days (IQR 42-182) across 67 patients. The maximum first-line duration reached 730 days, demonstrating that some patients achieved remarkable durable disease control despite the limited therapeutic arsenal. Age-stratified analysis revealed that elderly patients (n=14) maintained longer median first-line duration at 127 days (IQR 63-186) compared to younger patients (n=53) at 84 days (IQR 28-169).

Second-line treatment duration was uniformly brief, with a median of 63 days overall (IQR 43-117), representing rapid treatment failure across all second-line approaches. Among the 17 patients receiving second-line chemotherapy, the median duration of only 63 days. The variability in treatment duration showed distinct age-related patterns. Younger patients demonstrated extreme heterogeneity in first-line pembrolizumab duration (range 37-730 days), encompassing both rapid progressors and exceptional responders. In contrast, elderly patients showed remarkable uniformity (range 87-379 days).

6.4.2 Treatment Sequence Patterns – Rigshospitalet/Herlev Hospital

Longitudinal treatment sequences were analyzed for the entire Rigshospitalet/Herlev Hospital cohort (N=489) to characterize real-world treatment pathways following first-line immune checkpoint inhibitor initiation in metastatic NSCLC (Table 33 and 34).



Table 31: Most common treatment sequences by age group – Rigshospitalet/Herlev Hospital

Treatment Sequence	Overall (N=489) n (%)	Age <75 (n=378) n (%)	Age ≥75 (n=111) n (%)
Single Line Therapy			
Pembrolizumab monotherapy	171 (35.0)	120 (31.7)	51 (45.9)
Chemotherapy + Pembrolizumab	79 (16.2)	63 (16.7)	16 (14.4)
Atezolizumab monotherapy	26 (5.3)	16 (4.2)	10 (9.0)
Durvalumab monotherapy	26 (5.3)	23 (6.1)	<5*
Two-Line Sequences			
Pembrolizumab → Chemotherapy	56 (11.5)	47 (12.4)	9 (8.1)
Chemo + Pembrolizumab → Chemotherapy	51 (10.4)	45 (11.9)	6 (5.4)
Chemo + Pembrolizumab → Pembrolizumab	30 (6.1)	22 (5.8)	8 (7.2)
Atezolizumab → Chemotherapy	12 (2.5)	11 (2.9)	<5*
Durvalumab → Chemotherapy	8 (1.6)	6 (1.6)	<5*
Three-Line Sequences			
Chemo + Pembrolizumab → Pembrolizumab → Chemo + Pembrolizumab	6 (1.2)	<5*	<5*

* Values <5 masked for patient privacy

Table 32: Duration of treatment by line of therapy - Rigshospitalet/Herlev Hospital

Treatment Line	Overall Median days [IQR] (n)	Age <75 Median days [IQR] (n)	Age ≥75 Median days [IQR] (n)
Line 1 (All treatments)	86 [42-215] (489)	116 [43-245] (378)	72 [43-212] (111)
Pembrolizumab monotherapy	116 [26-316] (239)	129 [30-322] (178)	67 [22-259] (61)
Chemo + Pembrolizumab	95 [44-188] (170)	106 [43-188] (138)	89 [62-176] (32)
Atezolizumab monotherapy	79 [64-178] (39)	80 [62-220] (28)	65 [64-112] (11)
Durvalumab monotherapy	111 [59-320] (39)	137 [60-324] (34)	62 [59-82] (5)
Line 2 (All treatments)	79 [41-156] (186)	74 [33-150] (156)	108 [71-184] (30)



Treatment Line	Overall Median days [IQR] (n)	Age <75 Median days [IQR] (n)	Age ≥75 Median days [IQR] (n)
Chemotherapy	71 [35-138] (131)	67 [29-133] (112)	79 [70-162] (19)
Pembrolizumab monotherapy	108 [73-206] (42)	97 [69-215] (32)	149 [100-184] (10)
Chemo + Pembrolizumab	155 [22-278] (12)	154 [22-213] (11)	- (<5)*
Line 3 (All treatments)	110 [29-166] (23)	85 [27-162] (21)	- (<5)*
Chemotherapy	160 [66-225] (10)	160 [66-225] (10)	-
Pembrolizumab monotherapy	43 [25-110] (6)	43 [25-110] (6)	-
Chemo + Pembrolizumab	98 [72-127] (6)	- (<5)*	- (<5)*

* Data not available or masked for groups <5 patients

First-Line Treatment Strategies and Monotherapy Predominance

The Rigshospitalet/Herlev Hospital cohort demonstrated a clear preference for immunotherapy monotherapy as the initial treatment strategy, with 252 patients (51.5%) receiving single-agent ICI as their only characterized treatment line. Pembrolizumab monotherapy emerged as the dominant approach, administered to 171 patients (35.0%) overall, representing the single most common treatment pathway.

Among elderly patients (≥75 years), pembrolizumab monotherapy utilization reached 45.9% (51/111), compared to 31.7% (120/378) in younger patients, representing a 14.2 percentage point difference. The median duration of first-line pembrolizumab monotherapy further reinforced these age-related considerations, with elderly patients maintaining treatment for only 67 days (IQR 22-259) compared to 129 days (IQR 30-322) in younger patients.

Alternative ICI monotherapies demonstrated distinct utilization patterns. Atezolizumab monotherapy was administered to 26 patients (5.3%) overall, with higher utilization in the elderly cohort (9.0% vs 4.2%). Durvalumab monotherapy, used in 26 patients (5.3%) overall, showed opposite age stratification with 6.1% utilization in younger patients versus <4.5% in elderly patients. The median treatment duration for durvalumab showed the most dramatic age-related disparity: 137 days in younger patients versus only 62 days in elderly patients.

Combination Therapy Approaches and Sequential Strategies

Upfront chemotherapy plus pembrolizumab combination represented the second most prevalent first-line strategy, administered to 79 patients (16.2%) overall. This combination approach showed relatively consistent utilization across age groups (16.7% in <75 years, 14.4% in ≥75 years). The median duration of first-line combination therapy was 95 days overall (IQR 44-188), shorter than pembrolizumab monotherapy.

The subsequent treatment patterns following initial combination therapy revealed two distinct management strategies. Among the 79 patients starting with chemotherapy plus



pembrolizumab, 51 (64.6%) proceeded to second-line chemotherapy. Conversely, 30 patients (38.0%) transitioned to pembrolizumab monotherapy as second-line. This de-escalation strategy was slightly more common in elderly patients (7.2% vs 5.8%).

Progressive Treatment Line Characteristics

The analysis revealed that 184 patients (37.6%) received at least two lines of characterized therapy, with marked age-related differences in progression to subsequent lines. Among younger patients, 40.5% (153/378) received second-line therapy compared to only 28.8% (32/111) in elderly patients, representing an 11.7 percentage point difference. This disparity likely reflects multiple factors including performance status deterioration, treatment tolerance, and clinical decision-making regarding the benefit-risk ratio of subsequent therapies in elderly populations.

Second-line treatment patterns demonstrated a shift toward chemotherapy predominance, with 131 patients receiving chemotherapy as second-line therapy, representing 70.4% of all second-line treatments. The median duration of second-line chemotherapy was notably brief at 71 days (IQR 35-138), substantially shorter than first-line treatments, suggesting limited efficacy of chemotherapy after immunotherapy progression. Interestingly, elderly patients who reached second-line therapy showed longer median treatment duration (108 vs 74 days).

Third-line therapy was characterized in only 23 patients (4.7% of the overall cohort), with 21 being younger patients and fewer than 5 elderly patients reaching this line. The median duration of third-line therapy at 110 days (IQR 29-166) showed substantial variability, with some patients achieving prolonged disease control while others rapidly progressed.

Duration Dynamics Across Treatment Lines

The temporal analysis of treatment duration revealed complex patterns that varied by both treatment line and patient age. First-line treatments showed the longest median duration at 86 days overall, with substantial heterogeneity reflected in the wide interquartile range (42-215 days). This variability encompassed both rapid progressors experiencing early treatment failure and exceptional responders maintaining treatment for over 500 days, with maximum durations reaching 1713 days for first-line pembrolizumab monotherapy.

The maximum treatment durations decreased substantially with each line: 1713 days (Line 1) → 1097 days (Line 2) → 426 days (Line 3). Age-stratified duration analysis revealed paradoxical patterns requiring careful interpretation. While younger patients showed longer first-line duration (116 vs 72 days), elderly patients who reached second-line demonstrated longer treatment duration (108 vs 74 days).

6.4.3 Treatment Sequence Patterns – HUS

Treatment Sequence Patterns

Longitudinal treatment sequences were analyzed for the HUS cohort (N=199) to characterize real-world treatment pathways following first-line immune checkpoint inhibitor initiation in metastatic NSCLC (table 35 and 36).



Table 33: Most common treatment sequences by age group - HUS

Treatment Sequence	Overall (N=199) n (%)	Age <75 (n=153) n (%)	Age ≥75 (n=46) n (%)
Single Line Therapy			
Pembrolizumab monotherapy	53 (26.6)	37 (24.2)	16 (34.8)
Chemotherapy + Pembrolizumab	31 (15.6)	23 (15.0)	8 (17.4)
Atezolizumab monotherapy	17 (8.5)	12 (7.8)	5 (10.9)
Nivolumab monotherapy	5 (2.5)	5 (3.3)	0 (0.0)
Cemiplimab monotherapy	<5*	<5*	<5*
Two-Line Sequences			
Chemo + Pembrolizumab → Pembrolizumab	22 (11.1)	18 (11.8)	<5*
Chemo + Pembrolizumab → Chemotherapy	21 (10.6)	18 (11.8)	<5*
Pembrolizumab → Chemotherapy	18 (9.0)	16 (10.5)	<5*
Atezolizumab → Chemotherapy	<5*	<5*	<5*
Three-Line Sequences			<5*
Chemo + Pembrolizumab → Pembrolizumab → Chemotherapy	<5*	<5*	<5*
Chemo + Pembrolizumab → Pembrolizumab → Chemo + Pembrolizumab	<5*	<5*	<5*

* Values <5 masked for patient privacy

Table 34: Duration of treatment by line of therapy - HUS

Treatment Line	Overall Median days [IQR] (n)	Age <75 Median days [IQR] (n)	Age ≥75 Median days [IQR] (n)
Line 1 (All treatments)	52 [22-111] (194)	55 [22-106] (150)	44 [16-124] (44)
Pembrolizumab monotherapy	46 [22-144] (76)	46 [22-138] (57)	54 [11-162] (19)
Chemo + Pembrolizumab	64 [22-92] (81)	69 [24-93] (64)	38 [22-65] (17)
Atezolizumab monotherapy	52 [1-85] (21)	52 [1-70] (15)	64 [12-179] (6)



Treatment Line	Overall Median days [IQR] (n)	Age <75 Median days [IQR] (n)	Age ≥75 Median days [IQR] (n)
Nivolumab monotherapy	27 [17-140] (6)	27 [17-140] (6)	-
Cemiplimab monotherapy	44 [28-85] (8)	40 [10-123] (6)	NA* (<5)
Line 2 (All treatments)	85 [49-147] (83)	88 [49-146] (69)	72 [46-167] (14)
Chemotherapy	78 [39-139] (50)	79 [39-144] (43)	51 [41-81] (7)
Pembrolizumab monotherapy	124 [58-184] (29)	113 [57-191] (23)	158 [85-182] (6)
Line 3 (All treatments)	43 [8-117] (10)	42 [12-106] (7)	NA* (<5)

* Data not available or masked for groups <5 patients

First-Line Treatment Strategies and Diverse ICI Landscape

The HUS cohort demonstrated diverse checkpoint inhibitor utilization. Pembrolizumab monotherapy as single-line therapy was administered to 53 patients (26.6%) overall, with age-stratified rates of 37 patients (24.2%) in younger patients versus 16 patients (34.8%) in elderly patients, representing a 10.6 percentage point age difference. Chemotherapy plus pembrolizumab as single-line therapy was utilized in 31 patients (15.6%), with 23 patients (15.0%) in younger and 8 patients (17.4%) in elderly cohorts, showing relatively consistent utilization across age groups.

