



Funded by the
Nordic Council
of Ministers

VALO2 non-small cell lung cancer (NSCLC) pilot study

Study report version 1.0

VALO2 NSCLC pilot study (Cross-border health
research project using SPEs)

Value from Nordic health data – VALO2 project

30 September 2026



Document info

0.1 Authors

Author	Partner
Argyri Iris Mathioudaki, Epidemiologist Consultant	IQVIA
Citadel Jungco Cabasag, Epidemiologist	IQVIA
Collaborator / Buyer Representative	
Mikael Rinnetmäki, Expert, Project Lead	Sitra
Persephone Doupi, Senior Medical Officer	THL
Principal Investigator	
Andreas Bjerrum, Main Principal Investigator	Rigshospitalet
IQVIA project responsible parties	
Elina Winblad, Project Director	IQVIA
Niina Leinonen, Project Manager	IQVIA

0.2 Keywords

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

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 © 2026 The VALO2 Project – contributing parties and authors identified in this report.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC), (<https://creativecommons.org/licenses/by-nc/4.0/>). Other uses require the written permission of THL and Sitra.

Table of Contents

Table of Contents.....	2
List of Abbreviations.....	4
Executive summary.....	6
1 Introduction	7
1.1 Disease burden and epidemiology	7
1.2 Molecular landscape and biomarker-driven treatment.....	7
1.3 First-line immunotherapy for patients with PD-L1 $\geq 50\%$ mNSCLC	8
1.4 Treatment challenges and patient heterogeneity	8
1.5 Study rationale	9
2 Objectives	11
2.1 Study aim	11
2.2 Primary objective	11
2.3 Secondary objectives.....	11
2.4 Exploratory objectives.....	11
2.5 Infrastructure objective	12
3 Methods.....	13
3.1 Study design	13
3.2 Data sources.....	13
3.3 Federated analysis framework	14
3.4 Study population	14
3.5 Study variables	16
3.6 Treatment variables	18
3.7 Study outcomes	18
3.8 Statistical analysis.....	19
3.9 Data quality assessment.....	23
3.10 Ethical and governance considerations	24
4 Results	25
4.1 Study population and eligibility attrition	25
4.2 Characteristics of the study population	27
4.3 Overall survival (primary objective)	31
4.4 Treatment patterns and treatment sequences (secondary objective)	36
4.5 Other clinical outcomes: TTNT and TTD (secondary objective).....	38
4.6 Exploratory Prognostic Analyses (Exploratory Objective)	44
4.7 Infrastructure feasibility findings (infrastructure objective).....	45
5 Discussion.....	47
5.1 Main findings.....	47
5.2 Primary objective: overall survival	47
5.3 Secondary objectives: treatment patterns, TTNT, and TTD	48
5.4 Exploratory objective: prognostic analyses	50
5.5 Infrastructure lessons learned	50
5.6 Limitations	51
6 Conclusions.....	54
6.1 Future research.....	54
References.....	55
Appendices	57
Appendix A. Study population and baseline characteristics	58
Appendix B. Country-specific treatment-pattern tables.....	65
Appendix C. Site-specific survival analyses	67
Appendix D. Model diagnostics and assumptions.....	77

7 Statistical annex.....	85
7.1 Treatment patterns.....	85
7.2 Time to event analysis	87
7.3 Average treatment effect	89
7.4 Sensitivity analysis: inverse probability of censoring weighting.....	90
7.5 Handling missing data	91
7.6 Pooled analysis.....	94
7.7 References.....	95

List of Abbreviations

Abbreviation	Definition
ALK	Anaplastic lymphoma kinase
BMI	Body mass index
BRAF	v-Raf murine sarcoma viral oncogene homolog B
CDM	Common Data Model
COPD	Chronic obstructive pulmonary disease
ECOG-PS	Eastern Cooperative Oncology Group performance status
EGFR	Epidermal growth factor receptor
EHDS	European Health Data Space
EHR	Electronic health record
HER2	Human epidermal growth factor receptor 2
HR	Hazard ratio
HUS	Helsinki University Hospital
ICI	Immune checkpoint inhibitor
IPTW	Inverse probability of treatment weighting
KM	Kaplan–Meier
KRAS	Kirsten rat sarcoma viral oncogene homolog
LIPI	Lung Immune Prognostic Index
LIPS-3	Lung Immuno-Oncology Prognostic Score
MCAR	Missing Completely At Random
MET	Mesenchymal–epithelial transition
mNSCLC	Metastatic non-small cell lung cancer
NA	Not available
NSCLC	Non-small cell lung cancer
NTRK	Neurotrophic tyrosine receptor kinase
OMOP	Observational Medical Outcomes Partnership
OS	Overall survival
OUS	Oslo University Hospital
PD-1	Programmed cell death protein 1
PD-L1	Programmed death-ligand 1

Abbreviation	Definition
RET	Rearranged during transfection
ROS1	c-ROS oncogene 1
RWE	Real-world evidence
SA	Statistical Annex
SACT	Systemic anti-cancer therapy
sHR	Subdistribution hazard ratio
SPEs	Secure processing environments
TRE	Trusted Research Environment
TTD	Time to treatment discontinuation
TTNT	Time to next line-of-therapy
VALO	Value from Nordic Health Data

Executive summary

Background: Metastatic non-small cell lung cancer is a serious disease with high mortality. In recent years, immune checkpoint inhibitors (ICIs) have become an important treatment option for patients whose tumours have high PD-L1 expression. However, information on how these treatments perform in routine clinical practice remains limited.

Objectives: VALO2 aimed to improve understanding of treatment outcomes among patients receiving first-line immunotherapy for metastatic lung cancer in routine Nordic healthcare. The study also evaluated whether health data from different Nordic countries can be analysed using a common research approach without transferring patient-level data between institutions.

Methods: VALO2 was a multicentre observational study conducted using electronic health record data from Rigshospitalet (Denmark), Helsinki University Hospital (Finland), and Oslo University Hospital (Norway). Data remained within each participating institution and were analysed through secure local environments. Researchers accessed and analysed harmonised Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)-formatted data remotely, while only aggregated study results were exported for reporting. Compared with the previous VALO1 pilot, VALO2 demonstrated that cross-border research can be conducted through direct analytical access to data within participating Secure Processing Environments (SPEs), while maintaining local data custody.

Results: The final study population included 887 patients from Denmark, 91 patients from Finland, and 76 patients from Norway. In the Danish cohort, the pembrolizumab policy-period group was associated with longer overall survival and more favourable treatment-related outcomes than the atezolizumab policy-period group. In the Finnish cohort, no statistically significant differences were observed between pembrolizumab and cemiplimab. In the Norwegian cohort, patients receiving pembrolizumab had longer overall survival than those receiving cemiplimab. However, this finding should be interpreted with caution because the cemiplimab group was small and the available information for adjustment was limited.

The study also showed that important prognostic factors could be evaluated when sufficient data were available. However, the availability of clinical variables differed between participating sites, demonstrating the continuing challenges of conducting comparable analyses across countries. The participating hospitals successfully implemented a common analytical framework, showing that real-world studies can be performed across national borders without transferring patient-level data between institutions.

Conclusions: VALO2 demonstrated that cross-border real-world data research is feasible using harmonised health data mapped to the OMOP Common Data Model, and local SPEs across Nordic healthcare institutions.

The study provided new information on treatment outcomes among the patients receiving first-line immunotherapy in routine clinical practice while highlighting the importance of data availability, cohort implementation and reporting requirements when comparing results across countries. The experience gained through VALO2 supports further development of Nordic federated research infrastructure and demonstrates how secure cross-border access to health data can enable future multinational research while maintaining local control of sensitive patient information.

1 Introduction

1.1 Disease burden and epidemiology

Lung cancer remains one of the most common cancers worldwide and is still the leading cause of cancer-related death. Non-small cell lung carcinoma (NSCLC) accounts for approximately 85% of lung cancer diagnoses and includes several tumour types, most commonly adenocarcinoma and squamous cell carcinoma [3,6].

Although screening, diagnostics, surgery, radiotherapy, systemic treatment, and supportive care have improved over time, many patients are still diagnosed with advanced or metastatic disease. Others progressed after treatment for earlier-stage disease. Outcomes in the metastatic non-small cell lung carcinoma (mNSCLC) setting remain poor, which highlights the need for improved prevention, effective treatment strategies, and a better understanding of patient outcomes in routine clinical practice. [3,5,6].

The treatment landscape for NSCLC has changed substantially over the past 2 decades. Molecularly targeted therapies and immune checkpoint inhibitors (ICIs) have improved outcomes for selected patient groups. However, benefits may vary across patient populations, and clinical trial participants do not always fully represent patients treated in routine oncology practice, including older individuals and those with comorbidities or poorer performance status [5,6].

1.2 Molecular landscape and biomarker-driven treatment

Treatment for metastatic non-small-cell lung cancer (mNSCLC) has moved from a mainly histology- and chemotherapy-based approach to one that is increasingly guided by molecular testing and predictive biomarkers. This change reflects a better understanding of the biological diversity of NSCLC and the identification of genomic alterations that can guide treatment choice [5,6].

In current clinical practice, biomarker testing is recommended for patients with advanced non-squamous NSCLC and for selected patients with squamous histology when clinically relevant. Testing commonly includes oncogenic drivers such as epidermal growth factor receptor (EGFR) mutations, anaplastic lymphoma kinase (ALK) rearrangements, c-ROS oncogene 1 (ROS1) rearrangements, and other alterations including v-Raf murine sarcoma viral oncogene homolog B (BRAF), mesenchymal–epithelial transition (MET), rearranged during transfection (RET), human epidermal growth factor receptor 2 (HER2), Kirsten rat sarcoma viral oncogene homolog (KRAS), neurotrophic tyrosine receptor kinase (NTRK), and related genomic markers where these are available [4,6,10].

Finding an actionable driver alteration is clinically important because targeted therapies can provide substantial benefit in molecularly defined patient groups. When an approved targeted option is available, targeted therapy is generally prioritised over immunotherapy-based treatment [4,6]. For the patients without actionable driver alterations, programmed death-ligand 1 (PD-L1) expression is a key biomarker for first-line immunotherapy selection, particularly when tumour PD-L1 expression is high [4,14].

In practice, implementation of biomarker-driven treatment still depends on how consistently biomarker testing is performed and recorded within routine-care. Factors such as testing availability, laboratory workflows, reimbursement, and documentation practices can all influence the completeness of real-world data. These issues are especially important in real-world studies, where incomplete or inconsistent biomarker data can affect cohort definition, treatment classification, and interpretation of outcomes [10].

1.3 First-line immunotherapy for patients with PD-L1 $\geq 50\%$ mNSCLC

ICIs targeting the programmed cell death protein 1 (PD-1) and PD-L1 pathways have changed first-line treatment for patients with advanced or metastatic NSCLC without actionable driver alterations. In patients with high tumour PD-L1 expression, commonly defined as PD-L1 expression $\geq 50\%$, ICI monotherapy is an established treatment option supported by pivotal clinical trials and international treatment guidelines [4,10].

Pembrolizumab was the first PD-1 inhibitor to show durable clinical benefit as first-line monotherapy in this setting. Long-term follow-up from the KEYNOTE clinical development programme showed sustained survival among patients with advanced NSCLC treated with pembrolizumab [2]. These findings supported the broad use of pembrolizumab as a standard first-line option for patients with PD-L1 $\geq 50\%$ mNSCLC and no targetable EGFR or ALK alterations [4,10].

Atezolizumab, a PD-L1 inhibitor, has also shown benefit as first-line treatment for patients with advanced NSCLC, including PD-L1-selected populations, and is included in current treatment pathways for appropriate patients [7,15]. Cemiplimab, another PD-1 inhibitor, expanded the first-line monotherapy options for patients with advanced NSCLC and PD-L1 expression of at least 50%, with evidence from the EMPOWER-Lung clinical trial programme supporting its use in this population [8,14].

Pembrolizumab, atezolizumab, and cemiplimab are all used as first-line monotherapy options for patients with PD-L1 $\geq 50\%$ mNSCLC [4,10,18,7,14]. Despite their widespread use, there are currently no randomised head-to-head trials directly comparing pembrolizumab, atezolizumab, and cemiplimab in patients with PD-L1 $\geq 50\%$ mNSCLC. In routine Nordic practice, treatment use and preferred PD-(L)1 inhibitor strategies have varied over time and between countries, reflecting national reimbursement decisions, hospital procurement processes, administrative pricing policies, clinical judgement, treatment schedules, and local implementation timelines.

Differences in treatment adoption across Nordic countries provide a natural opportunity to examine how these therapies perform in routine clinical practice. The Nordic healthcare setting is well suited for this type of work because it combines population-based healthcare systems, structured electronic health data, and variation in treatment adoption across sites. At the same time, this variation should not be viewed as a random treatment assignment. Any comparative findings need cautious interpretation, with attention to policy timing, local practice, and potential patient selection.

1.4 Treatment challenges and patient heterogeneity

Patients with mNSCLC that are treated in routine oncology care are highly heterogeneous and often differ from those enrolled in pivotal randomised trials. Age, performance status, comorbidities, smoking history, tumour burden, histology, biomarker availability, and prior treatment history can all influence treatment choice and clinical outcomes [4,5,6].

Older patients are particularly important in mNSCLC because lung cancer incidence increases with age, while older adults are often underrepresented in clinical trials. Evidence from systematic reviews and real-world studies suggests that ICIs can provide clinical benefit in older patients, although outcomes may be influenced by frailty, comorbidities, concomitant medicines, organ function, and differences in treatment discontinuation or monitoring [9,13,16,17].

This patient heterogeneity also creates challenges for observational research. Differences between treatment groups at baseline may reflect clinical decision-making, reimbursement rules, prescribing preferences, or treatment availability rather than treatment effects alone. For that reason, careful description of demographic, clinical, biomarker, and treatment related variables is essential when interpreting real-world outcomes [4,10].

Stratified analyses of clinically important patient subgroups, including older patients and patients with differing health status, can help determine whether outcomes are broadly consistent across routine-care populations. These analyses are especially useful for patients who are less well represented in trials but make up a substantial part of the real-world mNSCLC population [9,13,16,17].

1.5 Study rationale

The Value from Nordic Health Data (VALO)2 study builds upon the experience from VALO1, which showed that federated analyses using harmonised Nordic healthcare data are technically feasible. Prior VALO1 work also indicated that technical feasibility alone is not enough to generate clinically interpretable cross-site evidence. Important lessons included the need for transparent cohort attrition, structured reporting of data completeness, harmonised biomarker and prognostic covariate capture, and cautious interpretation of country-specific treatment and survival patterns when Eastern Cooperative Oncology Group performance status (ECOG-PS), smoking status, PD-L1 documentation, treatment-line reconstruction, or radiotherapy data are incomplete or inconsistently captured across sites. Where VALO1 mainly focused on infrastructure development and characterisation of metastatic NSCLC populations across participating sites, VALO2 extends this work to clinically relevant questions about treatment patterns and outcomes among patients receiving first-line ICI monotherapy while also testing and assessing cross-border access for remotely analysing data in local secure environments.

Real-World Evidence (RWE) is particularly relevant in this setting because patients treated in routine oncology practice may differ from those enrolled in pivotal clinical trials. Routine healthcare data can provide insights into treatment utilisation, treatment sequencing, and clinical outcomes in broader patient populations, including older patients and those with comorbidities or variable performance status who may be underrepresented in clinical trials [1]. Specifically, VALO2 aims to generate Nordic RWE for patients with PD-L1 $\geq 50\%$ mNSCLC treated with pembrolizumab, atezolizumab, or cemiplimab. The study uses country- and site-level variation in first-line ICI use, including variation related to policy, pricing, reimbursement, and local clinical practice, to describe treatment patterns and evaluate outcomes in routine-care.

Through collaboration between university hospitals in Denmark, Finland, and Norway, VALO2 aims to characterise treatment patterns, evaluate clinical outcomes, assess outcomes among elderly patients, and explore treatment comparisons where the data and methods support this. In parallel, the study evaluates the operational feasibility of federated observational research using cross-border access for Nordic Secure Processing Environments (SPEs) and harmonised Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) infrastructure. This supports the development of a sustainable framework for cross-border RWE generation in the Nordic region.

VALO2 therefore addresses 2 related questions: how patients receiving first-line ICI monotherapy perform in routine Nordic care, and whether the Nordic OMOP and SPEs infrastructure can support future multi-country RWE studies.

2 Objectives

2.1 Study aim

The aim of this study was to compare the effectiveness of PD-1 and PD-L1 inhibitor monotherapy as first-line treatment for mNSCLC within and across Nordic countries (Denmark, Finland, and Norway) and subgroups defined by demographics. The study evaluated pembrolizumab-, atezolizumab-, and cemiplimab-defined exposure groups individually and where appropriate across participating Nordic countries.

Objectives were evaluated by site and by ICI exposure-group according to the site-specific analytical definitions. To fulfil the study aims outlined above, the study objectives were explored and stratified by:

- Country
- ICI monotherapy treatment group (pembrolizumab, atezolizumab, and cemiplimab) based on intention-to-treat assignment.
- In accordance with the protocol, subgroup analyses were planned for patients aged <75 years and those aged ≥75 years at initiation of ICI treatment.

2.2 Primary objective

To compare the overall survival (OS) among patients with mNSCLC and PD-L1 ≥50% treated with pembrolizumab, atezolizumab, or cemiplimab as first-line monotherapy in routine clinical practice.

2.3 Secondary objectives

- To describe demographic and clinical characteristics of patients with mNSCLC and PD-L1 ≥50% initiating pembrolizumab, atezolizumab or cemiplimab.
- To describe treatment outcomes, including OS, time to next treatment (TTNT), and time to treatment discontinuation (TTD), overall and within predefined age subgroups, where feasible.
- To describe treatments patterns and subsequent treatments received after first-line PD-(L)1 inhibitor (programmed cell death 1 or programmed death-ligand 1) monotherapy (mainly pembrolizumab, atezolizumab, or cemiplimab).

2.4 Exploratory objectives

To assess whether Lung Immune Prognostic Index (LIPI) and/or Lung Immuno-Oncology Prognostic Score (LIPS-3) add explanatory value in risk adjustment between treatment subgroups. In accordance with the protocol, exploratory analyses by age category (<75 years and aged ≥75 years at initiation of ICI treatment), were undertaken where feasible. [19,20]

2.5 Infrastructure objective

In addition to the clinical objectives, this study assessed the feasibility of accessing Nordic Health Data sources through SPEs. The study documented process outcomes related to the SPEs infrastructure, including:

- Time from ethics approval to data-access.
- Proportion of protocol-specified variables successfully extracted.
- Cross-site data quality concordance metrics.

3 Methods

3.1 Study design

VALO2 was a retrospective observational multisite study designed to assess clinical outcomes in adult patients diagnosed with mNSCLC who received first-line ICI monotherapy. The study observation period spanned 01 January 2018 to 31 December 2024, with patient identification through 30 June 2024 allowing at least 6 months of potential follow-up period before the administrative end of study individual follow-up could end earlier because of death, loss to follow-up, or an outcome-specific censoring event.

The study included analytical cohorts from Rigshospitalet in Denmark, Helsinki University Hospital (HUS) in Finland, and Oslo University Hospital (OUS) in Norway. The study used routinely collected electronic health record (EHR) data harmonised to the OMOP CDM, enabling implementation of a standardised analytical approach across participating sites. Results were reported separately for each participating site, with aggregation performed only where methodologically appropriate and scientifically valid.

In addition to evaluating clinical outcomes, VALO2 assessed the feasibility of cross-border use of SPEs in the Nordics for RWE studies. The Danish research team analysed patient-level health data held by HUS in Finland and OUS in Norway through the secure environments used by those organisations. The Danish researchers also completed the corresponding analysis using Danish data within their established TRE environment.

Clinical outcomes were evaluated using site-specific analytical cohorts and exposure-group definitions. At Rigshospitalet, exposure groups were assigned according to predefined treatment-policy periods and did not necessarily represent the PD-(L)1 inhibitor administered. At HUS and OUS, exposure groups represented the first-line PD-(L)1 inhibitor initiated in clinical practice. The site-specific exposure definitions are described in Sections 3.5 and 3.7.

3.2 Data sources

Clinical data were obtained from three participating Nordic healthcare institutions:

- Rigshospitalet, Denmark, representing patients from the Capital Region of Denmark and Region Zealand.
- HUS, Finland, representing patients treated within the HUS catchment area, including Helsinki and other hospitals in the Uusimaa Region.
- OUS, Norway

The analytical cohorts represent participating institutional datasets and should not be interpreted as nationally representative populations.

Each participating institution retained patient-level information within its own secure infrastructure, and all respective databases were converted to the OMOP CDM. Patient-identifiable information was not shared between sites.

Analyses were performed using a federated approach whereby standardised analytical scripts were developed centrally by Danish researchers and executed locally within each participating SPE

by the Danish researchers through approved remote access mechanisms. Only aggregated and anonymised analytical outputs were exported for study reporting.

3.3 Federated analysis framework

The study employed a federated analytical model based on SPEs and Trusted Research Environment (TRE) principles. Under this framework:

- Patient-level data remained within each participating institution.
- Data-access was restricted to approved researchers.
- All analyses were conducted within local secure computing environments: Authorised Danish researchers analysed data remotely within the local environments, while patient-level information remained at the originating institution.
- Only anonymised and aggregated results were exported.
- Activities performed within each environment were logged and documented.

The federated cross-border design allowed Danish researchers to analyse data locally via remote connection while meeting each site's governance and data-protection requirements. Patient-level information remained at the originating institution throughout the analytical process and was not transferred between participating sites.

To facilitate cross-site consistency, all participating sites mapped their data to the OMOP CDM and applied common analytical definitions and study algorithms. Despite this common framework, the availability of individual variables and analytical outputs differed across sites. [11,12]

3.4 Study population

The study population comprised adult patients with histologically or cytologically confirmed mNSCLC with PD-L1 expression $\geq 50\%$ who received pembrolizumab, atezolizumab, or cemiplimab as first-line monotherapy at one of the participating Nordic University hospitals. The target population included both patients with de novo metastatic disease and patients who progressed to metastatic disease following an earlier-stage NSCLC diagnosis.

3.4.1 Inclusion criteria

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

- NSCLC diagnosis, defined as:
 - diagnosis of lung cancer, at any stage, documented at any time before or within 30 days after the metastatic disease diagnosis; and
 - histologically or cytologically confirmed NSCLC.
- diagnosed with PD-L1 $\geq 50\%$ mNSCLC and treated with at least one cycle of pembrolizumab, atezolizumab, or cemiplimab initiated in the mNSCLC setting.

Note: PD-L1 status was required for patient inclusion. The completeness of PD-L1 data and the potential effect of systematic PD-L1 missingness on cohort representativeness were assessed as part of study implementation.

- Metastatic disease, including:
 - de novo mNSCLC, defined as metastatic disease at the initial NSCLC diagnosis; or
 - progression to mNSCLC following an earlier-stage NSCLC diagnosis.
 - age ≥ 18 years at initiation of pembrolizumab, atezolizumab, or cemiplimab.
 - initiation of pembrolizumab, atezolizumab, or cemiplimab between 1 January 2018 and 30 June 2024.

For elderly-subpopulation analyses, the additional criterion was age ≥ 75 years at initiation of pembrolizumab, atezolizumab, or cemiplimab.

3.4.2 Exclusion criteria

Patients were excluded if they met any of the following exclusion criteria:

- Recorded opt-out from the use of their data for research purposes.
- Documentation in the medical record of any of the following genetic alterations:
 - EGFR, ALK, or ROS1 alterations.
- Receipt of targeted therapy at any time.
- For patients with other primary cancers, prior-malignancy exclusion criteria were applied according to the approach determined during feasibility assessment and intended for consistent implementation across participating sites.

3.4.3 Site-specific cohort implementation

The protocol-defined eligibility criteria were implemented locally using data available within each site-specific OMOP-formatted analytical dataset. Where a protocol-defined eligibility characteristic was not directly available, a predefined site-specific derivation approach was applied in consultation with the local site team.

At Rigshospitalet, selected cohort-defining and exclusion criteria were available in the exported aggregate outputs only as combined attrition steps. At HUS, metastatic status was derived using available treatment and surgical-history information. PD-L1 eligibility was inferred from local first-line ICI monotherapy prescribing practice because PD-L1 value results were not available as discrete records in the analytical output.

These site-specific implementation considerations affected the reporting of intermediate attrition stages and were considered when interpreting final cohort size, cohort composition, and cross-site comparability. Consequently, the final analytical cohorts reflect protocol eligibility applied within the available site-specific analytical datasets and do not necessarily represent all patients treated at the participating institutions.

3.4.4 Cohort definition & index date

Across participating sites, the index date corresponded to the first administration date of the relevant PD-(L)1 inhibitor and served as the starting point for all time-to-event analyses and treatment pathway evaluations. Treatment exposure-group definitions differed across participating sites based on local treatment practices and the predefined analytical approach.

In detail, in the Rigshospitalet cohort, patients were assigned to exposure groups according to predefined treatment-policy periods that reflected routine treatment adoption patterns in Danish clinical practice during the study period. Patients diagnosed and initiating first-line treatment during periods in which pembrolizumab was the recommended PD-(L)1 inhibitor were assigned to the pembrolizumab policy-period group, whereas patients diagnosed during periods in which atezolizumab was the recommended PD-(L)1 inhibitor were assigned to the atezolizumab policy-period group. The policy-period definitions were derived according to the protocol-specified treatment adoption approach and did not necessarily reflect the PD-(L)1 inhibitor ultimately administered.

At the same time, in both HUS and OUS cohorts, exposure groups were defined according to PD-(L)1 inhibitor monotherapy initiated in clinical practice (pembrolizumab or cemiplimab).

Patients were followed from the index date until the earliest occurrence of death, loss to follow-up, administrative end of study (31 December 2024), or outcome-specific censoring events. Follow-up and censoring rules varied according to the outcome under evaluation and are described in Section 3.8.

3.5 Study variables

The variables of interest were prespecified in the study protocol and have been selected to enable comprehensive characterisation of the patient population while ensuring feasible data collection across participating sites. Variables were grouped into demographic, clinical, disease-related, treatment-related, and prognostic domains. Variables collected for the study are summarised in Table 1.

3.5.1 Variable availability and implementation

Variables were derived from data standardised to the OMOP CDM. Variable availability differed across participating sites according to local data capture, documentation practices, and OMOP implementation. Variable availability reported in this study refers to availability within the OMOP-formatted analytical datasets and does not necessarily imply availability or absence of information in the underlying source systems.

Local disclosure-control and reporting requirements could affect which aggregated results were exportable for reporting. Such restrictions relate to reportable outputs and do not necessarily indicate that the corresponding variables were unavailable for analysis.

Although additional genomic biomarkers were specified in the protocol, participating datasets recorded only positive genomic test results. Consequently, an absent result could not be distinguished from a negative result, a missing result, or a test that had not been performed, and the denominator of tested patients could not be established. Following investigator review, genomic biomarkers other than PD-L1 were therefore excluded from baseline characteristics and analytical

models because test completeness and denominators could not be reliably established across participating datasets.

Eligibility for inclusion in the cohort required evidence of PD-L1 expression $\geq 50\%$. However, detailed quantitative PD-L1 expression values above this threshold were not consistently available in the reportable analytical datasets across participating sites. Consequently, PD-L1 status was used primarily for cohort eligibility but was not included as an adjustment variable in the analytical models. The availability of PD-L1 information differed across participating datasets and, at some sites, cohort eligibility was established using site-specific implementation procedures as described in Section 3.4.

Table 1: Protocol-specified study variables and domains

Domain	Variables
Demographic characteristics	Age at treatment initiation, sex, smoking status
Anthropometric characteristics	Height, weight, body mass index (BMI)
Clinical characteristics/comorbidities*	Cardiovascular disease, chronic obstructive pulmonary disease (COPD), diabetes mellitus
Disease characteristics	Histological subtype, disease stage at diagnosis, ECOG-PS, PD-L1 expression
Genomic biomarkers**	EGFR, ALK, ROS1, BRAF, MET, RET, HER2, KRAS, NTRK, and other protocol-specified genomic biomarkers
Anti-cancer treatment history	Prior systemic anti-cancer therapy (SACT), prior cancer-related surgery, prior radiotherapy
Treatment exposure	Pembrolizumab, atezolizumab, or cemiplimab as first-line monotherapy
Treatment patterns	Subsequent treatment lines, treatment duration, treatment sequencing
Prognostic variables	LIPI, LIPS-3

*Comorbidities were identified primarily from diagnosis records and supplemented with treatment information where appropriate. ** Availability and analytical use of genomic biomarker information is described in Section 3.6

3.5.2 Prognostic scores

Where sufficient laboratory and clinical information was available, the following prognostic scores were calculated:

- LIPI [19]
- LIPS-3 [20]

The proportion of patients with sufficient information to calculate each score was documented. Availability of LIPI and LIPS-3 differed across participating sites and is reported in the results section.

3.6 Treatment variables

The primary treatment exposure groups were defined according to the site-specific analytical approach described in Section 3.5.

In the Rigshospitalet analytical cohort, exposure groups were assigned according to the predefined treatment-policy periods and included pembrolizumab, atezolizumab policy-period groups. In HUS and OUS analytical cohorts, exposure groups reflected the first-line PD-(L)1 inhibitor actually initiated as first-line monotherapy for mNSCLC. Treatment history preceding metastatic disease and treatments received following first-line immunotherapy were also assessed.

Treatment-related variables specified in the protocol included:

- prior systemic anti-cancer therapy (SACT);
- prior cancer-related surgery;
- prior radiotherapy;
- first-line PD-(L)1 inhibitor exposure-group
- subsequent lines of anti-cancer treatment;
- treatment duration; and
- treatment sequences

Patients were assigned to exposure groups according to the site-specific analytical definitions described in Section 3.5. Exposure-group assignment remained fixed throughout follow-up irrespective of subsequent treatment modifications or treatment changes. Where data were available, patients were followed across subsequent treatment lines to describe observed treatment sequences.

3.7 Study outcomes

3.7.1 Primary outcome

The primary outcome was overall survival (OS). OS was defined as the time from index date until death from any cause. Patients without an observed death were censored at the earliest of the last known date alive or the end of the study period.

3.7.2 Secondary outcomes

Secondary outcomes included TTNT or death and TTD or death.

3.8.2.1 Time to Next Treatment (TTNT) or Death

TTNT was defined as the time from initiation of first-line ICI treatment until initiation of a subsequent line of anti-cancer therapy or death from any cause, whichever occurred first.

Subsequent treatment was defined according to the line-of-therapy algorithms specified in the Statistical Annex ([SA](#)). Patients not experiencing TTNT were censored according to the outcome-specific rules defined in the SA.

3.8.2.2 Time to Treatment Discontinuation or Death (TTD)

TTD was defined as the time from initiation of first-line ICI treatment until permanent discontinuation of treatment, treatment switch, treatment augmentation, or death from any because whichever occurred first.

Treatment discontinuation rules, including allowable treatment gaps and handling of treatment modifications, were defined according to the [SA](#).

3.7.3 Additional outcomes

Additional descriptive outcomes included:

- treatment duration;
- treatment sequencing;
- patterns of care following first-line immunotherapy.

Treatment duration was evaluated from initiation of first-line ICI therapy until treatment discontinuation according to protocol-defined treatment-line algorithms.

3.7.4 Censoring and competing-risk rules

Censoring and competing-risk rules were applied according to the outcome under evaluation.

For OS, death from any cause was considered the event of interest. Patients without an observed death were censored at the earliest of the last known date alive or administrative end of study period. Where identifiable, diagnosis of new primary malignancy resulted in censoring at the date of diagnosis.

For TTNT the event was defined as initiation of a qualifying subsequent line-of-therapy or death from any cause, whichever occurred first. Patients without a qualifying next line-of-therapy event were censored at the earliest of the last documented healthcare contact date or the end of the study period.

For TTD, the event was defined as permanent discontinuation of index therapy, treatment switch, treatment augmentation, or death from any cause, whichever occurred first. Patients without a TTD event were censored at the last date known to be receiving treatment, the last documented healthcare contact date, or the end of the study period, whichever occurred first.

Loss to follow-up was defined as absence of further information following the last documented healthcare contact date and resulted in censoring at that date. Because death formed part of the TTNT and TTD composite endpoint definitions, new primary malignancy was treated as the competing event where identifiable. Detailed endpoint definitions, treatment-line algorithms, censoring rules, and competing-risk specifications are provided in the [SA](#).