The median duration of first-line pembrolizumab monotherapy (n=76 total receiving this treatment) was 46 days (IQR 22-144), with similar medians between age groups: 46 days (IQR 22-138) in 57 younger patients and 54 days (IQR 11-162) in 19 elderly patients. First-line chemotherapy plus pembrolizumab (n=81 total) showed a median duration of 64 days (IQR 22-92) overall. However, marked age differences emerged: 69 days (IQR 24-93) in 64 younger patients versus only 38 days (IQR 22-65) in 17 elderly patients. This 31-day difference represents substantial age-related disparity in treatment duration.

Alternative ICIs achieved meaningful representation. Atezolizumab monotherapy was administered to 17 patients (8.5%) as single-line therapy, with 12 patients (7.8%) younger and 5 patients (10.9%) elderly. Among 21 total patients receiving atezolizumab, the median duration was 52 days (IQR 1-85), with extreme variability including patients with immediate discontinuation. Nivolumab appeared in 5 patients (2.5%), all in the younger cohort, with a median duration of 27 days (IQR 17-140) among 6 total recipients. Cemiplimab was used minimally in fewer than 5 patients.

6.4.4 Sequential Treatment Patterns and Maintenance Strategies

The analysis revealed that 83 patients (41.7%) received at least two lines of characterized therapy. In younger patients, 69 of 153 (45.1%) received second-line therapy, while elderly progression was more limited with 14 of 46 (30.4%) reaching second-line.

Among patients with two-line sequences, 22 (11.1%) followed chemotherapy plus pembrolizumab with pembrolizumab maintenance, while 21 (10.6%) proceeded from chemotherapy plus pembrolizumab to chemotherapy.

Pembrolizumab to chemotherapy sequencing occurred in 18 patients (9.0%), with 16 patients (10.5%) in the younger cohort. Other two-line sequences, including atezolizumab to chemotherapy, occurred in fewer than 5 patients.

Third-line therapy was characterized in 10 patients (5.0%), with specific sequences involving pembrolizumab re-challenge after intervening chemotherapy occurring in fewer than 5 patients each.

6.4.5 Treatment Duration Dynamics

The HUS cohort demonstrated notably brief first-line treatment durations, with an overall median of 52 days (IQR 22-111) across 194 patients. Age-stratified analysis showed 55 days (IQR 22-106) in 150 younger patients versus 44 days (IQR 16-124) in 44 elderly patients.

Among 83 patients reaching second-line, younger patients (n=69) showed a median of 88 days (IQR 49-146) while elderly patients (n=14) had 72 days (IQR 46-167). Chemotherapy as second-line (n=50) showed a median duration of 78 days (IQR 39-139), with 79 days (IQR 39-144) in 43 younger patients versus 51 days (IQR 41-81) in 7 elderly patients. Notably, pembrolizumab monotherapy as second-line maintenance (n=29) achieved a median duration of 124 days (IQR 58-184), exceeding first-line pembrolizumab duration. The 23 younger patients showed a median of 113 days (IQR 57-191), while 6 elderly patients demonstrated remarkable persistence at 158 days (IQR 85-182). Third-line therapy duration was brief at 43 days (IQR 8-117) among 10 patients, with 7 younger patients showing 42 days (IQR 12-106).

6.5 Objective 4: Clinical Outcomes - Overall Survival

6.5.1 Overall Survival Analysis - OUS

Overall survival (OS) was analyzed from the date of first ICI initiation following metastatic diagnosis (index date) until death from any cause (Table 37). Patients were censored at the end of the study period (December 31, 2023) or at last known follow-up. Survival probabilities were estimated using the Kaplan-Meier method with corresponding 95% confidence intervals (CI).



Table 35: Overall survival outcomes - OUS

Characteristic	Overall (N=67)	Age <75 (n=53)	Age ≥75 (n=14)
Number of deaths	45	35	10
Median survival, days (95% CI)	315 (250-476)	302 (207-754)	NA*
Restricted mean survival, days (95% CI)	619 (452-786)	645 (450-841)	517 (241-793)
Survival probability, % (95% CI)			
180 days	71.52 (61.46-83.22)	67.71 (56.18-81.62)	85.71 (69.21-100.00)
182 days	69.96 (59.76-81.90)	65.72 (54.04-79.93)	85.71 (69.21-100.00)
360 days	47.82 (36.79-62.15)	44.63 (32.69-60.93)	60.61 (38.37-95.74)
365 days	47.82 (36.79-62.15)	44.63 (32.69-60.93)	60.61 (38.37-95.74)
540 days	34.39 (24.00-49.29)	35.14 (23.78-51.94)	30.31 (12.06-76.16)
547 days	34.39 (24.00-49.29)	35.14 (23.78-51.94)	30.31 (12.06-76.16)
720 days	32.10 (21.85-47.14)	35.14 (23.78-51.94)	20.20 (5.96-68.46)
730 days	32.10 (21.85-47.14)	35.14 (23.78-51.94)	20.20 (5.96-68.46)

* NA: Data not available

Primary Survival Outcomes

The overall cohort experienced 45 death events among 67 patients during the follow-up period, representing a substantial mortality burden in this population of metastatic NSCLC patients receiving first-line ICI therapy. The median overall survival was 315 days (95% CI: 250-476), indicating that half of the cohort had died by approximately 10.4 months after ICI initiation. The confidence interval spanning from 250 to 476 days demonstrates considerable heterogeneity in survival outcomes within the cohort. The restricted mean survival time, which provides the average survival time up to the maximum follow-up period, was 619 days (95% CI: 452-786). This restricted mean substantially exceeded the median survival, suggesting the presence of a subset of long-term survivors whose extended survival influenced the mean calculation. The low number of patients in this cohort reduces the power in these analyses.

Age-stratified mortality analysis revealed distinct patterns between younger and elderly patients. The younger cohort (<75 years) experienced 35 death events among 53 patients, representing 66.0% mortality during follow-up. The median survival of 302 days (95% CI: 207-754) was numerically shorter than the overall cohort median, though the extremely wide



confidence interval spanning from 207 to 754 days indicates substantial variability in outcomes within this age group. The lower bound of 207 days suggests some younger patients experienced rapid disease progression, while the upper bound of 754 days indicates others achieved prolonged disease control. The restricted mean survival of 645 days (95% CI: 450-841) was the highest among all groups, suggesting that despite the lower median survival, a proportion of younger patients achieved extended survival durations that positively influenced the mean.

The elderly cohort (≥ 75 years) recorded 10 deaths among 14 patients, representing 71.4% mortality. The restricted mean survival for the elderly cohort was 517 days (95% CI: 241-793), the shortest restricted mean among all groups. The wide confidence interval from 241 to 793 days reflects both the small sample size of 14 patients and the consequent statistical uncertainty in survival estimates for this subgroup.

Figure 3: Kaplan-Meier survival curves by age group - OUS survival plot showing three curves with confidence intervals - overall (blue), <75 years (salmon), ≥ 75 years (green)

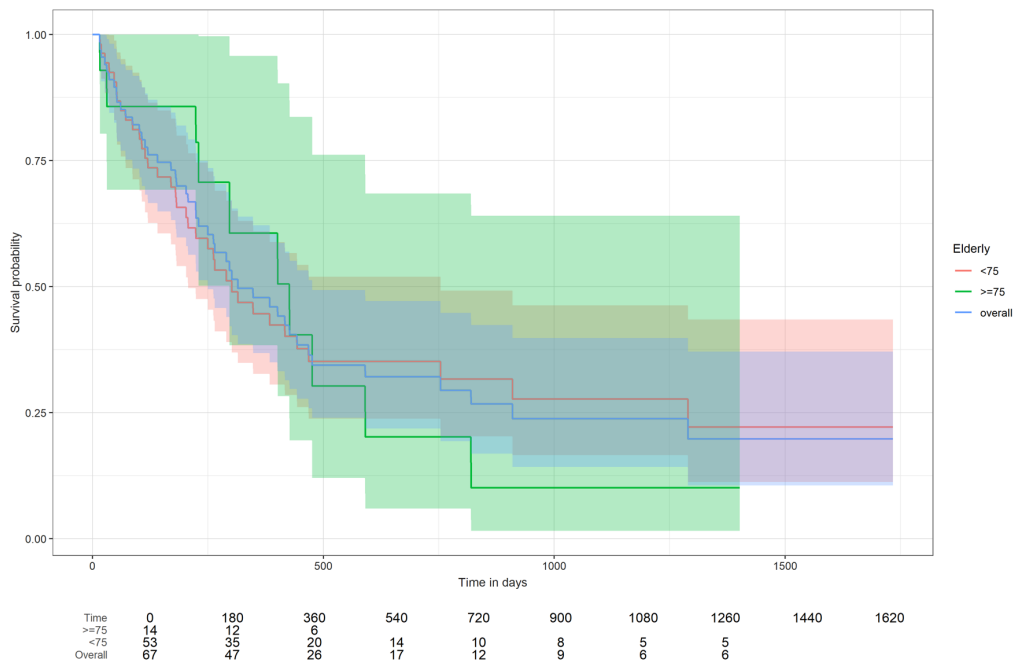


Table 36: Patients at risk and events over time - OUS

Time Point	Overall		Age <75		Age ≥ 75	
	At Risk	Events *	At Risk	Events *	At Risk	Events *
Days 0	67	0	53	0	14	0
180	47	19	35	17	12	-
360	26	14	20	11	6	-
540	17	7	14	-	-	-



Time Point	Overall		Age <75		Age ≥75	
	Events	Suppressed	Events	Suppressed	Events	Suppressed
720	12	-	10	0	-	-
900	9	-	8	-	-	-
1080	6	-	5	-	-	0
1260	6	0	5	0	-	0

* Events represent deaths occurring in the interval from the previous timepoint; "-" indicates suppressed or unavailable values

Survival Probability Trajectories

The temporal evolution of survival probability demonstrated distinct phases of mortality risk across the follow-up period (Table 38). At 180 days, the overall cohort maintained a survival probability of 71.52% (95% CI: 61.46-83.22). The confidence interval reflects moderate uncertainty in this estimate. A marginal decline to 69.96% (95% CI: 59.76-81.90) was observed by 182 days.