3.8 Statistical analysis

Analyses were conducted by Danish researchers separately within each participating site SPE using harmonised analytical code and standardised variable definitions. Patient-level data were not pooled across sites, and only aggregated and anonymised analytical outputs were exported from

the local environments. Results were therefore reported separately by site. Aggregation across sites was considered following scientific review of data quality, completeness, exposure definitions, variable availability, and analytical feasibility. No formal comparability assessment framework or scoring tool was applied. Because material differences remained across participating sites, results were reported separately and pooled estimates were not generated.

Detailed analytical specifications, including treatment-line algorithms, endpoint definitions, regression methods, covariate selection, missing-data handling, partial-date rules, subgroup feasibility criteria, model diagnostics, and sensitivity analyses, are provided in the VALO2 [SA](#).

Continuous variables were summarised using the number of patients with available data, mean, standard deviation, median, first and third quartiles, minimum, and maximum. Categorical variables were summarised using frequencies and percentages. The denominator used for each summary was reported, and missing-data were presented explicitly. Counts below the applicable site-specific disclosure threshold were suppressed in accordance with local disclosure-control requirements.

3.8.1 Primary objective: overall survival

OS was summarised using the Kaplan–Meier (KM) method. KM outputs included survival curves, numbers at risk, survival percentiles, median survival where estimable, survival probabilities at selected time points, and corresponding 95% confidence intervals. Treatment groups were compared using two-sided log-rank tests.

Treatment group associations with OS were evaluated using Cox proportional hazards regression models. Results were presented as hazard ratios (HRs), 95% confidence intervals, and p-values. Adjusted analyses incorporated baseline covariates selected according to the protocol, [SA](#), site-specific variable availability, missingness, the number of observed events, and investigator review.

Candidate covariates considered for adjustment, as prespecified in the [SA](#), included age, sex, ECOG performance status, smoking status, height and weight, disease stage, histology, cardiovascular disease, chronic obstructive pulmonary disease, diabetes, prior SACT, prior cancer-related surgery, prior radiotherapy, index year, and prognostic variables where available. Final model specification differed across participating sites according to variable availability, missingness, event counts, and protocol-defined feasibility criteria. As a result, identical adjustment sets could not be applied across participating sites, and model specification was determined separately for each analytical cohort.

3.8.2 Secondary objectives: treatment patterns and sequences

First-line treatment distributions were summarised using counts and percentages. Subsequent treatment lines of therapy (LoT) were identified using the treatment-line algorithms specified in the [SA](#).

A subsequent LoT was defined based on treatment switching, initiation of a previously unused anti-cancer therapy, addition of another therapy, or initiation of treatment after the prespecified allowable gap. Treatment sequences were summarised by the treatment group and LoT. Treatment duration was summarised using medians and interquartile ranges.

Treatment-pattern analyses were descriptive. No causal interpretation was assigned to differences in treatment distribution, sequencing, or duration across the site-specific analytical cohorts.

3.8.3 Secondary objectives: TTNT and TTD

Time to next line-of-therapy or death and TTD or death were evaluated as composite time-to-event outcomes.

Cumulative-incidence functions were estimated using the Aalen–Johansen estimator in the presence of competing events. New primary malignancy was treated as a competing event where identifiable. Fine–Gray subdistribution hazard regression models were used to evaluate treatment group and covariate associations with the cumulative-incidence of TTNT and TTD. Results were presented as subdistribution HRs (sHRs), 95% confidence intervals, and p-values.

TTNT was calculated from the index date to initiation of a qualifying subsequent line of anti-cancer treatment or death from any cause, whichever occurred first. TTD was calculated from the index date to permanent discontinuation of index monotherapy, treatment switch, treatment augmentation, or death from any cause, whichever occurred first. Treatment discontinuation was defined using the 60-day allowable treatment gap specified in the [SA](#).

Inverse probability of censoring weighting was prespecified as a potential sensitivity approach for informative censoring. It was not implemented in the available analyses because very few patients were non-administratively censored.

3.8.4 Exploratory prognostic analyses

Exploratory analyses assessed whether LIPI and, where available, LIPS-3 were associated with OS, TTNT, and TTD after adjustment for treatment group and other available covariates.

LIPI and LIPS-3 were included as covariates in Cox proportional hazards or Fine–Gray models where the scores could be calculated and the available number of events supported stable model estimation. Associations were summarised using HRs or sHRs, as appropriate, with corresponding 95% confidence intervals and p-values [19,20].

These analyses assessed whether categories of the prognostic scores were associated with the evaluated outcomes after adjustment. They were exploratory and were not designed to establish treatment-effect modification, validate the prognostic scores, or support causal inference.

3.8.5 Subgroup analyses

Analyses were conducted separately within each participating site-specific analytical cohort. Prespecified subgroup analyses were planned according to treatment group, participating site, and age category (<75 years and ≥75 years at treatment initiation).

Age-stratified analyses were performed when at least 10 events were observed per category, in accordance with the [SA](#). Analyses were omitted where this threshold was not met or where variable availability did not support valid estimation.

3.8.6 Handling of missing-data and partial dates

Missingness was summarised by participating site, treatment group, and variable. Missing values for categorical baseline characteristics were reported as a separate category where appropriate, and denominators were stated explicitly.

Complete-case analysis refers to analyses restricted to patients with available values for all covariates included in the relevant model. This approach was considered when missingness across included covariates was below 5%.

Clinically important baseline covariates considered for adjustment included protocol-specified demographic, clinical, disease-related, treatment history, and prognostic variables such as age, sex, ECOG performance status, smoking status, comorbidities, histology, stage, prior treatment history, and LIPI or LIPS-3 where available.

Multiple-imputation was considered for clinically important baseline covariates where missingness exceeded 5% but was generally no greater than 40%, and where a missing-at-random (MAR) assumption was considered plausible. Covariates with more than 40% missingness were not imputed and were excluded from multivariable adjustment models.

Where applied, missing baseline covariates were imputed using Multiple-Imputation by Chained Equations (MICE). The imputation model included the covariates used in the analytical model, treatment exposure, and outcome indicators without imputing the outcome itself. Imputation was performed separately within each site, consistent with the federated analytical framework. Estimates from imputed datasets were combined using Rubin's rules.

Missing-data patterns were assessed using descriptive summaries and, where feasible, Little's test of Missing-Completely-at-Random (MCAR test) of data Missing-Completely-at-Random. In this context, MCAR refers to missingness that is unrelated to either observed or unobserved patient characteristics. Results from these assessments informed the choice between complete-case and multiple-imputation approaches. Results from these assessments informed the choice between complete-case and imputation-based analyses. Partial dates were handled according to the prespecified rules in the [SA](#). Where the year and month were available, but the day was missing, the day was assigned as the 15th of the month. Time-to-event outcomes were not calculated where the month was missing, or the date was otherwise insufficient to support valid endpoint derivation. Imputed dates were constrained so that they did not conflict with other known dates, including birth and death dates.

3.8.7 Model diagnostics and sensitivity analyses

Model diagnostics were conducted to assess model assumptions, stability, and robustness. For Cox proportional hazards models, the proportional hazards assumption was evaluated using Schoenfeld residual tests and residual plots. Model fit was assessed using Cox-Snell residuals, and influential observations were evaluated using deviance residuals. Where evidence of non-proportional hazards was identified, additional time-stratified analyses were performed where appropriate.

For inverse probability of treatment weighting (IPTW) analyses, weight distributions and the positivity assumption were assessed. Extreme weights were reviewed and could be trimmed according to the prespecified approach. Post-weighting covariate balance was evaluated using standardised mean differences.

Sensitivity analyses included unadjusted and adjusted Cox models, multiple-imputation analyses, IPTW-adjusted models, Fine–Gray competing-risk models, and additional time-stratified analyses where model diagnostics indicated departures from proportional hazards assumptions.

3.8.8 Analytical considerations

The number and complexity of adjusted analyses depended on the size of each site-specific cohort, the number of events, and the availability and completeness of relevant covariates. Consequently, identical adjustment sets could not always be applied across sites. Analyses involving small treatment groups, limited event counts, or wide confidence intervals were interpreted as exploratory.

Potential limitations arising from non-randomised treatment allocation, residual confounding, site-specific variable availability, different treatment comparisons, missing-data, and incomplete feasibility of subgroup analyses are discussed in Section 5.6. Although pooled analyses were prespecified in the Statistical Annex, pooled estimates were not generated. Differences in exposure definitions, variable availability, cohort implementation, and analytical feasibility across participating sites limited comparability of the site-specific estimates and precluded meaningful pooling of results.

3.9 Data quality assessment

Data quality was evaluated throughout the study to assess the completeness, consistency, and comparability of site-specific datasets. The assessment was descriptive and study-specific; no formal data-quality scoring framework, QUANTUM-aligned tool, or OMOP mapping-consistency audit was applied.

Key evaluation areas included:

- completeness of mandatory variables;
- observed differences in site-specific OMOP implementation and mapped-variable availability;
- assessment of differences in variable availability despite use of a common OMOP structure;
- cross-site concordance of variable definitions;
- feasibility of federated analyses;
- availability of protocol-specified data elements.

Data quality assessment included evaluation of completeness, consistency, and cross-site comparability of protocol-specified variables.

Consistent with lessons from prior VALO1 work, cohort attrition, variable missingness, small-cell suppression, and site-specific implementation differences were considered central to assessing cross-site comparability and the interpretability of country-specific findings.

3.10 Ethical and governance considerations

The study was conducted in accordance with applicable national regulations, institutional requirements, and recognised pharmacoepidemiological principles. Only pseudonymised data were available for analysis within participating environments. While analyses were conducted by Danish researchers with SPEs access to pseudonymised patient-level data, no data were transferred between institutions/SPEs, and only aggregated results were included in study outputs.

Ethical approvals and governance clearances were obtained according to local requirements at participating sites prior to study initiation. For Norway, a conditional Regional Ethics Committee approval was first received. A patient information sheet was requested for all living patients to provide the option to opt-out of study participation, after which the study could be initiated. No patients chose the option to opt-out of study participation.

4 Results

The results section first describes cohort construction, eligibility attrition, and the final analytical study populations, followed by baseline characteristics and data availability. Subsequent results are organised according to the prespecified study objectives: treatment patterns, clinical outcomes, feasible subgroup and exploratory prognostic analyses, and infrastructure feasibility findings.

4.1 Study population and eligibility attrition

4.1.1 Analytical cohorts and treatment groups represented

The analytical cohorts and represented first-line exposure groups at Rigshospitalet, HUS, and OUS are summarised in Table 2.

Rigshospitalet contributed 1341 initially extracted patients and after all attrition steps 887 patients remained in the final analytical cohort. The final Rigshospitalet cohort included 694 (78%) patients assigned to the pembrolizumab policy-period group and 193 (21.8%) patients in the atezolizumab policy-period group. Exposure-group assignment was based on predefined treatment-policy periods and did not necessarily reflect the PD-(L)1 inhibitor administered. Cemiplimab was not represented in the current Rigshospitalet analytical cohort.

HUS contributed 156 extracted patients to the initial population represented in the reportable attrition output (non-reportable source population n=13611). After application of the eligibility criteria and key exclusions, 96 patients remained in the final analytical cohort. Five patients who initiated atezolizumab were excluded from comparative analyses because the treatment group was too small for meaningful comparison. The final HUS cohort included 91 patients, comprising 74 (81.3%) patients initiating pembrolizumab and 17 patients initiating cemiplimab as first-line monotherapy (18.7%).

At OUS, 79 remained after cohort selection. Three patients who initiated atezolizumab were excluded from comparative analyses because the treatment group was too small for meaningful comparison. The final analytical cohort included 76 patients, comprising 64 patients (84.2%) who initiated pembrolizumab and 12 patients (15.8%) who initiated cemiplimab as first-line monotherapy.

Table 2: Site-specific analytical cohorts and first-line treatment groups represented

Site	Source population	Post eligibility cohort	Treatments in final cohort	Not represented	Comparison
Rigshospitalet	1365	887	Pembro: 694 (78.2%); Atezo: 193 (21.8%)	Cemi	Pembro vs Atezo
HUS	13611	96 ¹	Pembro: 74 (81.3%); Cemi: 17 (18.7%)	Atezo not included	Pembro vs Cemi
OUS	6897	79 ²	Pembro: 64 (84.2%); Cemi: 12 (15.8%)	Atezo not included	Pembro vs Cemi

¹ Five patients who initiated atezolizumab were excluded from final cohort and comparative analyses because the treatment group was too small. ² Three patients who initiated atezolizumab were excluded from final cohort and comparative analyses because the treatment group was too small. Exposure-group definitions differed by site. Treatment distributions should not

be interpreted as complete descriptions of local prescribing practice. **Abbreviations:** Pembro = pembrolizumab; Atezo = atezolizumab; Cemi = cemiplimab.

4.1.2 Eligibility attrition

Eligibility attrition from the initially extracted populations to the final analytical cohorts is summarised in Table 3, with the detailed site-specific analytical outputs in Appendix A.

At Rigshospitalet, the initially extracted cohort included 1341 patients. Of these, 1248 met the core cohort-defining criteria reported in the current aggregate output (Table 3). At this site, mNSCLC and PD-L1 $\geq 50\%$ eligibility was reported together as a combined attrition step. After the application of additional key exclusion criteria, 887 patients remained in the final analytical cohort and formed the population used in subsequent baseline, treatment-pattern, and outcome analyses ([Appendix A](#)). The individual contribution of prior targeted therapy and prior-malignancy exclusions could not be quantified because these counts were combined in the reportable outputs as a result of local small-cell disclosure-control requirements.

At HUS, the initial output included 156 patients, of whom 139 were identified as having metastatic NSCLC. After application of the previous/concurrent malignancy exclusion, 96 patients remained. At the same time, 5 atezolizumab-treated patients were excluded from treatment-comparative analyses because the group was insufficient for comparative analysis. As a result, the final analytical cohort comprised 91 patients treated with either pembrolizumab or cemiplimab ([Appendix A](#)), and the atezolizumab group was not included in comparative analyses.

The OUS attrition output resulted in a final analytical cohort of 79 patients. Detailed intermediate attrition values are presented in the site-specific attrition diagram in Appendix A. The initial population included 6897 patients, of whom 340 had received inhibitor checkpoint immunotherapy relevant to the study. Also, 3 atezolizumab-treated patients were excluded from treatment-comparative analyses because the group was insufficient for comparative analysis. Following application of the PD-L1 $\geq 50\%$ eligibility criterion and exclusion of patients with the protocol-defined genomic or prior/concurrent malignancies, 76 patients remained in the final analytical cohort and were included in the subsequent analyses.

Full site-specific attrition tables are provided in [Appendix A](#). Intermediate attrition stages are not directly comparable across participating sites because eligibility criteria were implemented using site-specific analytical datasets, derivation procedures, and reporting frameworks. Disclosure-control requirements further affected the level of detail available in reportable outputs. Consequently, the final analytical cohorts reflect protocol eligibility criteria applied within the available site-specific analytical datasets and should not be assumed to represent all patients treated at the participating institutions.

Table 3: Country-specific analytical cohorts and first-line treatment groups represented

Attrition stage	Rigshospitalet	HUS	OUS
Initial extracted reportable population	1341	156	6897
Cohort-defining criteria met	1248	139	215
After key exclusion criteria	887	96	79
Final analytical cohort	887	91	76

Abbreviations: HUS = Helsinki University Hospital; OUS = Oslo University Hospital

4.1.3 Final analytical cohort by site and treatment

Final analytical cohort composition by site and exposure-group is summarised in Table 3. Detailed cohort characteristics and site-specific treatment group distributions are presented in [Appendix A](#).

4.2 Characteristics of the study population

Baseline characteristics were summarised descriptively for the Rigshospitalet, HUS, and OUS analytical cohorts. Detailed baseline characteristics of the final analytical cohorts by exposure-group are provided in [Appendix A](#). Variable availability differed across sites, and selected values were subject to local small-cell suppression requirements.

4.2.1 Demographic characteristics

The final Rigshospitalet cohort included 887 patients. Over half ($n=508$) (57.3%) of the patients were female. The median (Q1, Q3) age at index was 70 (64, 76) years and 277 (31.2%) patients were aged ≥ 75 years at treatment initiation. Among patients with available smoking-status information ($n=580$), 548 (94.5%) had a recorded history of smoking and 32 (5.5%) were recorded as never smokers. Smoking-status information was missing for 307 patients (34.6%).

The final HUS cohort included 91 patients. Female patients accounted for 54 (59.3%) patients and male patients for 37 (40.7%) patients. The median (Q1, Q3) age at index was 70 (62, 75) years and 26 (28.6%) patients were aged ≥ 75 years at treatment initiation. Smoking status information was not available in the analytical output.

The final OUS analytical cohort included 76 patients. The median (Q1, Q3) age at index was 71 (64, 77) years, and 27 (35.5%) patients were aged ≥ 75 years at treatment initiation. Sex counts for the overall OUS cohort were suppressed, although sex-specific counts were reportable for the pembrolizumab group. Smoking status information was missing for all 76 patients. Height and weight summaries were not reportable because the corresponding values were subject to disclosure-control requirements.

Details on demographic characteristics by first-line treatment group are presented in [Appendix A](#), Table [A1](#) to Table [A3](#).

Table 4: Demographic characteristics by site final cohorts

Characteristic	Rigshospitalet N = 887	HUS N = 91	OUS N=76
Country	Denmark	Finland	Norway
Sex			
Female, n (%)	508 (57.3%)	54 (59.3%)	Suppressed
Male, n (%)	379 (42.7%)	37 (40.7%)	Suppressed
Missing	0 (0%)	0 (0%)	Suppressed
Age, years, median (Q1, Q3)	70 (64, 76)	70 (62, 75)	71 (64, 77)
Missing, n (%)	0 (0%)	0 (0%)	0 (0%)
Age group, years			
<55	36 (4.1%)	NA	Suppressed
55–64	197 (22.2%)	NA	Suppressed
<65	NA	28 (30.8%)	Suppressed
65–74	377 (42.5%)	37 (40.7%)	Suppressed
≥75	277 (31.2%)	26 (28.6%)	27 (35.5%)
Missing	0 (0%)	0 (0%)	0 (0%)
Smoking status			
Ever	548 (94.5%)	NA	NA
Never	32 (5.5%)	NA	NA
Missing	307 (34.6%)	NA	76 (100%)
Height, cm, median (Q1, Q3)	169 (163, 176)	167 (160, 174)	Not reportable
Missing	156 (17.6%)	15 (16.5%)	Suppressed
Weight, kg, median (Q1, Q3)	70 (59, 81)	67 (59, 79)	Not reportable
Missing	54 (6.1%)	12 (13.2%)	Suppressed

This is an abbreviated version of the table. The full contents of this table are included in Appendix A. Values are presented as n (%) or median (Q1, Q3), unless otherwise specified. Small-cell suppression has been applied in accordance with local disclosure-control requirements. Suppressed values should not be interpreted as missing-data. Percentages for smoking categories are based on patients with available smoking-status information. Missing percentages are based on the total cohort denominator. **Abbreviations:** HUS = Helsinki University Hospital; NA = not available; OUS = Oslo University Hospital.

4.2.2 Clinical characteristics and prior treatment

Clinical characteristics and prior treatment history by site is presented in [Table 5](#). Details on clinical characteristics and prior treatment history by first-line treatment group are presented in [Appendix A](#), Table [A1](#) to Table [A3](#).

Among the 887 patients in the Rigshospitalet cohort, histology data was missing in 162 (18.3%) patients. Among patients with available histology information, 572 patients were classified as non-squamous NSCLC and 147 patients as squamous NSCLC. Percentages for histology categories are based on patients with available histology information. ECOG-PS 1 was the most frequently reported performance status category (n=451), followed by ECOG-PS 0 (n=297) and ECOG-PS 2 (n=65). Cardiovascular comorbidity, chronic obstructive pulmonary disease (COPD), and diabetes were reported in 355, 129, and 53 patients, respectively. Similarly, LIPI category 0 (n=408) and category 1 (n=333) were the most recorded LIPI categories. For LIPS-3, category 0 (n=539) was the most common, followed by category 1 (n=188) and category 2 (n=66). In accordance with local disclosure-control requirements, small-cell suppression was applied to selected ECOG-PS, LIPI, and LIPS-3 categories. Prior systemic anti-cancer therapy (SACT), prior thoracic procedures, and prior radiotherapy were reported in 77 (8.7%), 105 (11.8%), and 180 (20.3%) patients, respectively.

Among the 91 patients in the HUS analytical cohort, histology and ECOG-PS data were not available in the current analytical output. Cardiovascular comorbidity was reported in 45 patients, whereas counts for COPD and diabetes were subject to local disclosure-control suppression. LIPI was available, while LIPS-3 was not available. In accordance with local disclosure-control requirements, selected LIPI categories and prior treatment variables were subject to small-cell suppression. Prior radiotherapy was reported in 22 (24.2%) patients. Additional clinical characteristics and prior treatment information are presented in [Table 5](#) and [Appendix A](#).

In the OUS analytical cohort, cardiovascular comorbidity was reported in 37 patients (48.7%). COPD and diabetes counts for the overall cohort were suppressed. Histology information was missing for 41 (53.9%) patients, and reportable histology distributions were limited by suppression. ECOG-PS was extensively incomplete, with ECOG-PS missing for 45 of 64 patients (70.3%) in the pembrolizumab group and corresponding overall and cemiplimab-group values subject to suppression. LIPI, LIPS-3, disease stage, and prior radiotherapy were unavailable for all 76 patients in the analytical output. Prior SACT and prior thoracic procedure counts for the overall cohort were also subject to suppression.

Table 5. Clinical characteristics and prior treatment history by site

Characteristic	Rigshospitalet n = 887	HUS n = 91	OUS n = 76
Histology, n (%)			
Non-squamous	572 (79.4%)	NA	Suppressed
Squamous	147 (20.6%)	NA	Suppressed
Missing	162 (18.3%)	NA	41 (53.9%)
Comorbidities, n (%)			
Cardiovascular	355 (40.0%)	45 (49.5%)	37 (48.7%)
COPD	129 (14.5%)	Suppressed	Suppressed
Diabetes	53 (6.0%)	Suppressed	Suppressed
ECOG-PS, n			
0	297	NA	Suppressed
1	451	NA	Suppressed
2	65	NA	Suppressed
3	Suppressed	NA	Suppressed
4	Suppressed	NA	Suppressed
Missing	Suppressed	NA	Suppressed
LIPI, n			
0	408	41	NA
1	333	Suppressed	NA
2	Suppressed	Suppressed	NA
Missing	Suppressed	Suppressed	76
LIPS3, n			
0	539	NA	NA
1	188	NA	NA
2	66	NA	NA
3	Suppressed	NA	NA
Missing	Suppressed	NA	76
Prior SACT, n (%)	77 (8.7%)	Suppressed	Suppressed
Prior thoracic procedure, n (%)	105 (11.8%)	Suppressed	Suppressed
Prior radiotherapy, n (%)	180 (20.3%)	22 (24.2%)	Suppressed

This is an abbreviated version of the table. The full contents of this table are included in Appendix A, Table A1 to A3. Values are presented as n (%) or median (Q1, Q3), unless otherwise specified. Small-cell suppression has been applied in accordance with local disclosure-control requirements. Suppressed values should not be interpreted as missing-

data. Histology percentages are based on patients with available histology information. Missing percentages are based on the total cohort denominator. Values were reproduced from the site-specific analytical outputs.

Abbreviations: COPD=chronic obstructive pulmonary disease; ECOG = Eastern Cooperative Oncology Group performance status; HUS = Helsinki University Hospital; LIPI = Lung Immune Prognostic Index; LIPS-3 = Lung Immuno-Oncology Prognostic Score; NA = not available; OUS = Oslo University Hospital; SACT = systemic anti-cancer therapy.

4.2.3 Data availability and completeness

Availability of baseline variables differed across participating analytical datasets and is summarised in Table 4 and Table 5, with detailed variable-level information provided in [Appendix A](#).

At Rigshospitalet, the greatest amounts of missing information were observed for smoking status (307 patients; 34.6%), histology (162 patients; 18.3%), and height (156 patients; 17.6%).

At HUS, age and sex were available for all patients, consistent with OMOP CDM requirements. Smoking status, histology, and ECOG-PS were not available in the analytical output. LIPS-3 could not be calculated because ECOG-PS, a required component of the score, was unavailable.

At OUS, age and sex were available for all patients. Smoking status was unavailable for all patients, and histology and ECOG-PS were affected by substantial missingness. LIPI and LIPS-3 were unavailable in the analytical output, and prior radiotherapy data were not available. Height and weight information was available for very few patients, limiting the interpretability of anthropometric summaries.

Across all participating sites, selected variables were affected by local disclosure-control requirements that limited the level of detail that could be exported and reported. These reporting restrictions should not be interpreted as evidence that the corresponding variables were unavailable for analysis, as variable availability refers only to information available within the reportable analytical outputs and not necessarily to information present in the underlying source systems.

4.3 Overall survival (primary objective)

Overall survival was the primary study outcome. Site-specific OS results for the currently available Rigshospitalet, HUS, and OUS analytical cohorts are presented below. Detailed KM estimates, adjusted survival analyses, age-stratified analyses, and sensitivity analyses are provided in [Appendix C](#).

4.3.1 Overall survival

In the Rigshospitalet analytical cohort, assignment to the pembrolizumab policy-period group was associated with a lower hazard of death than assignment to the atezolizumab policy-period group following multivariable adjustment and multiple-imputation of missing baseline covariates (HR 0.72, 95% CI 0.58–0.90; $p=0.004$; Appendix C, [Table C6](#)). A similar finding was observed in the IPTW-adjusted sensitivity analysis (HR 0.76, 95% CI 0.60–0.95; $p=0.016$; [Appendix C, Table C7](#)). In the unadjusted analysis, median OS was 566 (95% CI 525–641) days among patients assigned to the pembrolizumab policy-period group and 401 (95% CI 335–555) days among patients assigned to the atezolizumab policy-period group ([Appendix C, Table C1](#)). The unadjusted log-rank test indicated a statistically significant difference between treatment groups ($p=0.013$). The KM curves are presented in Figure 1.

In the HUS analytical cohort, multivariable-adjusted analyses did not identify a statistically significant association between treatment group with OS (HR for pembrolizumab versus cemiplimab 1.48, 95% CI 0.60–3.67; $p=0.391$; Appendix C, [Table C6](#)). Similar findings were observed in the IPTW-adjusted sensitivity analysis (HR 1.41, 95% CI 0.56–3.55; $p=0.469$; Appendix C, [Table C7](#)). In the unadjusted analysis, median OS was not reached in either treatment group during the available follow-up period (Figure 2 and Appendix C, [Table C3](#)). Estimated survival at 500 days was 55.4% (95% CI 45.2%–68.0%) among pembrolizumab-treated patients and 57.4% (95% CI 35.1%–93.6%) among cemiplimab-treated patients. Kaplan–Meier curves are presented in Figure 2. The unadjusted log-rank test did not indicate a statistically significant difference between treatment groups ($p=0.33$).

In the OUS analytical cohort, the available multivariable Cox model indicated a lower estimated hazard of death among patients who initiated pembrolizumab than among those who initiated cemiplimab (HR 0.42, 95% CI 0.20–0.88; $p=0.022$; Appendix C, [Table C8](#)). The estimate should be interpreted cautiously because the cemiplimab comparator group comprised only 12 patients and adjustment was limited by variable availability. Multiple-imputation was not performed for the OUS analysis. In the unadjusted analysis, the HR for pembrolizumab versus cemiplimab was 0.38 (95% CI 0.18–0.78; $p=0.009$). Median OS was 192 days (95% CI 148 to not estimable) among patients who initiated cemiplimab and was not reached among patients who initiated pembrolizumab. The Kaplan–Meier curves are presented in Figure 3. The OUS estimates should be interpreted in the context of the small cemiplimab-group, which included 12 patients, the limited covariate availability, and the absence of multiple-imputation and IPTW analyses.

Figure 1 Overall survival in Rigshospitalet. Kaplan–Meier estimates of overall survival among patients initiating first-line pembrolizumab or atezolizumab in the Rigshospitalet analytical cohort. Curves show overall survival estimates with 95% confidence intervals and numbers at risk displayed below the plot.

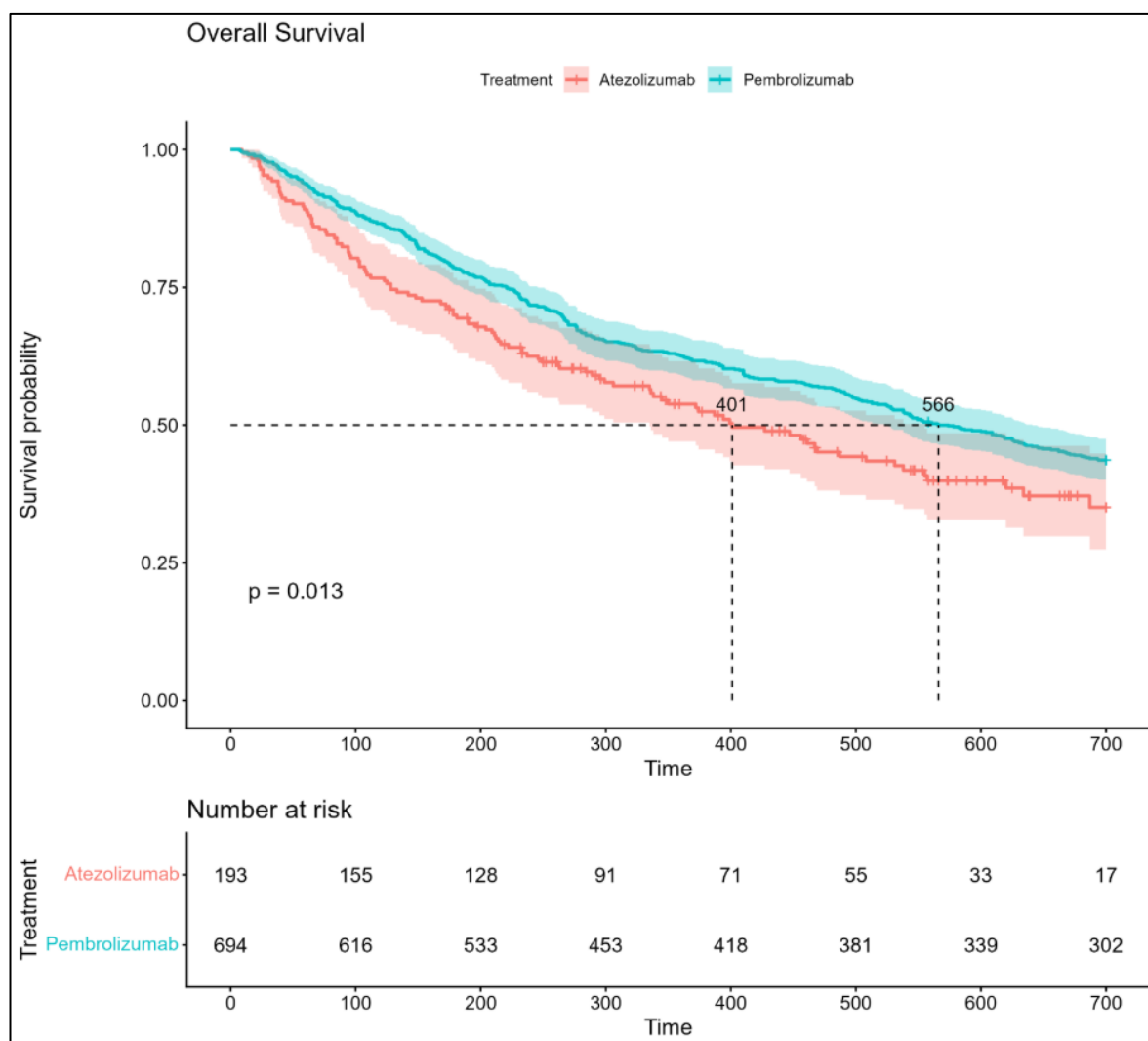


Figure 2 Overall survival in HUS. KM estimates of overall survival among patients initiating first-line pembrolizumab or cemiplimab in the HUS analytical cohort. Curves show overall survival estimates, with 95% confidence intervals and numbers at risk displayed below the plot.