The survival probability demonstrated a more substantial decline by the 1-year mark. At 360 days, survival probability was 47.82% (95% CI: 36.79-62.15), representing a 23.70 percentage point decline from the 180-day estimate. This survival probability remained stable through 365 days at 47.82% (95% CI: 36.79-62.15), indicating no additional mortality events in this 5-day window. The confidence interval widened considerably at 1 year compared to 6 months, reflecting increasing uncertainty as the risk set diminished.

Long-term survival outcomes showed continued attrition with stabilization patterns. At 540 days (approximately 18 months), survival probability was 34.39% (95% CI: 24.00-49.29), representing a 13.43 percentage point decline from the 1-year estimate. This probability remained unchanged at 547 days (34.39%, 95% CI: 24.00-49.29). By 720 days (approximately 2 years), survival probability had declined to 32.10% (95% CI: 21.85-47.14), with no further change observed through 730 days.

Age-Stratified Survival Patterns

The younger cohort (<75 years) demonstrated distinct survival dynamics throughout the observation period. Initial 180-day survival was 67.71% (95% CI: 56.18-81.62), lower than the overall cohort, with further decline to 65.72% (95% CI: 54.04-79.93) at 182 days. This 1.99 percentage point decrease over 2 days indicates ongoing early mortality risk. By 365 days, survival probability had decreased to 44.63% (95% CI: 32.69-60.93), representing a 23.08 percentage point decline from 180 days.

A notable pattern emerged in the younger cohort's long-term survival. At both 540-547 days and 720-730 days, the survival probability remained stable at 35.14% (95% CI: 23.78-51.94).

The elderly cohort (≥75 years) exhibited markedly different survival trajectories. Early survival was superior to both the overall and younger cohorts, with 180-day and 182-day survival probability of 85.71% (95% CI: 69.21-100.00). The absence of change between these timepoints and the upper confidence limit reaching 100% reflects the small sample size and limited early events. At 365 days, the elderly cohort maintained higher survival at 60.61% (95% CI: 38.37-95.74), representing a 25.10 percentage point decline from 6 months but still exceeding the younger cohort's 1-year survival by 15.98 percentage points.

However, the elderly cohort experienced accelerated mortality between 1 year and 18 months. The 540–547-day survival probability dropped to 30.31% (95% CI: 12.06-76.16), a 30.30 percentage point decline from 1 year. This represented the steepest decline in any



time interval for any group. By 720-730 days, survival probability had further decreased to 20.20% (95% CI: 5.96-68.46). The extremely wide confidence intervals at these later timepoints (ranging from 5.96% to 68.46% at 2 years) reflect the very small number of elderly patients remaining at risk and the consequent statistical uncertainty.

Risk Set Evolution and Event Distribution

The risk set demonstrated progressive attrition throughout the follow-up period, with distinct patterns of event clustering. From the initial 67 patients, 47 remained at risk at 180 days following 19 events in the first 6 months, representing 28.4% early mortality. This high early event rate underscores the aggressive nature of metastatic NSCLC even in the era of immunotherapy. By 360 days, only 26 patients remained at risk after 14 additional events between 6 and 12 months. The risk set continued to decline to 17 patients at 540 days following 7 more events, and 12 patients at 720 days.

The younger cohort's risk set evolution showed 35 of 53 patients (66.0%) remaining at risk at 180 days after 17 events. This declined to 20 patients (37.7% of original cohort) at 360 days following 11 additional events, 14 patients (26.4%) at 540 days, and 10 patients (18.9%) at 720 days. Notably, no events were recorded at 720 days for this group, consistent with the observed plateau in survival probability.

The elderly cohort maintained 12 of 14 patients (85.7%) at risk at 180 days, demonstrating better early retention than the younger cohort. By 360 days, 6 patients (42.9% of original cohort) remained at risk. Beyond this timepoint, the elderly cohort data showed suppressed or unavailable values for patients at risk, limiting precise quantification of the risk set evolution in later periods. The available event data at 1080 and 1260 days showed 0 events, suggesting either survival stabilization among the remaining elderly patients or complete censoring.

6.5.2 Overall Survival Analysis - Rigshospitalet/Herlev Hospital

Overall Survival Analysis

Overall survival (OS) was analyzed from the date of first ICI initiation following metastatic diagnosis (index date) until death from any cause. Patients were censored at the end of the study period (December 31, 2023) or at last known follow-up.

Table 37: Overall survival outcomes - Rigshospitalet/Herlev Hospital

Characteristic	Overall (N=489)	Age <75 (n=378)	Age ≥75 (n=111)
Number of deaths	290	223	67
Median survival, days (95% CI)	542 (459-645)	583 (447-725)	516 (368-666)
Restricted mean survival, days (95% CI)	867 (783-951)	895 (800-990)	758 (585-932)
Survival probability, % (95% CI)			
180 days	79.90 (76.42-83.53)	80.12 (76.19-84.25)	79.12 (71.89-87.09)
182 days	79.69 (76.20-83.34)	79.86 (75.91-84.01)	79.12 (71.89-87.09)
360 days	60.60 (56.34-65.19)	60.93 (56.13-66.15)	59.42 (50.68-69.65)



Characteristic	Overall (N=489)	Age <75 (n=378)	Age ≥75 (n=111)
365 days	60.13 (55.86-64.74)	60.34 (55.52-65.58)	59.42 (50.68-69.65)
540 days	50.00 (45.54-54.91)	51.11 (46.11-56.66)	45.97 (36.86-57.32)
547 days	49.72 (45.25-54.63)	51.11 (46.11-56.66)	44.65 (35.55-56.08)
720 days	43.25 (38.68-48.36)	44.80 (39.67-50.60)	37.56 (28.55-49.42)
730 days	42.19 (37.60-47.34)	43.48 (38.31-49.34)	37.56 (28.55-49.42)

Primary Survival Outcomes

The overall Rigshospitalet/Herlev Hospital cohort experienced 290 death events among 489 patients during the follow-up period, representing 59.3% mortality (Table 39). The median overall survival was 542 days (95% CI: 459-645). This median survival of approximately 17.8 months indicates that half of the Rigshospitalet/Herlev Hospital cohort survived beyond one and a half years after ICI initiation. The restricted mean survival time was 867 days (95% CI: 783-951), considerably exceeding the median and suggesting the presence of a substantial proportion of long-term survivors.

Age-stratified mortality analysis revealed differential patterns between age groups. The younger cohort (<75 years) experienced 223 death events among 378 patients, representing 59.0% mortality. The median survival was 583 days (95% CI: 447-725), approximately 19.2 months, with a notably wide confidence interval spanning from 447 to 725 days. This broad range reflects considerable heterogeneity in outcomes within the younger population. The restricted mean survival of 895 days (95% CI: 800-990) was the highest among all groups, indicating a subset of younger patients achieving particularly extended survival.

The elderly cohort (≥75 years) recorded 67 deaths among 111 patients, representing 60.4% mortality. The median survival was 516 days (95% CI: 368-666), shorter than the younger cohort by 67 days. The confidence interval ranging from 368 to 666 days demonstrates substantial variability in elderly patient outcomes. The restricted mean survival for the elderly cohort was 758 days (95% CI: 585-932), the lowest among all groups but still representing over 2 years of average survival time.



Figure 4: Kaplan-Meier survival curves by age group - Rigshospitalet/Herlev Hospital survival plot showing three curves with confidence intervals - overall (blue), <75 years (salmon), ≥75 years (green)

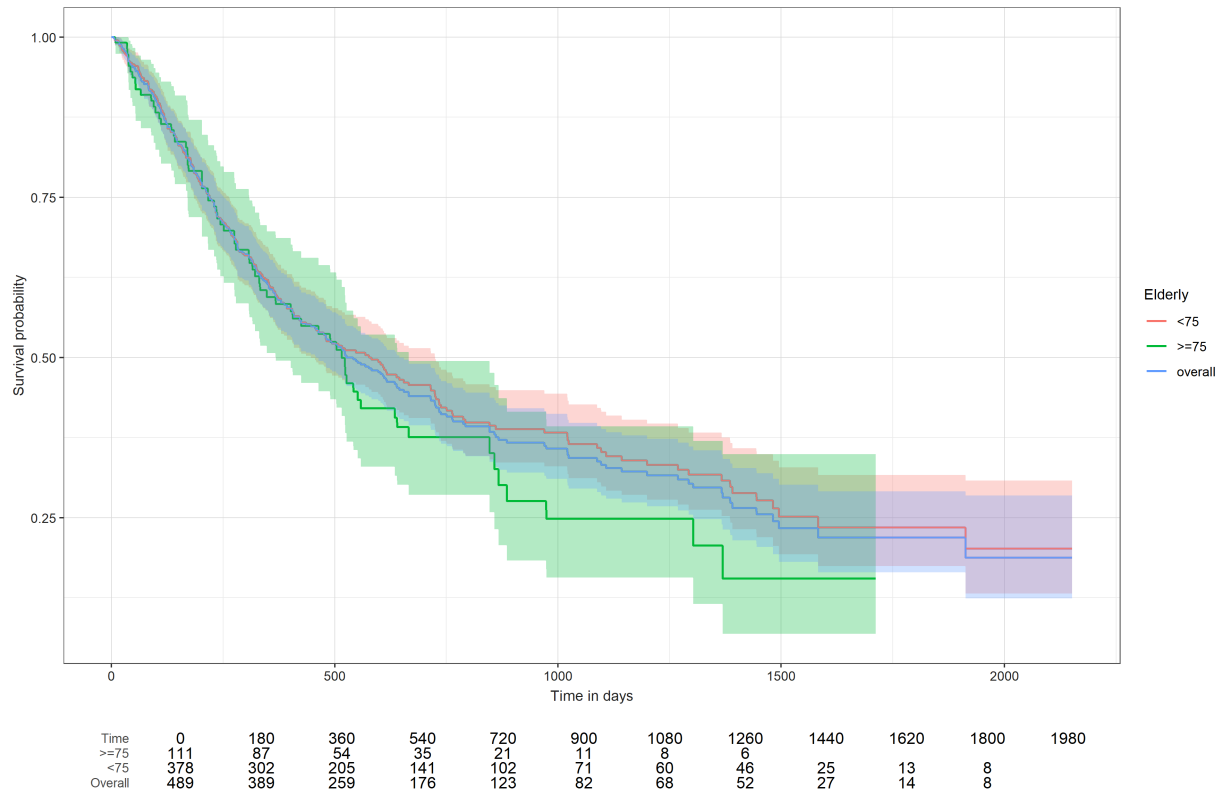


Table 38: Overall survival outcomes - Rigshospitalet/Herlev Hospital

Time Point	Overall		Age <75		Age ≥75	
	At Risk	Events*	At Risk	Events*	At Risk	Events*
0	489	0	378	0	111	0
180	389	98	302	75	87	23
360	259	89	205	69	54	20
540	176	42	141	31	35	11
720	123	22	102	16	21	6
900	82	17	71	13	11	-
1080	68	5	60	-	8	-
1260	52	5	46	5	6	0
1440	27	7	25	5	-	-
1620	14	-	13	-	-	0



Time Point	Overall		Age <75		Age ≥75	
1800	8	0	8	0	-	-

* Events represent deaths occurring in the interval from the previous timepoint; "-" indicates suppressed or unavailable values

Survival Probability Trajectories

At 180 days, the overall cohort maintained a survival probability of 79.90% (95% CI: 76.42-83.53) (Table 40). The minimal decline to 79.69% (95% CI: 76.20-83.34) at 182 days represents only a 0.21 percentage point decrease over 2 days, suggesting relatively stable mortality risk in this immediate period.

By the 1-year mark, survival probability had decreased to 60.13% (95% CI: 55.86-64.74) at 365 days, representing a 19.77 percentage point decline from 180 days. At 547 days (approximately 18 months), survival probability was 49.72% (95% CI: 45.25-54.63), maintaining nearly half of the cohort alive. By 730 days (2 years), survival probability was 42.19% (95% CI: 37.60-47.34).

Age-Stratified Survival Patterns

The younger cohort (<75 years) demonstrated robust survival throughout the observation period. Initial 180-day survival was 80.12% (95% CI: 76.19-84.25), with minimal change to 79.86% (95% CI: 75.91-84.01) at 182 days. This early survival exceeded both the overall cohort and the elderly group. By 365 days, survival probability was 60.34% (95% CI: 55.52-65.58), representing a 19.78 percentage point decline from 6 months.

Long-term outcomes in the younger cohort showed continued favorable survival. At 540 and 547 days, survival probability remained stable at 51.11% (95% CI: 46.11-56.66), indicating that more than half of younger patients survived beyond 18 months. By 730 days, survival probability was 43.48% (95% CI: 38.31-49.34).

The elderly cohort (≥75 years) exhibited distinct survival dynamics. Early survival at 180 and 182 days was 79.12% (95% CI: 71.89-87.09), with no change between these timepoints. Despite the older age, this 6-month survival was comparable to the younger cohort (79.12% vs 80.12%). At 365 days, survival remained similar between age groups at 59.42% (95% CI: 50.68-69.65).

However, divergence became apparent at later timepoints. By 547 days, elderly survival had declined to 44.65% (95% CI: 35.55-56.08), compared to 51.11% in the younger cohort. The 2-year survival of 37.56% (95% CI: 28.55-49.42) represented a 5.92 percentage point disadvantage compared to younger patients.

6.5.3 Overall Survival Analysis - HUS

Overall Survival Analysis

Overall survival (OS) was analyzed from the date of first ICI initiation following metastatic diagnosis (index date) until death from any cause.



Table 39: Overall survival outcomes - HUS

Characteristic	Overall (N=198)	Age <75 (n=152)	Age ≥75 (n=46)
Number of deaths	137	103	34
Median survival, days (95% CI)	380 (322-440)	404 (326-500)	284 (164-614)
Restricted mean survival, days (95% CI)	629 (518-740)	661 (531-790)	508 (307-710)
Survival probability, % (95% CI)			
180 days	64.96 (58.62-71.99)	66.87 (59.76-74.83)	58.70 (46.06-74.80)
182 days	64.96 (58.62-71.99)	66.87 (59.76-74.83)	58.70 (46.06-74.80)
360 days	52.56 (45.88-60.21)	54.51 (46.97-63.26)	46.37 (33.67-63.86)
365 days	50.75 (44.05-58.47)	52.18 (44.60-61.04)	46.37 (33.67-63.86)
540 days	36.84 (30.23-44.89)	37.27 (29.76-46.68)	35.46 (23.42-53.68)
547 days	36.84 (30.23-44.89)	37.27 (29.76-46.68)	35.46 (23.42-53.68)
720 days	26.69 (20.46-34.83)	27.16 (20.12-36.67)	25.07 (14.05-44.74)
730 days	25.81 (19.62-33.94)	26.03 (19.06-35.54)	25.07 (14.05-44.74)

Primary Survival Outcomes

The overall HUS cohort experienced 137 death events among 198 patients during the follow-up period, representing 69.2% mortality (Table 41). The median overall survival was 380 days (95% CI: 322-440). This median survival of approximately 12.5 months indicates that half of the HUS cohort had died within just over one year of ICI initiation. The restricted mean survival time was 629 days (95% CI: 518-740), substantially exceeding the median and indicating the presence of long-term survivors despite the overall high mortality rate.

Age-stratified mortality analysis revealed differential patterns between age groups. The younger cohort (<75 years) experienced 103 death events among 152 patients, representing 67.8% mortality. The median survival was 404 days (95% CI: 326-500), approximately 13.3 months, with a confidence interval spanning from 326 to 500 days. The restricted mean survival of 661 days (95% CI: 531-790) was the highest among age groups, suggesting a subset of younger patients achieving extended survival despite the overall high mortality.

The elderly cohort (≥75 years) recorded 34 deaths among 46 patients, representing 73.9% mortality, the highest mortality rate among all age groups across all three Nordic sites. The median survival was 284 days (95% CI: 164-614). The extremely wide confidence interval ranging from 164 to 614 days reflects both the small sample size and substantial heterogeneity in elderly patient outcomes. The restricted mean survival for the elderly cohort was 508 days (95% CI: 307-710), the lowest among all groups.



Figure 5: Kaplan-Meier survival curves by age group - HUS survival plot showing three curves with confidence intervals - overall (blue), <75 years (salmon), ≥75 years (green)

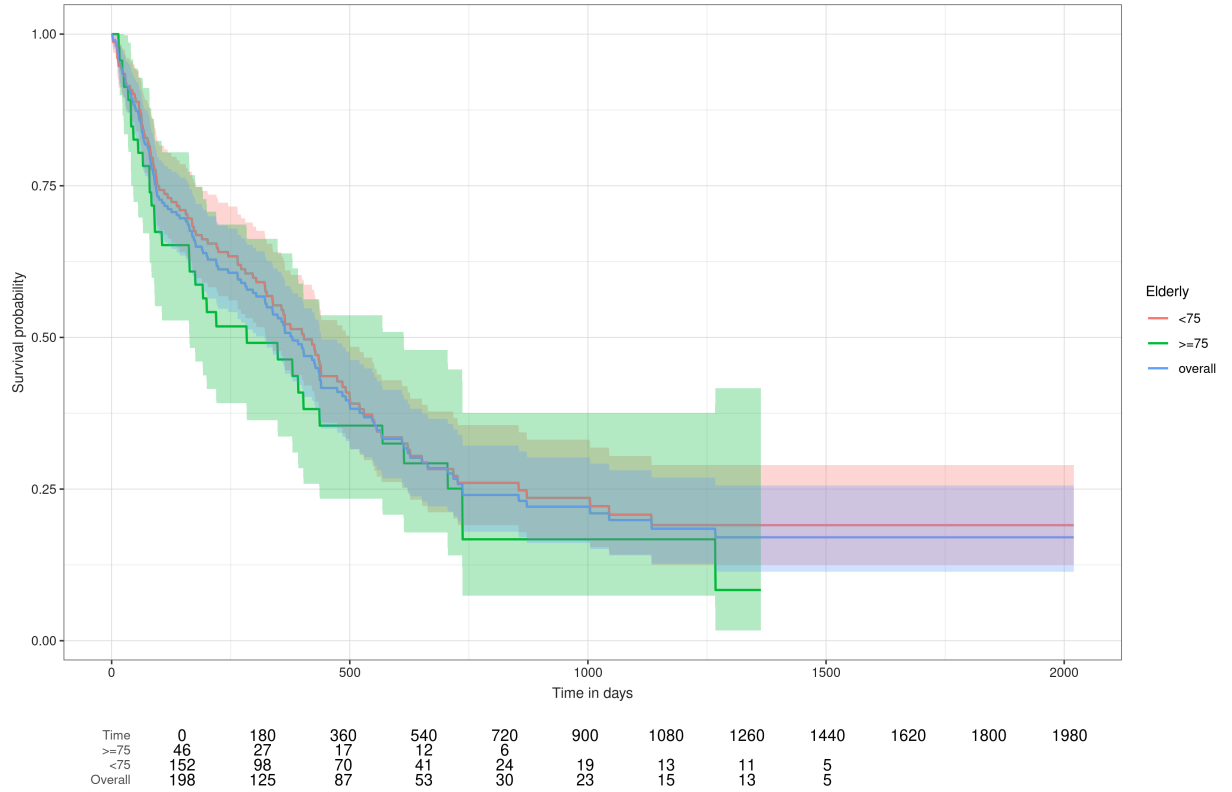


Table 40: Patients at risk and events over time - HUS

Time Point Days	Overall		Age <75		Age ≥75	
	At Risk	Events*	At Risk	Events*	At Risk	Events*
0	198	0	152	0	46	0
180	125	69	98	50	27	19
360	87	22	70	17	17	5
540	53	24	41	20	12	-
720	30	13	24	10	6	-
900	23	5	19	-	-	-
1080	15	-	13	-	-	0
1260	13	-	11	-	-	0
1440	5	-	5	0	-	-

* Events represent deaths occurring in the interval from the previous timepoint; "-" indicates suppressed or unavailable values



Survival Probability Trajectories

At 180 days, the overall cohort maintained a survival probability of 64.96% (95% CI: 58.62-71.99) (Table 42). The survival probability remained stable at 182 days (64.96%, 95% CI: 58.62-71.99), with no change over this 2-day interval.

By the 1-year mark, survival probability had declined to 50.75% (95% CI: 44.05-58.47) at 365 days, representing a 14.21 percentage point decline from 180 days. This represented a more gradual decline. At 547 days (approximately 18 months), survival probability was 36.84% (95% CI: 30.23-44.89), and by 730 days (2 years), survival probability had decreased to 25.81% (95% CI: 19.62-33.94).

Age-Stratified Survival Patterns

The younger cohort (<75 years) demonstrated modestly better survival throughout the observation period compared to the elderly. Initial 180-day survival was 66.87% (95% CI: 59.76-74.83), remaining unchanged at 182 days. By 365 days, survival probability was 52.18% (95% CI: 44.60-61.04), representing a 14.69 percentage point decline from 6 months.