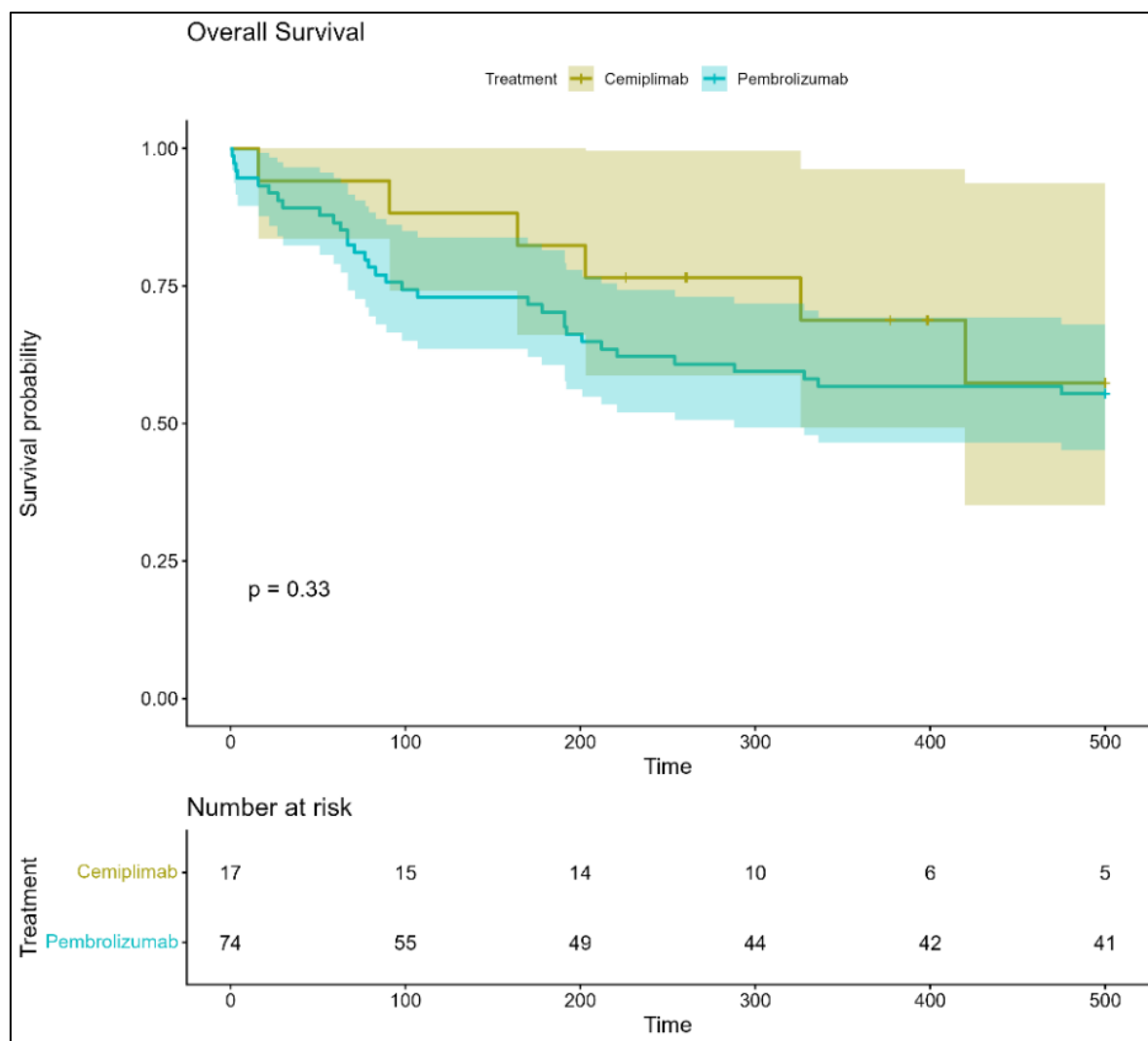
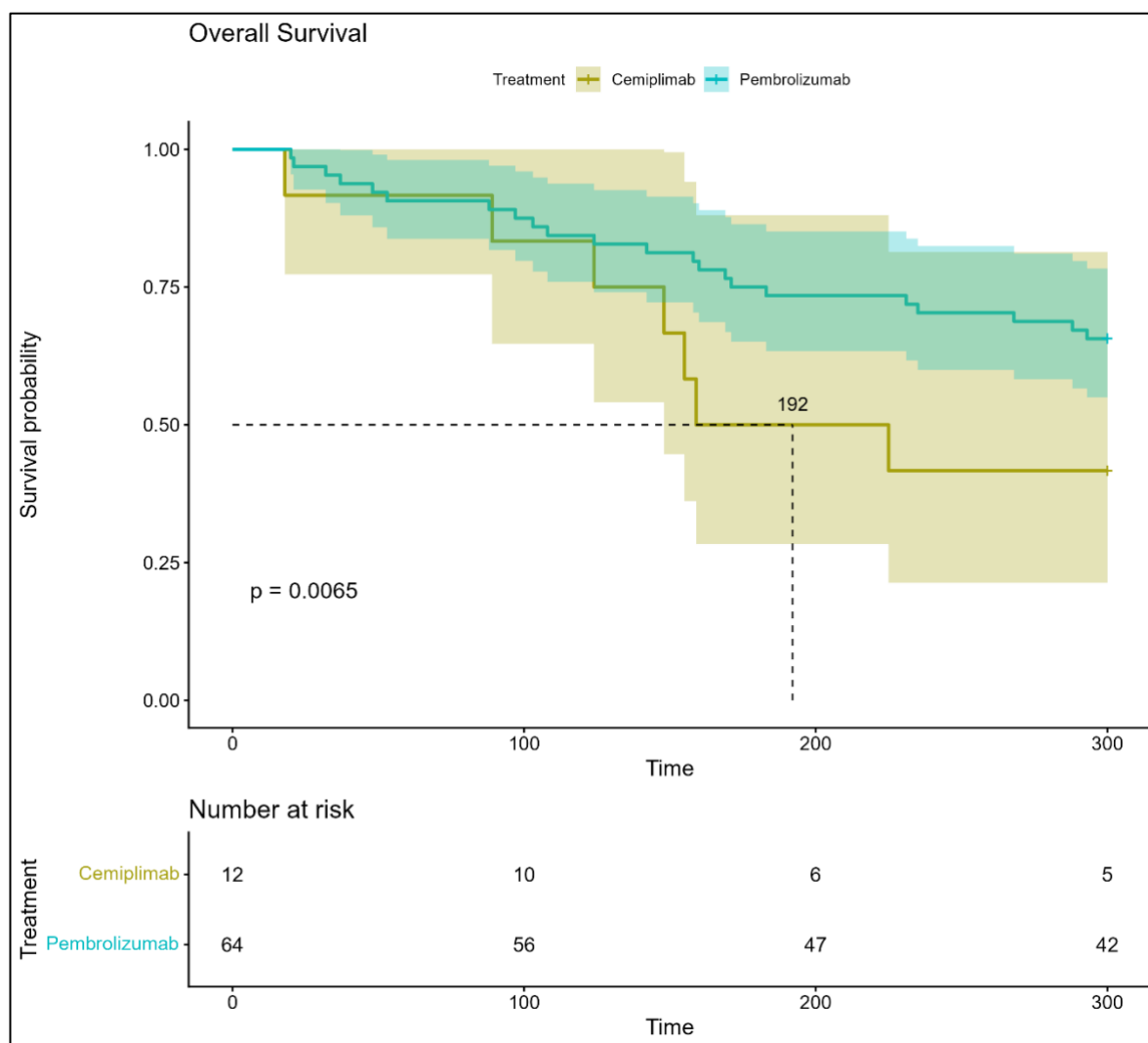


Figure 3: Overall survival in OUS. KM estimates of overall survival among patients who initiated first-line pembrolizumab or cemiplimab in the OUS cohort. Curves show overall survival estimates with 95% confidence intervals and numbers at risk displayed below the plot.



4.3.2 Age-stratified overall survival

Age-stratified overall survival analyses were performed among patients aged ≥ 75 years and < 75 years. Detailed KM plots and adjusted Cox model results are presented in Appendix C; Figures C3–C6 and [Table C5a](#).

In the Rigshospitalet cohort, assignment to the pembrolizumab policy-period group was associated with a lower estimated hazard of death than assignment to the atezolizumab policy-period group in both age strata. Among patients aged ≥ 75 years the adjusted HRs for pembrolizumab versus atezolizumab was 0.72 (95% CI 0.49–1.07; $p=0.104$) and among patients aged < 75 years was 0.74 (95% CI 0.56–0.97; $p=0.031$). Statistical significance was observed only in the < 75 -year subgroup. In the HUS cohort, adjusted HRs for pembrolizumab versus cemiplimab were 0.72 (95% CI 0.10–5.24; $p=0.698$) among patients aged ≥ 75 years and 2.22 (95% CI 0.60–8.18; $p=0.221$)

among patients aged <75 years. Confidence intervals were wide in both age strata, reflecting limited subgroup sample-size and event counts, and no statistically significant association between treatment group and OS was observed. Age-stratified OS analyses were not performed for the OUS cohort because feasibility criteria were not met.

4.3.3 Sensitivity analyses

Sensitivity analyses using unadjusted and IPTW-adjusted Cox proportional hazards models are summarised in Appendix C, [Table C7](#). In the Rigshospitalet cohort, the unadjusted Cox model comparing the pembrolizumab policy-period group with the atezolizumab policy-period group estimated a hazard ratio of 0.76 (95% CI 0.61–0.94; $p=0.013$). The corresponding IPTW-adjusted model estimated a hazard ratio of 0.76 (95% CI 0.60–0.95; $p=0.016$). These findings were consistent with the primary adjusted analysis. In the HUS cohort, the unadjusted Cox model comparing pembrolizumab with cemiplimab estimated a hazard ratio of 1.53 (95% CI 0.65–3.62; $p=0.329$). The corresponding IPTW-adjusted model estimated a hazard ratio of 1.41 (95% CI 0.56–3.55; $p=0.469$). These findings were also consistent with the primary adjusted analysis. At OUS, the unadjusted Cox model estimated an HR of 0.38 (95% CI 0.18–0.78; $p=0.009$) for pembrolizumab versus cemiplimab. No IPTW analysis was performed. The corresponding available-case multivariable result is presented with the primary OUS analysis. Findings from sensitivity analyses were generally consistent with the primary adjusted analyses.

4.4 Treatment patterns and treatment sequences (secondary objective)

Treatment-pattern results for the available Rigshospitalet, HUS, and OUS analytical cohorts are presented below. Results include first-line treatment distribution, subsequent LoT, and treatment duration estimates. Detailed site-specific treatment-pattern outputs are provided in [Appendix B](#).

4.4.1 First-line treatment distribution

First-line treatment distributions are summarised in Table 6 and detailed in [Appendix B](#) (Table [B1](#)). At Rigshospitalet, 694 (78.2%) patients were assigned to the pembrolizumab policy-period group and 193 (21.8%) patients were assigned to the atezolizumab policy-period group. Exposure-group assignment was based on predefined treatment-policy periods and did not necessarily reflect the PD-(L)1 inhibitor ultimately administered. At the same time, at HUS 74 (81.3%) patients initiated pembrolizumab and 17 (18.7%) initiated cemiplimab as first-line monotherapy ([Table 6](#)). At OUS, 64 (84.2%) patients initiated pembrolizumab and 12 (15.8%) patients initiated cemiplimab as first-line monotherapy ([Table 6](#)).

4.4.2 Subsequent lines of therapy

Subsequent treatment patterns are summarised in Table 6 and detailed in [Appendix B](#) (Tables [B2–B3](#)), including treatment distributions across later LOT. Across the analytical cohorts, platinum-based and non-platinum chemotherapy were the most frequently reported second-line treatment categories ([Table 6](#)).

At Rigshospitalet, second-line platinum-based chemotherapy was the most frequently reported subsequent treatment in both exposure groups, accounting for 186 of 694 patients (26.8%) in the pembrolizumab policy-period group and 58 of 193 patients (30.1%) in the atezolizumab policy-period group. Second-line non-platinum chemotherapy was reported for 79 patients (11.4%) and 25 patients

(13.0%), respectively Danish exposure groups were assigned according to predefined treatment-policy periods. Therefore, these summaries should not be interpreted as treatment sequences defined exclusively by the first-line PD-(L)1 inhibitor administered.

Similar patterns were observed at HUS, although the smaller cohort size resulted in fewer observed treatment transitions. Platinum-based and non-platinum-based chemotherapy were the most frequently reported second-line treatment categories among both pembrolizumab-treated and cemiplimab-treated patients ([Table 6](#)).

At OUS, reportable second-line treatment counts in the pembrolizumab group included non-platinum chemotherapy in 13 patients and platinum-based chemotherapy in 11 patients. Other second-line counts in the pembrolizumab group and all corresponding counts in the cemiplimab-group were below the disclosure threshold. Treatment distributions across subsequent lines of therapy are presented in Appendix B.

Table 6. Overview of treatment utilisation and duration across participating sites

Characteristic	Rigshospitalet (n=887)	HUS (n=91)	OUS (n=76)
First-line treatment distribution			
Pembrolizumab	694 (78.2%)	74 (81.3%)	64 (84.2%)
Atezolizumab	193 (21.8%)	Not represented	Not represented
Cemiplimab	Not represented	17 (18.7%)	12 (15.8%)
Second-line therapy after first-line pembrolizumab			
Platinum-based chemotherapy	186/694 (26.8%)	19/74 (25.7%)	11/64 (17.2%)
Non-platinum chemotherapy	79/694 (11.4%)	23/74 (31.1%)	13/64 (20.3%)
PD-L1 inhibitor	8	<5	<5
Other	8	<5	<5
Second-line therapy after first-line atezolizumab/cemiplimab			
Platinum-based chemotherapy	58/193 (30.1%)	6/17 (35.3%)	<5
Non-platinum chemotherapy	25/193 (13.0%)	6/17 (35.3%)	<5
PD-1 inhibitor	9	<5	<5
PD-L1 inhibitor	<5	<5	<5
Median duration of first-line therapy, months (IQR)			
Pembrolizumab	4.88 (1.52–13.41)	2.44 (0.70–6.84)	4.18 (1.63–17.78)
Atezolizumab	2.55 (1.00–7.37)	Not represented	Not represented
Cemiplimab	Not represented	3.60 (1.37–5.81)	2.74 (0.51–5.68)

Note: This is an abbreviated table. The full contents of this table are provided in Appendix B, Tables B2-B4. Percentages are calculated using the site-specific analytical cohort denominator shown in the column header unless otherwise specified. At Rigshospitalet, exposure groups represent predefined treatment-policy periods and do not necessarily reflect the PD-(L)1 inhibitor administered. **Abbreviations:** HUS = Helsinki University Hospital; IQR = interquartile range; OUS = Oslo University Hospital; PD-1 = programmed cell death protein 1; PD-L1 = programmed death-ligand 1.

4.4.3 Treatment Duration

Treatment duration estimates are summarised in Table 6 with detailed treatment sequencing outputs provided in [Appendix B](#), Tables [B2-B4](#).

At Rigshospitalet, the median (IQR) duration of first-line therapy was 4.88 (1.52–13.41) months in the pembrolizumab policy-period group and 2.55 (1.00–7.37) months in the atezolizumab policy-period group ([Table 6](#)). The median (IQR) durations of second-line therapy were 2.30 (1.45–3.80) months and 2.17 (0.92–3.20) months, respectively. Exposure-group assignment in Denmark was based on predefined treatment-policy periods. Accordingly, these duration estimates should be interpreted according to the assigned policy-period exposure groups rather than the specific PD-(L)1 inhibitor administered.

At HUS, the median (IQR) duration of first-line therapy was 2.44 (0.70–6.84) months among pembrolizumab-treated patients and 3.60 (1.37–5.81) months among cemiplimab-treated patients ([Table 6](#)). The median (IQR) duration of second-line therapy was 2.55 (1.68–5.26) months and 1.93 (1.54–3.40) months, respectively.

At OUS, the median (IQR) duration of first-line therapy was 4.18 (1.63–17.78) months among patients who initiated pembrolizumab and 2.74 (0.51–5.68) months among patients who initiated cemiplimab. The median (IQR) duration of second-line therapy was 2.40 (0.98–2.96) months and 2.80 (1.75–3.85) months, respectively. Interpretation of treatment duration estimates for the cemiplimab-group should be made with caution because of the small number of patients included in that cohort.