Long-term outcomes in the younger cohort showed continued attrition. At 540 and 547 days, survival probability was 37.27% (95% CI: 29.76-46.68), indicating that approximately one-third of younger patients survived beyond 18 months. By 730 days, survival probability was 26.03% (95% CI: 19.06-35.54). Early survival at 180 and 182 days was 58.70% (95% CI: 46.06-74.80), the lowest 6-month survival among all elderly cohorts studied. At 365 days, survival had declined to 46.37% (95% CI: 33.67-63.86), remaining stable through 365 days. This represented a 12.33 percentage point decline from 6 months, a more gradual decline than observed in the younger cohort.

However, the elderly cohort showed minimal further decline after 1 year. By 547 days, survival was 35.46% (95% CI: 23.42-53.68), and at 730 days, it remained at 25.07% (95% CI: 14.05-44.74). The wide confidence intervals throughout (ranging from 14.05% to 44.74% at 2 years) reflect the small sample size of 46 patients.

6.6 Incomplete Study Objectives

Three pre-specified objectives could not be completed due to data limitations or time constraints:

Objective 2d(i): Cranial Irradiation Analysis

The planned analysis to describe the proportion of patients receiving cranial irradiation for mNSCLC post index date could not be completed. Initial data exploration revealed that specific radiotherapy site information, including cranial irradiation records, was not available in the OMOP-mapped databases. Radiotherapy data, where captured, did not include anatomical site specifications necessary to identify cranial irradiation.

Objective 3: Cox Regression Analysis

The planned Cox proportional hazards regression analysis to identify factors associated with overall survival could not be performed due to systematic unavailability of critical prognostic covariates across participating sites. The absence of these key prognostic factors precluded meaningful multivariable survival analysis, as crude survival comparisons across countries would be confounded by unmeasured differences in patient case-mix, disease severity, and performance status.



Objective 4: Healthcare Resource Utilization and Cost Analysis

The planned analysis of healthcare resource utilization (HCRU) and associated costs among patients diagnosed with mNSCLC receiving ICI as first-line treatment could not be performed. All participating sites reported that actual cost data were not available in their OMOP-mapped databases, as Nordic healthcare systems do not systematically capture procedure-level costs in clinical data warehouses. This objective was therefore not pursued following initial data quality assessment.

Alternative approaches for future research: Future Nordic federated HCRU studies could employ: (1) linkage to national administrative registries for comprehensive hospital admission data; (2) proxy-based estimation using documented clinical events (radiotherapy administrations, laboratory testing frequency) as surrogate markers for healthcare intensity; (3) prospective HCRU data enrichment protocols with standardized VISIT_OCCURRENCE mapping; or (4) diagnosis-related group (DRG)-based cost weighting derived from national tariffs.

7 Discussion

The execution of this pilot validates the technical infrastructure for federated OMOP-based analytics. The observed variations cannot be interpreted as reflecting true clinical or healthcare system differences due to fundamental data completeness gaps. Rather than generating hypotheses, these findings diagnostically identify critical prerequisites for future Nordic collaboration: standardized capture of prognostic covariates (ECOG, smoking history, PD-L1), validated algorithms for key classifications (emergent versus de novo disease), harmonized treatment intent and radiotherapy documentation, and prospective data quality monitoring. This represents progress in Nordic data collaboration, with pathways identified for scaling to larger cohorts, additional cancer types, and prospectively harmonized data collection protocols. The study's primary value lies in establishing the infrastructure and identifying specific data quality requirements for future Nordic comparative effectiveness research(44-46).

7.1 Principal Findings

Patient population: The final cohort comprised 755 patients (Rigshospitalet/Herlev Hospital 64.8%, HUS 26.4%, OUS 8.9%). Sex distribution varied, with Rigshospitalet/Herlev Hospital and HUS demonstrating female predominance (55.0% and 56.3% female, respectively), while OUS demonstrated male predominance (70.1% male). These differences may reflect combinations of site-specific referral patterns, temporal enrollment variations, or differences in catchment area demographics(13, 47).

BMI data completeness varied across sites, with OUS capturing 73.1% (49/67 patients), Rigshospitalet/Herlev Hospital 77.9% (381/489), and HUS 97.5% (194/199 patients), representing the highest BMI data completeness among the three Nordic sites. Within available BMI data, elderly patients in Rigshospitalet/Herlev Hospital demonstrated substantially greater variability (SD 11.1) compared to younger counterparts (SD 4.5). Median BMI values ranged from 23.7 kg/m² to 25.1 kg/m² across sites and age groups(48-50).

Comorbidity assessment using OMOP CDM medication-based proxies revealed differential performance across condition types. Cardiovascular medication documentation was near-universal (83.6% OUS, 97.8% Rigshospitalet/Herlev Hospital, 99.5% HUS), substantially higher than diagnosis-based cardiovascular disease prevalence (20.9%, 23.9%, 32.7% respectively). This pattern likely reflects broad-spectrum medication capture including supportive oncology agents such as antiemetics, corticosteroids, and medications for chemotherapy-associated complications, demonstrating the importance of condition-specific validation when using medication-based comorbidity proxies. In contrast, COPD demonstrated strong diagnosis-medication concordance in Rigshospitalet/Herlev Hospital and HUS, with both approaches yielding 11.9% and 18.6% prevalence respectively. This concordance suggests reliable capture for conditions with highly specific medication profiles. COPD prevalence varied across sites: 18.6% (HUS), 13.4% (OUS), and 11.9% (Rigshospitalet/Herlev Hospital), consistent with documented variations in Nordic smoking epidemiology patterns (51, 52). Diabetes medication prevalence varied from 10.4% (Norway) to 23.6% (HUS), likely reflecting site-specific documentation and coding practices rather than true population differences in diabetes burden. These patterns illustrate the variable performance of automated comorbidity algorithms across different condition types within

federated OMOP networks, highlighting the value of diagnosis-medication concordance validation for establishing data quality benchmarks (53).

Treatment patterns: Pembrolizumab-based regimens predominated across all sites with variation: 91.0% of OUS treatment sequences utilized pembrolizumab, 95.5% in HUS' first-line treatment, and 51.5% in Rigshospitalet/Herlev's first-line use. The pembrolizumab predominance in OUS and HUS may reflect alignment with guideline recommendations for PD-L1 positive mNSCLC (10, 54). Minimal or absent use of nivolumab and atezolizumab in some sites' treatment patterns (OUS: 0% nivolumab, 0% atezolizumab; Finland: 4.0% nivolumab, 0.5% atezolizumab) may represent site-specific formulary decisions, referral patterns for clinical trial participation, or temporal factors given the study period (2018-2023) spanning guideline evolution(55, 56). Interpreting these patterns requires caution given incomplete treatment intent documentation and incomplete data conversion to OMOP CDM.

Radiotherapy integration across healthcare systems: Documented radiotherapy utilization varied substantially across sites. Rigshospitalet/Herlev Hospital demonstrated radiotherapy documentation in 75.5% of patients, including use during first-line treatment (10.8%), between treatment lines, and during later therapies. HUS documented radiotherapy utilization in 66.3% of patients as first treatment, though temporal relationships between systemic therapy and radiotherapy initiation could not be definitively established from available sequence data.

OUS's 0% radiotherapy documentation represents a data capture limitation rather than absence of clinical radiotherapy practice, as explicitly acknowledged by the OUS site. The documented radiotherapy utilization in Rigshospitalet/Herlev Hospital and HUS may reflect multimodal care strategies including consolidation following chemoradiotherapy, palliative radiotherapy for symptomatic metastases, and management of oligometastatic disease (57). The durvalumab-containing sequences observed in Rigshospitalet/Herlev Hospital may represent consolidation therapy following concurrent chemoradiotherapy, consistent with established evidence (58).

The observed association between documented radiotherapy utilization and survival outcomes cannot be interpreted given OUS's complete absence of radiotherapy documentation representing a data quality limitation rather than clinical practice pattern. Future studies require complete treatment documentation across all sites, explicit radiotherapy indication capture (palliative versus consolidative versus oligometastatic-directed), and adjustment for performance status and disease characteristics.

Treatment duration and intensity patterns: Median first-line treatment durations varied across sites: OUS 100 days, HUS 52 days (Rigshospitalet/Herlev Hospital duration not available in extracted data). These duration calculations rely on proxy definitions using temporal gaps between administrations, which cannot distinguish between planned treatment protocols, dose delays due to toxicity, maintenance therapy strategies, or data recording practices(59). The apparent variation in first-line duration may reflect true treatment continuation differences, differing approaches to maintenance therapy, documentation practices affecting calculated duration, incomplete capture of treatment discontinuation events at all sites, or differences in patient populations across institutional settings(60).

The observed inverse relationship between first-line duration and second-line progression rates (HUS: 52 days median first-line duration with 46.7% receiving second-line; OUS: 100 days median first-line duration with 34.3% receiving second-line) may represent different



treatment approaches whose comparative effectiveness requires evaluation in controlled studies with complete performance status documentation (61, 62). However, incomplete data capture at all three sites, heterogeneous patient populations reflecting different institutional characteristics (tertiary referral center, regional comprehensive center, single-center referral hospital), and absence of prognostic covariate data preclude causal interpretation of these patterns.

Chemotherapy integration: Chemotherapy co-administration with immunotherapy varied: 86.7% (Rigshospitalet/Herlev Hospital), intermediate rates in HUS, and 56.5% (OUS). And third-line therapy was documented in 9.0% (OUS), 16.1% (Rigshospitalet/Herlev Hospital), and 31.2% (HUS) of patients. These variations may reflect differences in data capture completeness across institutional settings (tertiary center versus regional center versus single-center referral hospital), referral patterns for complex cases, follow-up duration variations, or treatment philosophies regarding fitness requirements for third-line therapy(63, 64). All three sites may experience incomplete patient journey capture when patients receive later-line therapies at community oncology settings, though the magnitude of this effect may differ by institutional integration. These variations cannot be definitively attributed to any single factor given current data limitations across all sites.

Age-stratified treatment approaches and outcomes: Elderly patients (≥ 75 years) received less intensive treatment across all three sites, with chemotherapy co-administration 14-18 percentage points lower than younger patients (< 75 years). This age-related treatment pattern was consistent across all Nordic sites despite similar universal healthcare frameworks.

- Chemotherapy co-administration: 14-18 percentage point reduction in elderly across all sites
- Second-line therapy progression: Younger patients 1.4-1.6 times more likely to receive subsequent therapy across sites
- Third-line therapy: Lower documentation rates in elderly populations across all sites
- Elderly patients demonstrated lower survival across all three sites. Median survival in elderly populations ranged from 197 days (HUS) to 427 days (Rigshospitalet/Herlev Hospital), with intermediate values in OUS

Potential explanations for age-related treatment patterns: In OUS's small elderly subgroup ($n=14$), documented first-line ICI duration appeared longer in elderly versus younger patients (127 vs 84 days), differing from patterns documented in other countries. This observation requires cautious interpretation given: (1) small sample size ($n=14$) with wide confidence intervals, (2) potential survival bias where only fitter elderly patients received extended treatment, (3) acknowledged incomplete patient journey capture similar to limitations affecting all sites, (4) data masking requirements preventing detailed age-stratified sequence analysis, and (5) possibility of different treatment regimen intensity allowing longer single-agent continuation. This observation should not inform clinical practice and requires validation in larger cohorts with complete treatment documentation and performance status capture across all sites.