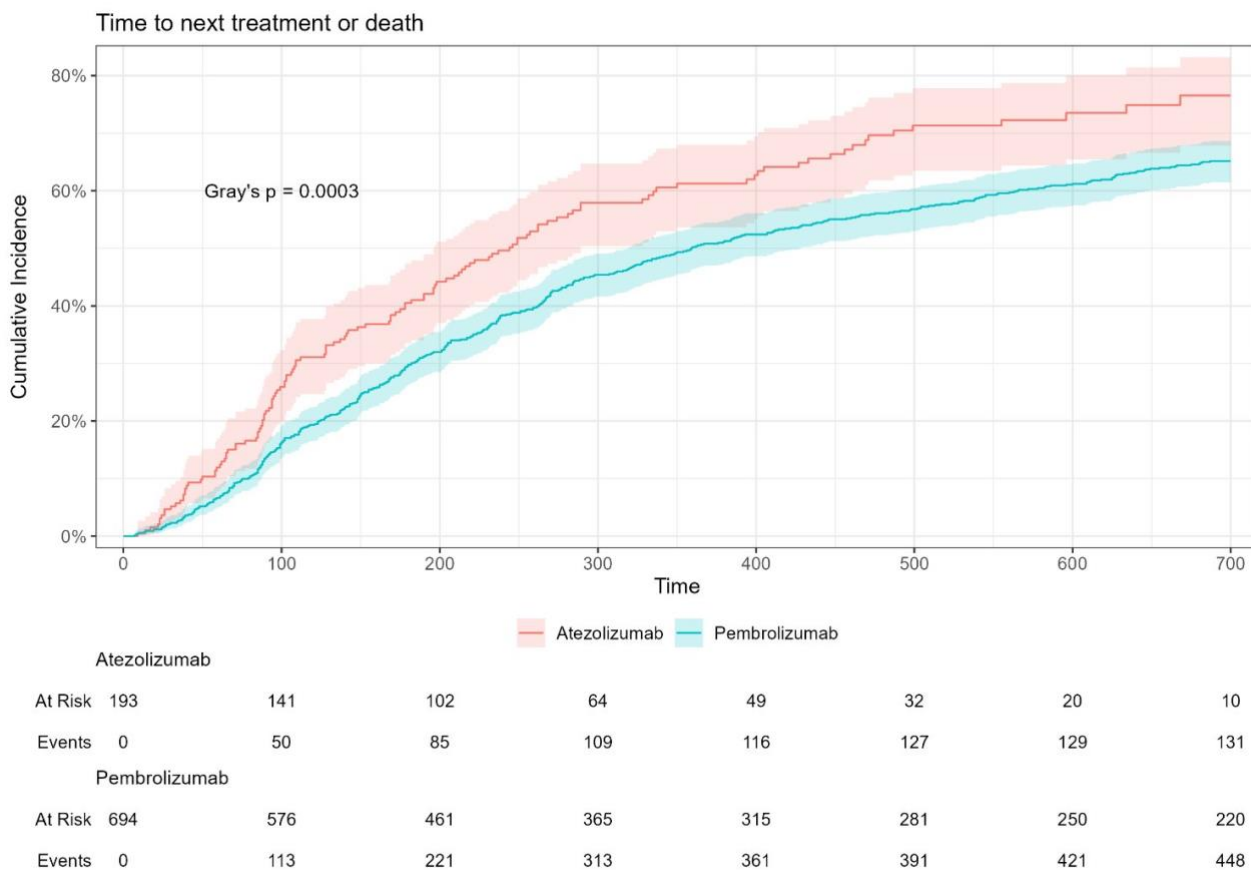
4.5 Other clinical outcomes: TTNT and TTD (secondary objective)

TTNT and TTD were evaluated as secondary composite time-to-event outcomes. Site-specific results for the Rigshospitalet, HUS, and OUS analytical cohorts are presented below. Detailed competing-risk analyses and model outputs are provided in [Appendix C](#).

4.5.1 Time to Next Treatment or Death (TTNT)

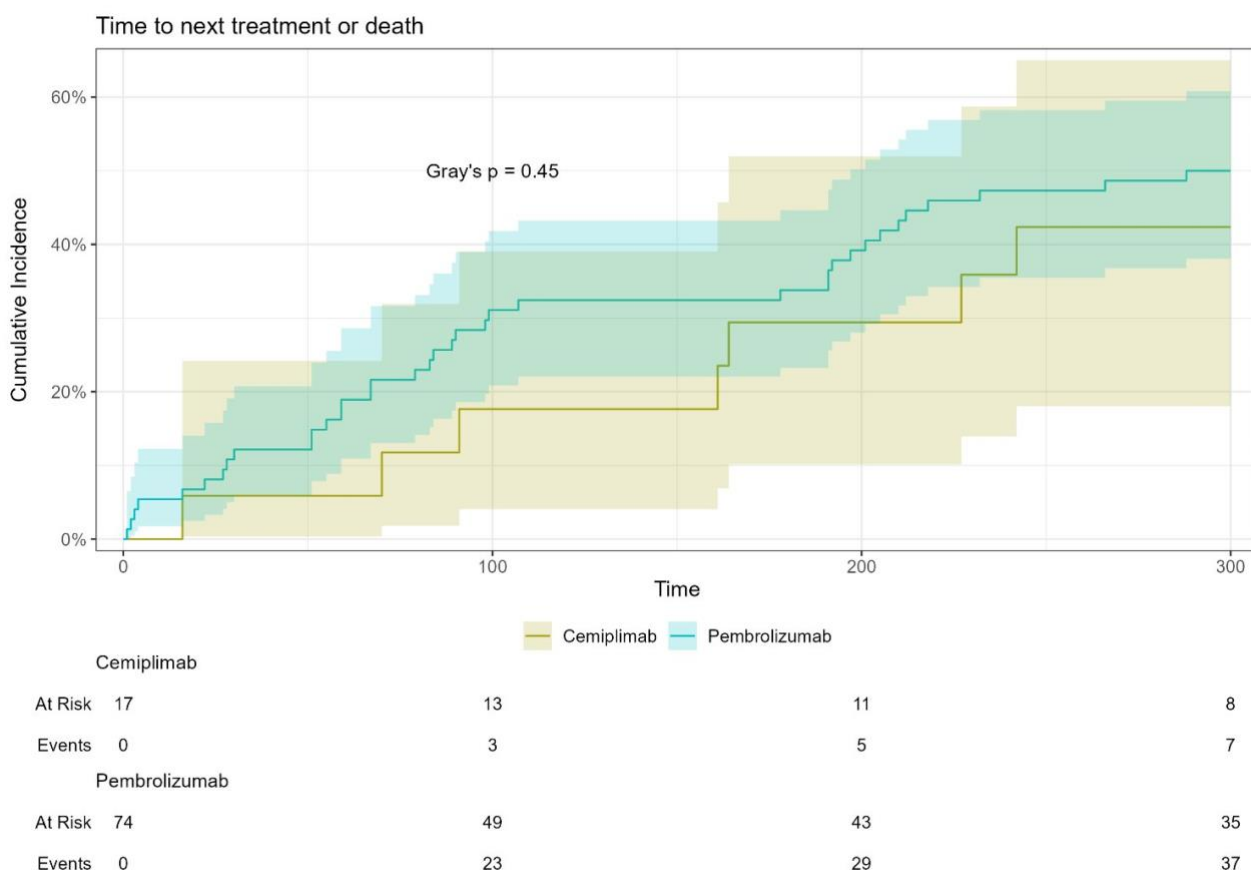
TTNT results for the available analytical cohorts are presented below. Detailed competing-risk analyses are provided in [Appendix C, Table C10](#). In the Rigshospitalet cohort, assignment to the pembrolizumab policy-period group was associated with a lower hazard of TTNT than assignment to the atezolizumab policy-period group (sHR 0.71, 95% CI 0.57–0.87; $p=0.001$; Appendix C, [Table C8](#)). The corresponding cumulative-incidence curves are shown in Figure 4.

Figure 4: Time to next line of treatment or death (TTNT) in Rigshospitalet. Cumulative-incidence curves showing time from first-line treatment initiation to initiation of subsequent anti-cancer treatment or death. Numbers at risk are displayed below the plot.



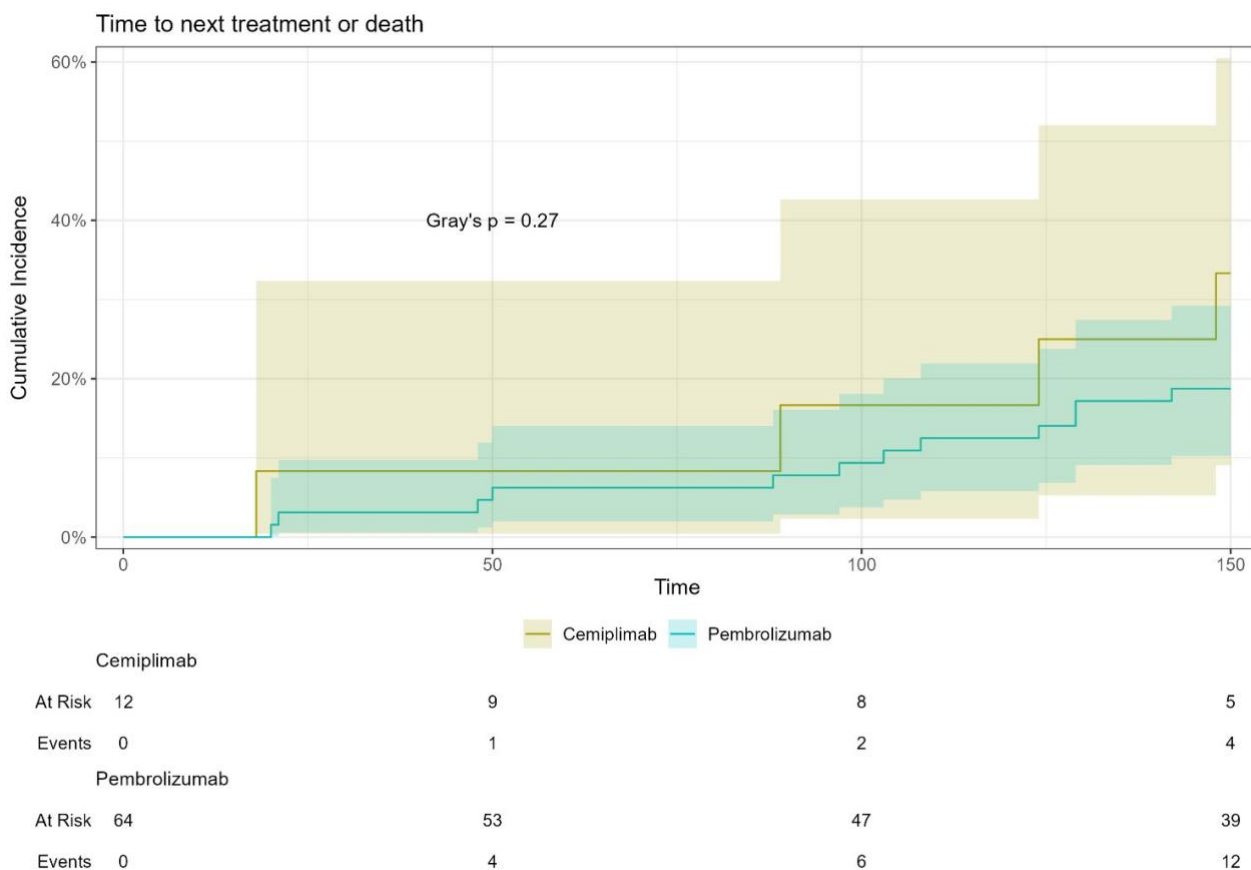
In the HUS cohort, no statistically significant difference in TTNT was observed between the treatment groups (sHR for pembrolizumab versus cemiplimab 1.45, 95% CI 0.67–3.14; $p=0.346$; Appendix C, [Table C8](#)). The corresponding cumulative-incidence curves are shown in Figure 5.

Figure 5: Time to next line of therapy or death (TTNT) in HUS. Cumulative-incidence curves showing time from initiation of first-line pembrolizumab or cemiplimab to initiation of subsequent anti-cancer treatment or death. Numbers at risk are displayed below the plot.



In the OUS cohort, no statistically significant association between treatment group and TTNT was observed in the adjusted competing-risk analysis (sHR for pembrolizumab versus cemiplimab 0.65, 95% CI 0.26–1.62; $p=0.360$). The model included treatment group, age, sex, prior SACT, and prior thoracic procedure. No multiple-imputation analysis was performed because of limited covariate availability. The corresponding cumulative-incidence curves are presented in Figure 6.

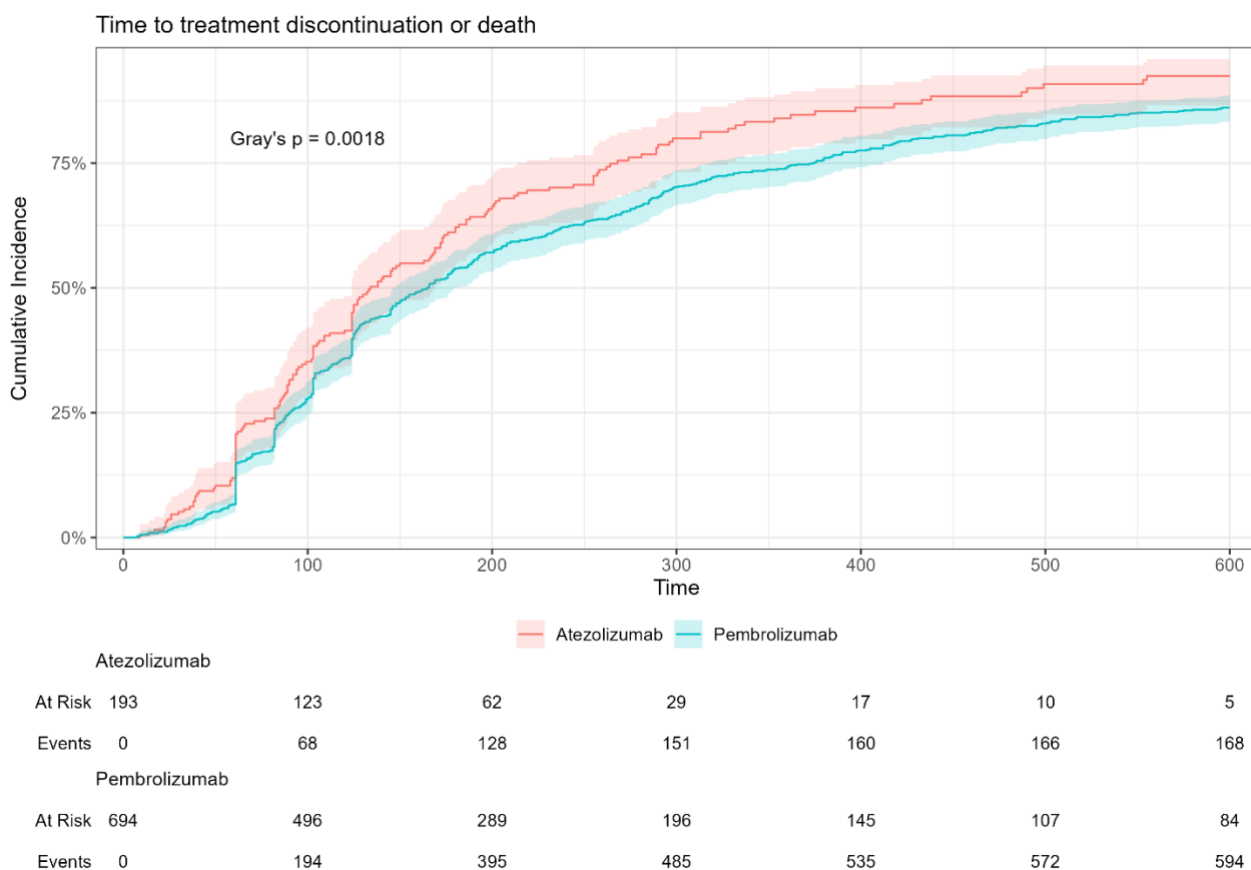
Figure 6. Time to next line of treatment or death (TTNT) in OUS. Cumulative-incidence curves showing time from initiation of first-line pembrolizumab or cemiplimab to initiation of subsequent anti-cancer treatment or death. Numbers at risk are displayed below the plot.