The consistent age-related treatment patterns across Nordic sites—despite universal healthcare access—raise questions about optimal treatment selection for elderly mNSCLC patients that require prospective investigation. Studies incorporating comprehensive geriatric assessment(65-67) are warranted to develop evidence-based frameworks for treatment



selection. The absence or near-absence of performance status data across all sites in this pilot study prevented fitness-based analyses and represents a data quality gap for future iterations.

Survival outcomes: Median overall survival varied: 542 days in Rigshospitalet/Herlev Hospital (95% CI: 465-648), 380 days in HUS (95% CI: 309-504), and 315 days in OUS (95% CI: 220-485). Six-month survival probability ranged from 63.5% (OUS) to 80.0% (Rigshospitalet/Herlev Hospital). Two-year survival ranged from 20.9% (OUS) to 33.8% (Rigshospitalet/Herlev Hospital). These differences occurred within populations with comparable age distribution (median 67-70 years) and sex distribution (63.3%-70.1% male). However, the absence of ECOG performance status data in Rigshospitalet/Herlev Hospital and HUS, universal lack of smoking history documentation, and heterogeneous PD-L1 reporting preclude definitive attribution of survival differences to healthcare system performance versus patient selection or disease severity. In addition, the differences in attribution between the different sites might impact the selection of patients and hence the results. These variations may reflect complex interactions between healthcare delivery differences, unmeasured confounders, and data capture heterogeneity(68).

7.2 Limitations

The limitations described below must be interpreted within the dual-objective framework of this pilot study. While these constraints substantially affect the interpretability of clinical findings, they simultaneously represent the primary methodological achievement of the study: systematic identification and empirical quantification of data infrastructure enhancement requirements for future Nordic collaborative research. Each limitation documented here provides actionable intelligence for network development, transforming abstract data quality concepts into specific, measurable improvement targets.

The technical success of this pilot lies in demonstrating that federated OMOP analytics can reliably execute standardized queries across heterogeneous Nordic healthcare systems, systematically profile data completeness, and identify precise gaps preventing adjusted comparative effectiveness analyses. The limitations enumerated below should therefore be read as both constraints on current interpretation and specifications for future capability development.

Data Completeness and Variable Availability

This study faced substantial limitations in data completeness that affected both planned analyses and interpretation of findings. Performance status, a critical prognostic factor in NSCLC, was completely absent in Rigshospitalet/Herlev Hospital and HUS and only 53.7% complete in OUS, precluding risk-adjusted survival analyses and limiting assessment of treatment selection drivers. The universal absence of smoking history data prevented evaluation of this fundamental lung cancer risk factor and its relationship to treatment response. PD-L1 expression, the primary biomarker guiding ICI selection, could not be harmonized across sites due to heterogeneous testing platforms, reporting thresholds, and substantial missing data—with OUS having no evaluable records, Rigshospitalet/Herlev Hospital showing incompatible range definitions, and HUS capturing only 5.5% of patients.

The inability to complete planned Cox regression analyses due to incomplete variable characterization represents a major analytical limitation. Without multivariable adjustment for performance status, PD-L1 expression, detailed staging, and other prognostic factors, the observed survival differences between sites cannot be definitively attributed to healthcare



system factors versus patient selection or disease characteristics. Prospective harmonization of PD-L1 data capture—including standardized OMOP mapping conventions, documentation of testing platforms, and structured extraction workflows—should be prioritized in future Nordic collaborative studies to enable the adjusted comparative effectiveness analyses that were precluded in this pilot.

Comorbidity Assessment Methodology

The dual comorbidity assessment approach—using both diagnosis codes and medication prescriptions—was implemented to address known limitations in administrative data capture. Some healthcare systems demonstrate more complete medication documentation than diagnosis coding, particularly for chronic conditions managed in primary care. Medication records are typically captured comprehensively for reimbursement purposes, while diagnosis coding may be incomplete if not directly relevant to the oncology encounter. This could explain why cardiovascular medication prevalence exceeded 95% while diagnosis-based cardiovascular disease ranged from 20.9% to 32.7%.

However, this medication-based approach introduces its own biases. The near-universal cardiovascular medication use likely captures supportive care medications (antiemetics with QT-prolongation risk requiring cardiac monitoring, steroids requiring gastric protection counted as gastrointestinal medications, or prophylactic medications) rather than true cardiovascular comorbidity (69). The perfect concordance between diagnosis and medication-based COPD assessment in Rigshospitalet/Herlev Hospital and HUS suggests disease-specific reliability varies, with respiratory conditions requiring both specialist diagnosis and specific medications showing better agreement than cardiovascular conditions where medications have broader indications.

Emergent Versus De Novo Classification Bias

The 120-day threshold for distinguishing emergent from de novo metastatic disease, while providing operational consistency, systematically biases toward overestimating de novo presentation. Sites lacking complete patient journey capture or with poorly recorded initial diagnoses will misclassify emergent cases as de novo if the original non-metastatic diagnosis falls outside the captured timeframe. This particularly affects OUS, where acknowledged incomplete patient journey documentation means patients diagnosed at outside facilities and later referred for metastatic treatment would appear as de novo cases.

This misclassification bias likely explains part of the inverse relationship between cohort size and emergent disease prevalence (HUS 14.6%, OUS 11.9%, Rigshospitalet/Herlev Hospital 7.8%). Larger, more comprehensive databases like Rigshospitalet/Herlev Hospital's regional system capture more complete longitudinal histories, correctly identifying emergent cases, while smaller or referral-based databases may only capture patients at metastatic presentation. The wide range in time to metastatic progression in HUS (125-1959 days) suggests variable capture of historical diagnoses, with some patients having extensive historical documentation and others appearing de novo due to missing early-stage records.

Treatment Data Limitations and Proxy Measures

Several critical treatment-related analyses relied on proxy measures with inherent limitations. Chemotherapy cycles were estimated using gap-based proxies (14-day and 28-day



thresholds) rather than actual cycle documentation (Section 5.5.2.2.3). This approach may overestimate cycles when dose delays exceed thresholds or underestimate when planned treatment breaks fall within thresholds. Maintenance therapy poses particular challenges, as continuous weekly administration might be counted as multiple cycles while extended-interval dosing could be undercounted. Additionally, the proxy definitions for chemotherapy episodes cannot distinguish between different treatment intentions (neoadjuvant, adjuvant, palliative) or specific regimen types (dose-dense vs conventional). Small emergent cohort sizes (OUS N=8, Rigshospitalet/Herlev Hospital N=38, HUS N=29) result in extensive data masking, particularly affecting age-stratified analyses where elderly subgroups are limited (Rigshospitalet/Herlev Hospital emergent elderly n=10, OUS emergent elderly n=5).

The complete absence of radiotherapy data in OUS (0% documentation) represents a critical data infrastructure failure rather than clinical reality. Radiotherapy is a standard component of metastatic NSCLC care in OUS, but these services are documented in clinical systems not interfaced with the OMOP CDM-mapped database queried for this study or delivered at external facilities whose records are not captured in the federated analysis framework. This data gap entirely precludes assessment of actual radiotherapy utilization patterns in the OUS cohort and fundamentally limits the interpretability of cross-country treatment pattern comparisons. This missing radiotherapy data may partially explain OUS's shorter documented survival outcomes and apparently simplified treatment sequences, as radiotherapy often serves critical roles in metastatic NSCLC management including oligometastatic disease treatment, symptomatic palliation, and enabling subsequent systemic therapies. This missing data may partially explain OUS's shorter survival outcomes and simplified treatment sequences, as radiotherapy often serves as a bridge enabling subsequent systemic therapies.

Selection Bias and Generalizability Constraints

University hospital populations inherently differ from community-treated patients in disease complexity, performance status, and treatment eligibility. Each site exhibited distinct selection patterns that limit cross-national comparability. OUS's documented bias toward earlier-stage patients referred for curative treatment creates a paradox—the healthier baseline population showed poorer outcomes, suggesting either incomplete outcome capture or selection of patients with aggressive biology despite early stage. HUS's high prevalence of prior treatment exposure suggests referral of complex, potentially refractory cases, while Rigshospitalet/Herlev Hospital's comprehensive regional capture may provide a broader spectrum of care delivery across multiple facilities, potentially providing more representative population-level data than single-institution cohorts.

The fragmentation of care across participating and non-participating centers differentially affects sites. OUS explicitly acknowledged incomplete capture when patients received portions of care elsewhere, potentially missing radiotherapy, later-line treatments, or accurate survival data. Rigshospitalet/Herlev Hospital's regional integration suggests more complete capture, though patients seeking care outside the capital region would be missed. HUS's status as a tertiary referral center might capture complex cases while missing community-managed patients.

Methodological Constraints of Federated Analysis

The federated approach, while essential for data governance, prevented sophisticated statistical adjustments. Without individual-level data pooling, propensity score matching to balance baseline characteristics, multilevel modeling to account for site clustering, or competing risk analyses were impossible. Summary statistics aggregation limited exploration of effect modification and prevented post-hoc subgroup discoveries.

Privacy-mandated masking of cells with <5 patients created analytical dead zones, particularly affecting OUS's elderly population (n=14) and uncommon treatment sequences. The prohibition on deriving masked values prevented complete treatment pattern characterization. For instance, knowing that 8 emergent cases existed in OUS with 5 elderly patients should allow derivation of 3 younger patients, but reporting requirements prevented this, limiting comparative analyses.

Temporal and Practice Evolution Considerations

The study period encompassed substantial practice evolution including new drug approvals, changing PD-L1 testing standards, and COVID-19 disruptions. The pandemic (2020-2022) likely influenced treatment selection toward monotherapy and oral agents to minimize facility exposure, potentially differentially affecting sites based on local pandemic severity and healthcare system strain. These temporal factors were not addressed in analyses and may confound cross-site comparisons if sites experienced different pandemic impacts.

Documentation and Coding Variations

Systematic documentation differences created fundamental comparability challenges. HUS's capture of earlier-stage progression (Stages I-III: 9.5%) suggests either more comprehensive historical tracking or different patient populations than sites capturing only Stage IV disease. The reliance on administrative coding for biomarker status may miss results documented in unstructured notes or external laboratory systems.

The study could not distinguish between treatment intent (curative vs palliative), specific ICI dosing schedules, or discontinuation reasons (progression vs toxicity vs preference), limiting interpretation of duration differences. The absence of patient-reported outcomes prevented assessment of quality versus quantity of survival trade-offs inherent in treatment decisions.

Statistical and Analytical Limitations

The observational design precludes causal inference—unmeasured confounding by performance status, molecular markers, or socioeconomic factors likely contributes to observed differences. Disparate cohort sizes (Rigshospitalet/Herlev Hospital n=489, HUS n=199, OUS n=67) created unequal statistical precision, with OUS's wide confidence intervals preventing definitive conclusions. The lack of adjustment for multiple comparisons across numerous endpoints increases false discovery risk, though the consistent patterns across related outcomes suggest true differences.

Implications for Result Interpretation

These limitations indicate that observed variations reflect complex interactions between true healthcare system differences, selection biases, data capture variability, and unmeasured confounders. The 7.4-month survival difference between Rigshospitalet/Herlev Hospital and

OUS, while clinically meaningful, cannot be causally attributed to any single factor and may reflect differences in baseline disease severity, performance status, or other patient characteristics that were not captured uniformly across sites. Without adjustment for ECOG performance status (completely absent in Rigshospitalet/Herlev Hospital), smoking history (universally missing), and PD-L1 expression (heterogeneously reported), the survival differences could equally represent patient selection factors as healthcare delivery variations. Any apparent correlation between documented radiotherapy and outcomes cannot be assessed given fundamental radiotherapy data infrastructure limitations (0% capture in OUS, unknown completeness in Rigshospitalet/Herlev Hospital and HUS) and potential confounding by unmeasured factors.

Despite these constraints, this pilot demonstrates the technical feasibility of federated analytics infrastructure within Nordic healthcare systems. The observed differences reflect unmeasured confounding and data capture heterogeneity rather than interpretable clinical patterns. Future research with complete covariate capture is required before these data sources can support hypothesis generation or inform practice recommendations.

The limitations documented above preclude causal interpretation of observed survival differences and treatment pattern variations, rendering this study primarily a demonstration of federated analytical feasibility and a roadmap for future data quality enhancement rather than a basis for immediate clinical practice recommendations.

7.3 Conclusion

This pilot study successfully demonstrates the technical feasibility of federated OMOP-based analytics across Nordic healthcare systems, analyzing 755 patients with metastatic NSCLC receiving first-line immunotherapy. Cross-national variation was observed in median survival (315-542 days) captured radiotherapy data ranging from 0% (OUS - data not available) to 75.5% (Rigshospitalet/Herlev Hospital), documented radiotherapy utilization (0%-75.5%), and treatment sequencing approaches. However, critical data completeness gaps prevent these variations from supporting hypothesis generation. The findings serve primarily to establish data quality benchmarks and identify essential infrastructure improvements for future Nordic real-world evidence research.

The observed variations identify priorities for future Nordic collaboration: standardized capture of prognostic covariates to enable adjusted analyses, validated algorithms for key classifications including emergent versus de novo disease, harmonized treatment intent and radiotherapy documentation, and prospective data quality monitoring. This pilot's primary achievement lies in establishing the infrastructure and identifying specific data quality requirements for future Nordic comparative effectiveness research. Future studies with prospectively standardized covariate capture—particularly ECOG performance status, smoking history, and PD-L1 expression—will be positioned to investigate whether observed associations between treatment patterns and survival outcomes reflect causal relationships or result from unmeasured confounding by patient characteristics and disease severity. Additionally, systematic documentation of adherence to national versus institutional treatment protocols would enable assessment of whether observed practice variations reflect guideline implementation differences or clinician-level decision-making heterogeneity.



From Limitations to Network Enhancement Roadmap

The limitations documented throughout this pilot translate directly into a structured enhancement roadmap for the Nordic OMOP network. The data completeness gaps, while precluding adjusted analyses in this pilot phase, are tractable through prospective standardization protocols. Specifically, the universal absence of ECOG performance status can be addressed through structured EHR template modifications and clinical documentation workflows at participating sites. The heterogeneous PD-L1 capture requires network-level coordination to establish standardized OMOP mapping conventions, documentation of testing platforms, and harmonized extraction protocols that accommodate the reality of varied laboratory information systems across institutions.

The treatment data limitations, particularly the proxy-based chemotherapy cycle definitions and incomplete radiotherapy capture, highlight the need for enhanced procedural coding standards and inter-institutional data linkage protocols. These are not insurmountable technical barriers but rather reflect the early stage of Nordic healthcare system integration for research purposes. Denmark's comprehensive regional capture (75.5% radiotherapy documentation) demonstrates that complete treatment pathway documentation is achievable within Nordic systems when appropriate infrastructure investments are made.

The methodological constraints identified during this pilot provide the empirical foundation for a phased network enhancement strategy. Near-term priorities include: (1) standardized ECOG and smoking history capture protocols implemented at point of care, (2) harmonized PD-L1 data extraction with documented testing platform metadata, (3) validated algorithms for emergent versus de novo disease classification incorporating lookback period optimization, and (4) treatment intent and radiotherapy documentation standards enabling complete multimodal therapy characterization. Medium-term enhancements should focus on procedural coding standardization and inter-institutional linkage enabling comprehensive care pathway reconstruction even when patients receive treatment at multiple facilities.

Critically, none of these enhancements require abandonment of the federated OMOP architecture validated through this pilot. The technical infrastructure for distributed query execution, privacy-preserving analytics, and standardized results aggregation has been successfully demonstrated. Future studies building on this infrastructure with prospectively implemented data quality protocols will be positioned to generate the adjusted comparative effectiveness evidence that pilot-phase data completeness gaps prevented. This represents the intended trajectory of pilot research: initial infrastructure validation revealing specific enhancement requirements, followed by iterative improvement cycles yielding progressively robust evidence generation capabilities.

The successful execution of this pilot, despite documented data limitations, validates the core premise that Nordic healthcare systems can collaborate effectively through federated OMOP infrastructure. The clear articulation of enhancement requirements provides a roadmap for consortium decision-making regarding resource allocation, capability development timelines, and realistic expectations for subsequent study generations. This positions the Nordic OMOP network for sustainable maturation rather than disillusionment from unrealistic expectations about pilot-phase analytical capabilities. Subsequent studies with prospectively standardized covariate capture, particularly ECOG performance status, smoking history, and PD-L1 expression, will be positioned to generate clinically actionable evidence regarding optimal mNSCLC treatment strategies across Nordic healthcare systems.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2024;74(3):229-63.
2. Guo L, Zhu C, Cai L, Zhang X, Fang Y, Chen H, Yang H. Global burden of lung cancer in 2022 and projected burden in 2050. *Chinese Medical Journal*. 2024;137(21):2577-82.
3. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JH, Beasley MB, et al. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *Journal of thoracic oncology*. 2015;10(9):1243-60.
4. Schabath MB, Cote ML. Cancer progress and priorities: lung cancer. *Cancer epidemiology, biomarkers & prevention*. 2019;28(10):1563-79.
5. Kratzer TB, Bandi P, Freedman ND, Smith RA, Travis WD, Jemal A, Siegel RL. Lung cancer statistics, 2023. *Cancer*. 2024;130(8):1330-48.
6. Guo H, Li H, Zhu L, Feng J, Huang X, Baak JP. “How long have i got?” in stage IV NSCLC patients with at least 3 months up to 10 years survival, accuracy of long-, intermediate-, and short-term survival prediction is not good enough to answer this question. *Frontiers in Oncology*. 2021;11:761042.
7. Casal-Mouriño A, Ruano-Ravina A, Lorenzo-González M, Rodríguez-Martinez A, Giraldo-Osorio A, Varela-Lema L, et al. Epidemiology of stage III lung cancer: frequency, diagnostic characteristics, and survival. *Translational lung cancer research*. 2021;10(1):506.
8. Riely GJ, Wood DE, Ettinger DS, Aisner DL, Akerley W, Bauman JR, et al. Non–small cell lung cancer, version 4.2024, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network*. 2024;22(4):249-74.
9. Hendriks L, Kerr K, Menis J, Mok T, Nestle U, Passaro A, et al. Oncogene-addicted metastatic non-small-cell lung cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Annals of Oncology*. 2023;34(4):339-57.
10. Hendriks L, Kerr K, Menis J, Mok T, Nestle U, Passaro A, et al. Non-oncogene-addicted metastatic non-small-cell lung cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up☆. *Annals of Oncology*. 2023;34(4):358-76.
11. Friedlaender A, Perol M, Banna GL, Parikh K, Addeo A. Oncogenic alterations in advanced NSCLC: a molecular super-highway. *Biomarker research*. 2024;12(1):24.
12. Huang Q, Li Y, Huang Y, Wu J, Bao W, Xue C, et al. Advances in molecular pathology and therapy of non-small cell lung cancer. *Signal Transduction and Targeted Therapy*. 2025;10(1):186.
13. Christensen NL, Jekunen A, Heinonen S, Dalton SO, Rasmussen TR. Lung cancer guidelines in Sweden, Denmark, Norway and Finland: a comparison. *Acta Oncologica*. 2017;56(7):943-8.
14. Thunnissen E, Weynand B, Udovicic-Gagula D, Brcic L, Szolkowska M, Hofman P, et al. Lung cancer biomarker testing: perspective from Europe. *Translational lung cancer research*. 2020;9(3):887.