4.5.2 Time to Treatment Discontinuation or Death (TTD)

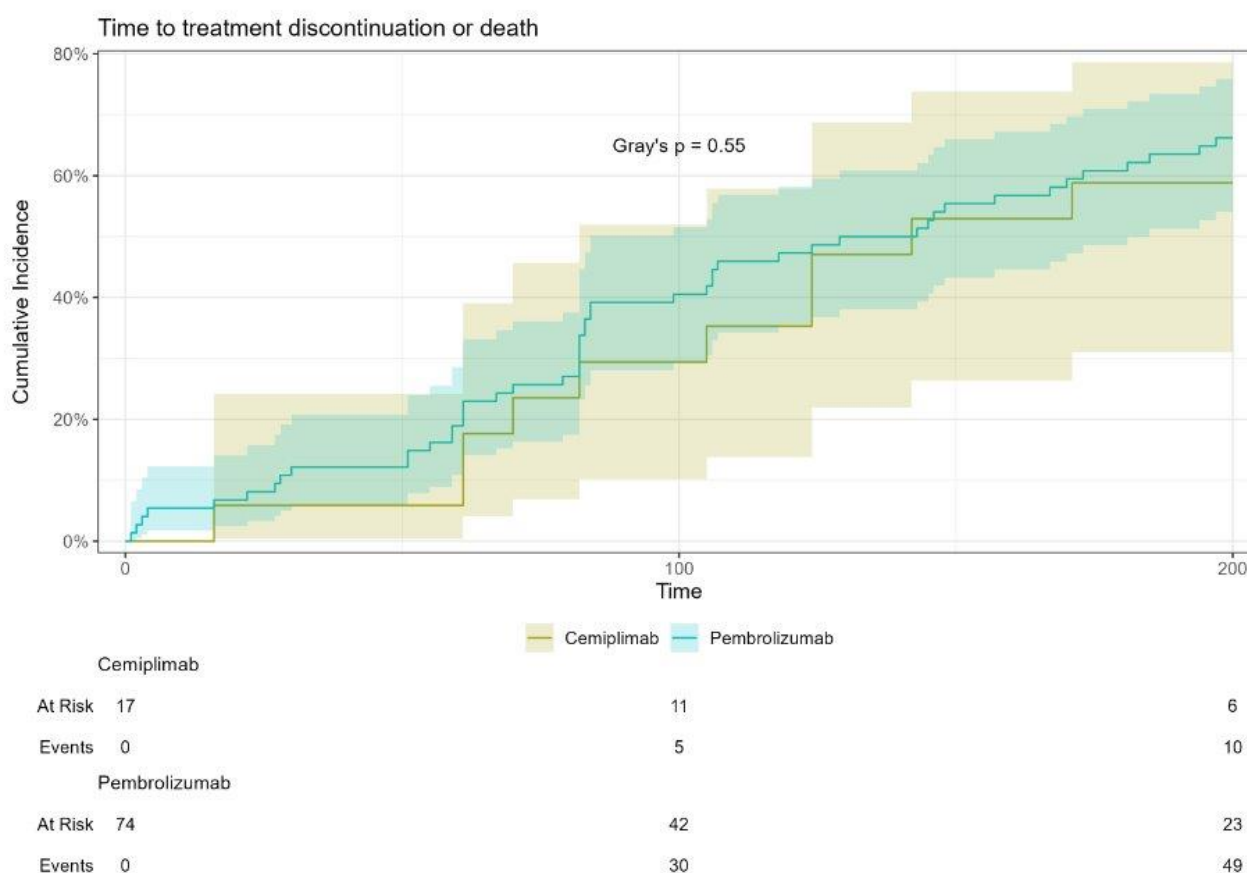
TTD results for the site analytical cohorts are presented below. Detailed competing-risk analyses are provided in Appendix C, [Table C10](#). In the Rigshospitalet cohort, assignment to the pembrolizumab policy-period group was associated with a lower adjusted hazard of TTD than assignment to the atezolizumab policy-period group (sHR 0.79, 95% CI 0.65–0.95; p=0.012; Appendix C, [Table C10](#)).

Figure 7: Time to treatment discontinuation or death (TTD) in Rigshospitalet. Cumulative-incidence curves showing time from first-line treatment initiation to treatment discontinuation or death. Numbers at risk are displayed below the plot.



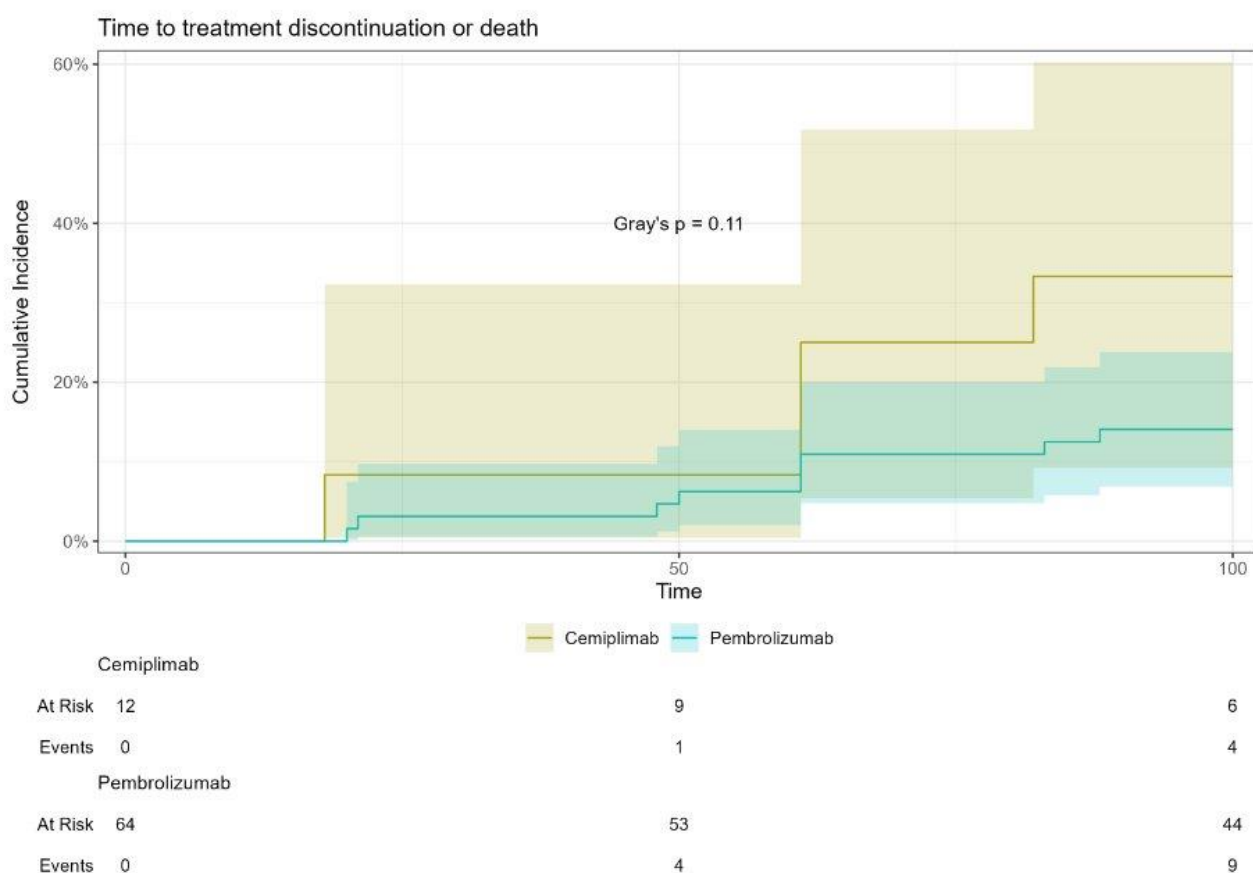
In the HUS cohort, the cumulative-incidence of TTD at 200 days was 66.2% among pembrolizumab-treated patients and 58.8% among cemiplimab-treated patients. Also, no statistically significant association between treatment group and TTD was observed in the adjusted competing-risk analysis (sHR for pembrolizumab versus cemiplimab 1.06, 95% CI 0.63-1.79; $p=0.817$; Appendix C, [Table C10](#)). The corresponding cumulative-incidence curves are presented in Figure 8.

Figure 8: Time to treatment discontinuation or death (TTD) in HUS. Cumulative-incidence curves showing time from first-line treatment initiation to treatment discontinuation or death. Numbers at risk are displayed below the plot.



In the OUS cohort, the adjusted Fine–Gray competing-risk analysis did not identify a statistically significant association between treatment group and TTD (sHR for pembrolizumab versus cemiplimab 0.97, 95% CI 0.41–2.31; $p=0.950$; Appendix C, [Table C10](#)). The model included treatment group, age, sex, prior SACT, and prior thoracic procedure. No multiple-imputation analysis was performed because of limited covariate availability. The corresponding cumulative-incidence curves are presented in Figure 9.

Figure 9. Time to Treatment Discontinuation or Death (TTD) in OUS. Cumulative-incidence curves showing time from initiation of first-line pembrolizumab or cemiplimab to treatment discontinuation or death. Numbers at risk are displayed below the plot.



4.6 Exploratory Prognostic Analyses (Exploratory Objective)

Exploratory prognostic analyses evaluated LIPI and, where available, LIPS-3 as potential prognostic variables. Detailed model outputs are provided in [Appendix C, Table C11](#).

In the Rigshospitalet cohort, compared with LIPI category 0, LIPI categories 1 and 2 were associated with increased hazards of death (HR 1.44, 95% CI 1.19–1.73 and HR 1.88, 95% CI 1.40–2.51, respectively; [Appendix C, Table C11](#)). LIPI category 1 and category 2 were also associated with higher subdistribution hazards for TTD (sHR 1.24, 95% CI 1.06–1.46 and sHR 1.65, 95% CI 1.23–2.22, respectively; [Appendix C, Table C11](#)). In the HUS cohort, compared with LIPI category 0, LIPI category 2 was associated with an increased hazard of TTD (sHR 2.50, 95% CI 1.12–5.56; [Appendix C, Table C11](#)). No statistically significant associations were observed for the other evaluated outcomes.

LIPS-3 was available in the Rigshospitalet cohort and was included in exploratory adjusted analyses. However, no consistent independent association with OS, TTNT, or TTD was observed ([Appendix C, Table C11](#)). LIPS-3 could not be evaluated in the HUS cohort because ECOG performance status, a required component of the score, was unavailable in the analytical output. Similarly, exploratory

analyses of LIPI and LIPS-3 were not performed in the OUS cohort because the information required to derive these prognostic scores was unavailable (Appendix C, Table C11).

4.7 Infrastructure feasibility findings (infrastructure objective)

The cross-border access for Danish researchers was established at participating sites in Rigshospitalet, HUS, and OUS. Harmonised analytical scripts were executed locally within the participating SPEs via approved remote access, and site-specific analytical outputs were generated from those analyses. Key infrastructure feasibility measures and variable availability in the OMOP-formatted analytical datasets are summarised in [Table 7](#).

4.7.1 Infrastructure implementation

Analyses were conducted by authorised Danish researchers within local SPEs, while patient-level data remained at the originating institutions throughout the study workflow [11,12]. Information provided by participating sites indicated variation in approval, data-access, and implementation timelines. Existing approvals from previous project enabled continued data-access at Rigshospitalet. At HUS, initial research approval was received on 26 January 2026, and data became available within HUS Acamedic on 31 March 2026. At OUS, Regional Ethics Committee (REC) conditional approval was received on May 2026 with the condition to send out patient information sheets to all living patients with the option to opt-out from the study. After Patient Information Sheets were sent to the identified study cohort, a 4-week opt-out period started after which the study could be initiated. No patients exercised the option to opt out of study participation. Data became available within the Trusted Secure Data (TSD) environment on 10 August 2026 ([Table 7](#)).

4.7.2 Variable availability in OMOP formatted data

Variable availability in the OMOP-formatted analytical datasets differed across sites (Table 7). In the Rigshospitalet analytical dataset, demographic variables, treatment information and outcome variables were available. ECOG-PS was derived from chart-note information using automated rule-based extraction methods. LIPI and LIPS-3 were calculated from the required component variables and were not directly available as pre-existing OMOP variables. Limitations were identified for genomic biomarkers and selected variables with missing data.

At HUS, demographic and outcome variables were available in the analytical dataset. PD-L1 values above the protocol eligibility threshold were not available as discrete reportable variables and LIPS-3 could not be calculated because ECOG-PS, a required component of the score was not available. Selected variables were also affected by local disclosure-control requirements.

At OUS, age, sex, treatment and outcome variables were available. Smoking status, LIPI, LIPS-3 disease stage information, and prior radiotherapy were unavailable in the analytical output. Histology and ECOG-PS were affected by substantial missingness and the level of reportable detail for a few variables was limited by local disclosure-control requirements.

Differences in variable availability reflected the content of the site-specific OMOP-formatted analytical datasets available for the study. Genomic biomarker analyses were limited because the available datasets recorded positive genomic findings without consistently providing a denominator of tested patients. Consequently, protocol-specified biomarkers were used for cohort inclusion and exclusion where applicable but were not included in baseline or adjusted analytical models. PD-L1

status was required for cohort eligibility; however, exact values above the eligibility threshold were not consistently available across sites and PD-L1 was therefore not included as an adjustment variable.

4.7.3 Cohort implementation considerations

Implementation of protocol-defined eligibility criteria differed across the available OMOP-formatted datasets.

At Rigshospitalet, selected cohort-defining criteria and exclusion criteria could not be reported separately in exported outputs because of local disclosure-control requirements. These reporting restrictions applied to the exported aggregate outputs and should not be interpreted as limitations of the underlying analytical implementation. At HUS, metastatic status was derived using treatment and surgical-history information, and PD-L1 eligibility was inferred from local first-line ICI monotherapy prescribing practice because PD-L1 results were not available as discrete records in the analytical output. Across participating sites, local disclosure-control requirements influenced the level of detail that could be exported and reported. Consequently, selected variables and intermediate attrition counts were suppressed or aggregated in the report outputs. These reporting restrictions did not necessarily indicate that the corresponding variables were unavailable for analysis.

Table 7. Infrastructure feasibility and variable availability in OMOP-formatted datasets

Measure	Rigshospitalet	HUS	OUS
Data approval	Existing approval from previous project enabled immediate data-access	Research approval received 26 January 2026	REC conditional approval received 8 May 2026
Data-access	Detailed timing metrics not available.	Data available in HUS Academic on 31 March 2026.	Data available in TSD on 10 August 2026
Variables requested	Protocol-specified set implemented.	Protocol-specified set implemented.	Protocol-specified set implemented
OMOP-formatted variables available	Demographic, treatment, and outcome variables available. ECOG-PS was derived from chart-note information using automated rule-based extraction methods. LIPI and LIPS-3 were calculated from component variables and were not directly available as OMOP variables. Limitations were identified for genomic biomarker information and selected variables affected by missing-data.	Demographics and outcomes available; PD-L1 not recorded as a discrete field; LIPS-3 could not be calculated because ECOG-PS was unavailable; selected variables suppressed.	Age, sex, treatment, and outcome variables available. Smoking status, LIPI, LIPS-3, disease stage, and prior radiotherapy variables were unavailable in the analytical output. Histology and ECOG-PS were affected by substantial missingness. Selected variables were subject to local disclosure-control restrictions.

Note: Variable availability refers to variables available in the OMOP-formatted analytical datasets used for the analyses and does not imply availability in the underlying electronic medical records.

5 Discussion

5.1 Main findings

This multinational Nordic study evaluated outcomes associated with first-line PD-(L)1 inhibitor monotherapy among patients with PD-L1 $\geq 50\%$ mNSCLC using harmonised analyses across Denmark, Finland, and Norway. The study also assessed the feasibility of conducting federated RWE analyses using SPEs and OMOP-based analytical datasets.

Clinical findings differed across participating site-specific analytical cohorts. In the Rigshospitalet cohort, assignment to the pembrolizumab policy-period group was associated with longer OS and lower subdistribution hazards of TTNT and TTD compared with the atezolizumab policy-period group. In contrast, no statistically significant differences in OS, TTNT, or TTD were observed between pembrolizumab and cemiplimab in the HUS cohort. In the OUS cohort, pembrolizumab was associated with longer OS than cemiplimab, although this finding was based on a small cemiplimab comparator group and limited covariate adjustment.