15. Penault-Llorca F, Socinski MA. Emerging molecular testing paradigms in non-small cell lung cancer management—current perspectives and recommendations. *The Oncologist*. 2025;30(3):oyae357.
16. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or chemotherapy for non–small-cell lung cancer with mutated EGFR. *New England journal of medicine*. 2010;362(25):2380-8.
17. Guo H, Zhang J, Qin C, Yan H, Luo X, Zhou H. Advances and challenges of first-line immunotherapy for non-small cell lung cancer: A review. *Medicine*. 2024;103(3):e36861.
18. Olivares-Hernández A, Del Portillo EG, Tamayo-Velasco Á, Figuero-Pérez L, Zhilina-Zhilina S, Fonseca-Sánchez E, Miramontes-González JP. Immune checkpoint inhibitors in non-small cell lung cancer: from current perspectives to future treatments—a systematic review. *Annals of Translational Medicine*. 2023;11(10):354.
19. Lim SW, Ahn M-J. Current status of immune checkpoint inhibitors in treatment of non-small cell lung cancer. *The Korean journal of internal medicine*. 2018;34(1):50.
20. Mathew M, Safyan RA, Shu CA. PD-L1 as a biomarker in NSCLC: challenges and future directions. *Annals of translational medicine*. 2017;5(18):375.
21. Patel SP, Kurzrock R. PD-L1 expression as a predictive biomarker in cancer immunotherapy. *Molecular cancer therapeutics*. 2015;14(4):847-56.
22. Niu M, Yi M, Li N, Luo S, Wu K. Predictive biomarkers of anti-PD-1/PD-L1 therapy in NSCLC. *Experimental hematology & oncology*. 2021;10(1):18.
23. Yin J, Song Y, Fu Y, Jun W, Tang J, Zhang Z, et al. The efficacy of immune checkpoint inhibitors is limited in elderly NSCLC: a retrospective efficacy study and meta-analysis. *Aging (albany NY)*. 2023;15(24):15025.
24. Voruganti T, Soulos PR, Mamtani R, Presley CJ, Gross CP. Association between age and survival trends in advanced non–small cell lung cancer after adoption of immunotherapy. *JAMA oncology*. 2023;9(3):334-41.
25. Conforti F, Pala L, Bagnardi V, Viale G, De Pas T, Pagan E, et al. Sex-based heterogeneity in response to lung cancer immunotherapy: a systematic review and meta-analysis. *JNCI: Journal of the National Cancer Institute*. 2019;111(8):772-81.
26. Mezquita L, Auclin E, Ferrara R, Charrier M, Remon J, Planchard D, 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*. 2018;4(3):351-7.
27. Zhang X, Beachler DC, Masters E, Liu F, Yang M, Dinh J, et al. Health care resource utilization and costs associated with advanced or metastatic nonsmall cell lung cancer in the United States. *Journal of Managed Care & Specialty Pharmacy*. 2022;28(2):255-65.
28. Nesline MK, Knight T, Colman S, Patel K. Economic burden of checkpoint inhibitor immunotherapy for the treatment of non–small cell lung cancer in US clinical practice. *Clinical therapeutics*. 2020;42(9):1682-98. e7.
29. Casaletto J, Bernier A, McDougall R, Cline MS. Federated analysis for privacy-preserving data sharing: a technical and legal primer. *Annual review of genomics and human genetics*. 2023;24(1):347-68.



30. Laugesen K, Ludvigsson JF, Schmidt M, Gissler M, Valdimarsdottir UA, Lunde A, Sørensen HT. Nordic health registry-based research: a review of health care systems and key registries. *Clinical epidemiology*. 2021;533-54.
31. Planchard D, Popat S, Kerr K, Novello S, Smit E, Faivre-Finn C, et al. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*. 2018;29:iv192-iv237.
32. Sarnola K, Koskinen H, Klintrup K, Astrup C, Kurko T. Uptake and availability of new outpatient cancer medicines in 2010–2021 in Nordic countries—survey of competent authorities. *BMC Health Services Research*. 2023;23(1):1437.
33. Bliddal M, Bjørk E, Karlstad Ø, Wastesson JW, Wesselhoeft R, Lindahl-Jacobsen R, et al. Comparison of sociodemographic factors, healthcare utilisation by general practitioner visits, somatic hospital admissions, and medication use in Norway, Sweden, and Denmark. *Annals of Epidemiology*. 2024;98:1-7.
34. Maret-Ouda J, Tao W, Wahlin K, Lagergren J. Nordic registry-based cohort studies: Possibilities and pitfalls when combining Nordic registry data. *Scandinavian journal of public health*. 2017;45(17_suppl):14-9.
35. Ludvigsson JF, Håberg SE, Knudsen GP, Lafolie P, Zoega H, Sarkkola C, et al. Ethical aspects of registry-based research in the Nordic countries. *Clinical epidemiology*. 2015:491-508.
36. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Annals of internal medicine*. 2010;153(9):600-6.
37. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Studies in health technology and informatics*. 2015;216:574.
38. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *Journal of the American Medical Informatics Association*. 2012;19(1):54-60.
39. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc*. 2015;22(3):553-64.
40. Reich C, Ostropelets A, Ryan P, Rijnbeek P, Schuemie M, Davydov A, et al. OHDSI Standardized Vocabularies—a large-scale centralized reference ontology for international data harmonization. *J Am Med Inform Assoc*. 2024;31(3):583-90.
41. Hripcsak G, Shang N, Peissig PL, Rasmussen LV, Liu C, Benoit B, et al. Facilitating phenotype transfer using a common data model. *J Biomed Inform*. 2019;96:103253.
42. Hripcsak G, Schuemie MJ, Madigan D, Ryan PB, Suchard MA. Drawing reproducible conclusions from observational clinical data with OHDSI. *Yearbook of medical informatics*. 2021;30(01):283-9.
43. (OHDSI) OHDSal. 2019 [Available from: <https://ohdsi.github.io/TheBookOfOhdsi/>].
44. Lee GH, Park J, Kim J, Kim Y, Choi B, Park RW, et al. Feasibility study of federated learning on the distributed research network of OMOP common data model. *Healthcare Informatics Research*. 2023;29(2):168-73.

45. González AR, Robles V, Mercado JJC, Pérez JMM, Mendez JLG, Ruiz EM, editors. ELADAIS: An Integrated Platform for High-Impact Clinical Data Extraction, Standardization and Advanced Analytics Using OMOP-CDM. 2025 IEEE 38th International Symposium on Computer-Based Medical Systems (CBMS); 2025: IEEE.
46. Kent S, Burn E, Dawoud D, Jonsson P, Østby JT, Hughes N, et al. Common problems, common data model solutions: evidence generation for health technology assessment. *Pharmacoeconomics*. 2021;39(3):275-85.
47. Gee K, Yendamuri S. Lung cancer in females—sex-based differences from males in epidemiology, biology, and outcomes: a narrative review. *Translational Lung Cancer Research*. 2024;13(1):163.
48. Martin L, Birdsell L, MacDonald N, Reiman T, Clandinin MT, McCargar LJ, et al. Cancer cachexia in the age of obesity: skeletal muscle depletion is a powerful prognostic factor, independent of body mass index. *Journal of clinical oncology*. 2013;31(12):1539-47.
49. Shoji F, Miura N, Matsubara T, Akamine T, Kozuma Y, Haratake N, et al. Prognostic significance of immune-nutritional parameters for surgically resected elderly lung cancer patients: a multicentre retrospective study. *Interactive cardiovascular and thoracic surgery*. 2018;26(3):389-94.
50. Morita-Tanaka S, Yamada T, Takayama K. The landscape of cancer cachexia in advanced non-small cell lung cancer: a narrative review. *Translational Lung Cancer Research*. 2023;12(1):168.
51. Backman H, Vanfleteren L, Lindberg A, Ekerljung L, Stridsman C, Axelsson M, et al. Decreased COPD prevalence in Sweden after decades of decrease in smoking. *Respiratory research*. 2020;21(1):283.
52. Knudsen AK, Allebeck P, Tollånes MC, Skogen JC, Iburg KM, McGrath JJ, et al. Life expectancy and disease burden in the Nordic countries: results from the Global Burden of Diseases, Injuries, and Risk Factors Study 2017. *The Lancet Public Health*. 2019;4(12):e658-e69.
53. Henke E, Zoch M, Peng Y, Reinecke I, Sedlmayr M, Bathelt F. Conceptual design of a generic data harmonization process for OMOP common data model. *BMC Medical Informatics and Decision Making*. 2024;24(1):58.
54. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, et al. NCCN guidelines insights: non–small cell lung cancer, version 2.2021: featured updates to the NCCN guidelines. *Journal of the National Comprehensive Cancer Network*. 2021;19(3):254-66.
55. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic non–small-cell lung cancer. *New England journal of medicine*. 2018;378(22):2078-92.
56. Mok TS, Wu Y-L, Kudaba I, Kowalski DM, Cho BC, Turna HZ, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *The Lancet*. 2019;393(10183):1819-30.
57. Palma DA, Olson R, Harrow S, Gaede S, Louie AV, Haasbeek C, et al. Stereotactic ablative radiotherapy for the comprehensive treatment of oligometastatic cancers: long-term results of the SABR-COMET phase II randomized trial. *Journal of Clinical Oncology*. 2020;38(25):2830-8.



58. Antonia SJ, Villegas A, Daniel D, Vicente D, Murakami S, Hui R, et al. Durvalumab after chemoradiotherapy in stage III non–small-cell lung cancer. *New England Journal of Medicine*. 2017;377(20):1919-29.
59. Griffith SD, Tucker M, Bowser B, Calkins G, Chang C-h, Guardino E, et al. Generating real-world tumor burden endpoints from electronic health record data: comparison of RECIST, radiology-anchored, and clinician-anchored approaches for abstracting real-world progression in non-small cell lung cancer. *Advances in therapy*. 2019;36(8):2122-36.
60. Liu Z, Lee K, Cohn D, Zhang M, Ai L, Li M, et al. Analysis of real-world data to investigate evolving treatment sequencing patterns in advanced non-small cell lung cancers and their impact on survival. *Journal of Thoracic Disease*. 2023;15(5):2438.
61. Tyczynski JE, Potluri R, Kilpatrick R, Mazumder D, Ghosh A, Liede A. Incidence and risk factors of pneumonitis in patients with non-small cell lung cancer: an observational analysis of real-world data. *Oncology and Therapy*. 2021;9(2):471-88.
62. Cramer-van der Welle CM, Verschueren MV, Tonn M, Peters BJ, Schramel FM, Klungel OH, et al. Real-world outcomes versus clinical trial results of immunotherapy in stage IV non-small cell lung cancer (NSCLC) in the Netherlands. *Scientific Reports*. 2021;11(1):6306.
63. Girard N, Jacoulet P, Gainet M, Elleuch R, Pernet D, Depierre A, et al. Third-line chemotherapy in advanced non-small cell lung cancer: identifying the candidates for routine practice. *Journal of Thoracic Oncology*. 2009;4(12):1544-9.
64. Sacher AG, Le LW, Lau A, Earle CC, Leigh NB. Real-world chemotherapy treatment patterns in metastatic non–small cell lung cancer: are patients undertreated? *Cancer*. 2015;121(15):2562-9.
65. Wildiers H, Heeren P, Puts M, Topinkova E, Janssen-Heijnen ML, Extermann M, et al. International Society of Geriatric Oncology consensus on geriatric assessment in older patients with cancer. *Journal of clinical oncology*. 2014;32(24):2595-603.
66. Mohile SG, Dale W, Somerfield MR, Schonberg MA, Boyd CM, Burhenn PS, et al. Practical assessment and management of vulnerabilities in older patients receiving chemotherapy: ASCO guideline for geriatric oncology. *Journal of clinical oncology*. 2018;36(22):2326-47.
67. Loh K, Liposits G, Arora S, Neuendorff N, Gomes F, Krok-Schoen J, et al. Adequate assessment yields appropriate care—the role of geriatric assessment and management in older adults with cancer: a position paper from the ESMO/SIOG Cancer in the Elderly Working Group. *ESMO open*. 2024;9(8):103657.
68. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *New England Journal of Medicine*. 2008;358(8):771-83.
69. Bang JH, Hwang S-H, Lee E-J, Kim Y. The predictability of claim-data-based comorbidity-adjusted models could be improved by using medication data. *BMC Medical Informatics and Decision Making*. 2013;13(1):128.