Treatment patterns were broadly similar across participating sites, with platinum-based and non-platinum chemotherapy representing the most frequently reported subsequent treatment categories following first-line PD-(L)1 inhibitor monotherapy. Exploratory analyses suggested prognostic value of LIPI where sufficient data were available, whereas evaluation of LIPS-3 was limited by data availability.

From an infrastructure perspective, VALO2 demonstrated that a common analytical framework could be successfully implemented across multiple participating Nordic healthcare institutions and that Danish researchers could analyse data remotely within local SPEs. Federated analysis was operationally feasible; however, use of OMOP Common Data Model (CDM) mapping and reusable analytical code did not ensure comparable cohorts, variables, or adjustment opportunities across participating sites. Differences in variable availability refer to the OMOP-mapped analytical datasets used in the study and do not necessarily imply absence of the corresponding information in the underlying EHR systems.

5.2 Primary objective: overall survival

Overall survival was the primary outcome of the VALO2 study and represented the principal endpoint to evaluate outcomes among patients receiving first-line ICI monotherapy for PD-L1 $\geq 50\%$ metastatic NSCLC.

The site-specific OS findings differed across the participating analytical cohorts. At Rigshospitalet, assignment to the pembrolizumab policy-period group was associated with a lower hazard of death than assignment to the atezolizumab policy-period group after multivariable adjustment (HR 0.72, 95% CI 0.58–0.90; $p=0.004$). In contrast, no statistically significant differences in OS were observed between pembrolizumab and cemiplimab in the HUS cohort (HR 1.48, 95% CI 0.60–3.67; $p=0.391$). At OUS, pembrolizumab was associated with a lower hazard of death than cemiplimab (HR 0.42, 95% CI 0.20–0.88; $p=0.022$), although this estimate was based on a cemiplimab comparator group comprising only 12 patients and a more limited set of adjustment variables.

Several factors should be considered when interpreting these findings. First, VALO2 was designed as an observational study rather than a randomised comparative trial. Treatment allocation may therefore have been influenced by patient characteristics, calendar-time treatment

uptake, reimbursement and procurement context, prescribing practices, and other factors that could not be fully captured in the available data. Although adjusted and sensitivity analyses produced findings that were broadly consistent with the primary analyses, residual confounding remains possible. Furthermore, the Danish comparison was based on policy-period exposure-group assignment rather than the PD-(L)1 inhibitor ultimately administered, whereas the Finnish and Norwegian analyses were based on the PD-(L)1 inhibitor initiated in routine clinical practice.

Interpretation of the HUS and OUS findings is constrained by the relatively small cemiplimab cohorts ($n=17$ and $n=12$, respectively), limited event counts, and reduced statistical precision. Consequently, the absence of statistically significant differences in the HUS should not be interpreted as evidence of treatment equivalence. Similarly, although a statistically significant association with OS was observed in the OUS cohort, the limited sample-size and restricted covariate availability increase uncertainty around the estimated effect size. These findings should therefore be considered exploratory and interpreted in the context of the broader evidence base.

The OS findings should also be considered in the context of the existing evidence base for immune checkpoint inhibitor monotherapy in patients with PD-L1 $\geq 50\%$ metastatic NSCLC. Pembrolizumab, atezolizumab, and cemiplimab have each demonstrated clinically meaningful survival benefits compared with chemotherapy in pivotal registration studies and subsequent long-term follow-up analyses [2,7,8,14,18]. The associations observed in the Rigshospitalet cohort are broadly consistent with the established evidence supporting first-line PD-(L)1 inhibitor monotherapy in PD-L1-high metastatic NSCLC. Likewise, the absence of statistically significant differences in the HUS cohort and the favourable outcomes observed in the OUS pembrolizumab cohort do not conflict with the available evidence supporting both pembrolizumab and cemiplimab as effective first-line treatment options for patients with PD-L1 $\geq 50\%$ metastatic NSCLC.

5.3 Secondary objectives: treatment patterns, TTNT, and TTD

The secondary analyses evaluated treatment patterns, treatment duration, TTNT, and TTD. Overall, the secondary endpoint findings were broadly aligned with the corresponding OS results and did not identify findings that materially altered the interpretation of the primary outcome analyses.

5.3.1 Treatment patterns

Treatment patterns were generally consistent across participating cohorts. Platinum-based and non-platinum chemotherapy represented the most frequently reported subsequent treatment categories following first-line PD-(L)1 inhibitor monotherapy. Treatment duration estimates varied between countries and exposure groups, with median first-line treatment duration ranging from 2.4 to 4.9 months across the evaluated cohorts. Although these findings are broadly consistent with contemporary management of metastatic NSCLC following progression on first-line immunotherapy, interpretation should consider differences in cohort size, exposure-group definition, and local disclosure-control requirements.

5.3.2 TTNT and TTD

TTNT and TTD findings were generally aligned with the OS results. In the Rigshospitalet cohort, assignment to the pembrolizumab policy-period group was associated with lower subdistribution hazards of both TTNT: sHR 0.71 (95% CI 0.57–0.87) and TTD: sHR 0.79 (95% CI 0.65–0.95) than

assignment to the atezolizumab policy-period group. In contrast, no statistically significant differences were observed in the HUS cohort for either TTNT (sHR 1.45, 95% CI 0.67–3.14) or TTD (sHR 1.06, 95% CI 0.63–1.79). Similarly, in the OUS cohort, no statistically significant differences were observed for TTNT (sHR 0.65, 95% CI 0.26–1.62) or TTD (sHR 0.97, 95% CI 0.41–2.31).

Interpretation of TTNT and TTD requires consideration of the composite nature of these endpoints. Both outcomes incorporated death as an event component and therefore reflect a combination of treatment trajectory and survival experience. Consequently, TTNT and TTD should be interpreted alongside OS rather than as independent measures of treatment effectiveness. The divergence observed in the OUS cohort, where an association with OS was observed but not with TTNT or TTD, illustrates the importance of considering these endpoints together. It is possible that mortality and treatment-transition patterns contributed differently to the composite endpoints, particularly in the context of small cohort sizes and limited event counts. Comparison with published evidence

The VALO2 findings should be interpreted within the context of the established evidence base supporting first-line PD-(L)1 inhibitor monotherapy for patients with PD-L1 $\geq 50\%$ metastatic NSCLC. Pembrolizumab, atezolizumab, and cemiplimab have each demonstrated clinically meaningful survival benefits compared with chemotherapy in pivotal registration trials and subsequent long-term follow-up analyses [2,7,8,14,18]. Consequently, all 3 agents are recognised treatment options for this patient population in contemporary clinical practice [4,10].

Importantly, direct comparative evidence between approved first-line PD-(L)1 inhibitor monotherapies remains limited. The pivotal studies supporting pembrolizumab, atezolizumab, and cemiplimab compared these agents against platinum-based chemotherapy rather than against each other [7,14,18]. As a result, treatment comparisons are often evaluated using observational real-world data and should be interpreted in the context of differences in patient populations, treatment selection, data availability, and analytical methodology.

Together, these studies form the evidence base supporting the use of pembrolizumab, atezolizumab, and cemiplimab as first-line treatment options in PD-L1-high metastatic NSCLC [4,7,8,10,14,18]. VALO2 findings complement this evidence by providing real-world observations generated within routine Nordic clinical practice and by evaluating outcomes using a federated analytical framework across multiple healthcare systems. The favourable outcomes observed in the Rigshospitalet cohort are consistent with the long-term survival benefit demonstrated for pembrolizumab in KEYNOTE-001 and KEYNOTE-024 [2,18]. Similarly, the absence of statistically significant differences in the HUS cohort does not contradict the evidence supporting both pembrolizumab and cemiplimab as effective first-line treatment options in patients with PD-L1 $\geq 50\%$ metastatic NSCLC [8,14,18]. The OUS findings should be interpreted cautiously because of the limited number of cemiplimab-treated patients; however, they remain broadly compatible with the overall evidence supporting the effectiveness of first-line immunotherapy in this population [7,8,14,18].

The Rigshospitalet findings are broadly consistent with the established effectiveness of pembrolizumab observed in both clinical trials and real-world studies [2,4,10]. In the available Danish cohort, patients assigned to the pembrolizumab policy-period group experienced more favourable observed outcomes across overall survival, TTNT, and TTD than patients assigned to the atezolizumab policy-period group. Because exposure-group assignment was based on predefined treatment-policy periods rather than necessarily on the PD-(L)1 inhibitor administered, these findings should be interpreted as associations between exposure groups rather than direct treatment-effect comparisons.

The differing findings across participating countries highlight the challenges of interpreting comparative-effectiveness results from relatively small observational cohorts, particularly when exposure definitions, cohort composition, variable availability, and adjustment strategies differ between sites. It is also important to recognise that the currently available VALO2 analyses do not represent a single treatment comparison. The Danish analysis compared policy-period exposure groups, whereas the Finnish and Norwegian analyses compared PD-(L)1 inhibitors initiated in routine clinical practice. Furthermore, pivotal clinical trials compared each immune checkpoint inhibitor with chemotherapy rather than with one another [7,8,14,18]. Direct comparison across these treatments is therefore inherently limited by differences in study populations, eligibility criteria, treatment context, and follow-up.

The findings observed in Denmark, Finland, and Norway are therefore more appropriately viewed as complementary RWE rather than confirmatory head-to-head comparisons between individual PD-(L)1 inhibitors.

5.4 Exploratory objective: prognostic analyses

Exploratory analyses evaluated the prognostic value of LIPI and, where available, LIPS-3 in patients receiving first-line ICI monotherapy for PD-L1 $\geq 50\%$ metastatic NSCLC. The findings suggest that LIPI may provide additional prognostic information beyond treatment assignment, with poorer LIPI categories generally associated with less favourable outcomes. For example, compared with LIPI category 0, LIPI category 2 was associated with a higher hazard of death (HR 1.88, 95% CI 1.40–2.51). These findings are consistent with previous studies reporting an association between systemic inflammatory status and outcomes among patients receiving immune checkpoint inhibitors [19].

In the Rigshospitalet cohort, higher LIPI categories were associated with increased risks of death and treatment discontinuation or death across multiple adjusted models. However, these associations were not observed consistently across all endpoints and cohorts, highlighting the exploratory nature of these analyses.

LIPS-3 could be derived from data in the Rigshospitalet OMOP-formatted analytical dataset but was not evaluated in HUS as the required availability of ECOG-PS was not available in the analytical dataset, limiting evaluation of its prognostic utility across participating sites. In the available Rigshospitalet analyses, LIPS-3 did not demonstrate a consistent independent association with OS, TTNT, or TTD after adjustment [20]. In the OUS cohort, neither LIPI nor LIPS-3 could be evaluated because the variables required to derive these scores were unavailable in the analytical output. Consequently, conclusions regarding the prognostic utility of LIPI and LIPS-3 across the full Nordic study population remain limited.

Overall, these analyses should be regarded as hypothesis-generating. Differences in variable availability, cohort size, event counts, and the restricted availability of LIPI and LIPS-3 across sites limit firm conclusions and highlight the need for validation in larger datasets and future studies [19,20].

5.5 Infrastructure lessons learned

The federated infrastructure component of VALO2 functioned as intended. Participating sites had successfully mapped local data to the OMOP Common Data Model and Danish researchers developed scripts executed harmonised analytical code within local SPEs and generated

aggregated study outputs without transfer of patient-level data. These findings demonstrate that federated RWE generation is operationally feasible across Nordic healthcare institutions.

The principal challenges encountered during VALO2 related to data availability and variable capture rather than analytical execution. Although a common protocol and analytical framework were applied across sites, differences in site-specific OMOP implementations, available mapped datasets, and local documentation practices influenced the availability and operationalisation of important study variables. Differences in variable availability should be interpreted as differences in the OMOP-formatted analytical datasets used for the study and do not necessarily imply absence of the corresponding variables in the underlying electronic medical records.

The study further demonstrated that harmonised analytical code does not automatically produce harmonised cohorts. Site-specific differences in variable availability influenced implementation of eligibility criteria, cohort construction, and attrition reporting. These findings highlight the importance of transparent documentation of cohort derivation and variable availability in federated studies.

Compared with the previous VALO1 pilot, VALO2 enabled Danish researchers to access OMOP-formatted analytical data through participating SPEs and to execute analyses directly within the local environments. This extended the federated operating model beyond site-run script execution and demonstrated the feasibility of cross-border analytical access while maintaining local data custody.

Future Nordic federated oncology studies, particularly mNSCLC would benefit from more standardised capture of biomarker testing, PD-L1 results, ECOG performance status, metastatic disease status, treatment-line information, and cohort attrition metrics. Improvements in these areas would strengthen cross-site comparability and support more robust interpretation of future federated RWE studies. All findings and recommendations related to the federated infrastructure, OMOP implementation, and operational lessons learned are described in the accompanying VALO2 Lessons Learned Report.

5.6 Limitations

5.6.1 Cohort construction and attrition

Intermediate attrition steps should be interpreted considering the site-specific cohort implementation considerations described in the [Methods](#). Eligibility criteria were implemented using locally available OMOP-formatted analytical datasets and, in some cases, required site-specific derivation approaches. Consequently, the actual analytical cohorts were smaller than anticipated during protocol development, and intermediate attrition stages are not directly comparable across sites. In addition, protocol-level cohort estimates were not always directly comparable to the final analytical populations available for analysis. Differences in cohort size across participating sites reflected differences in source populations, site-specific analytical datasets, eligibility implementation, and availability of reportable variables rather than a common sampling framework. For example, the HUS analytical dataset represented a subset of the Finnish population rather than national coverage, and approximately 750 ICI-treated patients were identified before application of study eligibility criteria.

The participating cohorts should also be considered in the context of representativeness. The VALO2 analyses were conducted using institutional and regional datasets rather than national populations. Consequently, the final analytical cohorts may not fully represent all patients receiving first-line PD-

(L)1 inhibitor monotherapy within Denmark, Finland, or Norway. In addition, patients may have been excluded because required eligibility information was unavailable in the analytical datasets, potentially further affecting representativeness.

5.6.2 Cross-country interpretability

The participating sites contributed different treatment comparisons, exposure definitions, analytical cohorts, and adjustment strategies.

In Denmark, exposure groups were assigned according to predefined treatment-policy periods, whereas the Finnish and Norwegian analyses were based on the PD-(L)1 inhibitor initiated in clinical practice. The Danish policy-period approach was selected to reduce potential selection associated with clinician choice during periods when treatment recommendations were changing and uptake occurred gradually over time.

Cross-site comparability was further influenced by differences in cohort size, OMOP-formatted variable availability, eligibility implementation, and healthcare-system context. Consequently, findings should be interpreted as country-specific observations rather than directly comparable treatment-effect estimates.

5.6.3 Observational design and residual confounding

VALO2 was designed as an observational real-world study and therefore remains subject to the limitations inherent to non-randomised treatment comparisons. Treatment allocation was determined in routine clinical practice and may have been influenced by patient characteristics and other clinical factors that could not be fully captured in the available datasets. Consequently, residual confounding cannot be excluded despite the use of adjusted analyses, multiple imputations, and sensitivity analyses.

Interpretation of the findings is further limited by the incomplete availability of important prognostic variables across participating sites. ECOG performance status was available in the Rigshospitalet analytical dataset but was not consistently available across all currently available cohorts. Because ECOG is an established prognostic factor in mNSCLC [4,10], residual confounding related to differences in baseline performance status cannot be excluded.

Comorbidity burden may also have contributed to residual confounding. Although selected comorbidities were assessed descriptively and incorporated into adjusted analyses where available, a harmonised cross-site measure of overall comorbidity burden was not consistently available across participating analytical datasets. Consequently, formal stratified analyses according to overall comorbidity burden were not performed, limiting assessment of whether treatment outcomes differed according to underlying patient health status. In addition, selected comorbidities may have been under-recorded in hospital EHR data when diagnosis and management occurred primarily in primary care settings, potentially contributing to residual confounding.

Taken together, these factors indicate that the currently available findings should be interpreted as observational associations rather than evidence of causal treatment effects. The consistency of findings across adjusted and sensitivity analyses strengthens confidence in the observed results; however, unmeasured and residual confounding remain possible.

5.6.4 Missingness, suppression, and biomarker limitations

Availability of prognostic variables differed across participating sites. ECOG-PS and LIPS-3 was available at Rigshospitalet but not at HUS, limiting the ability to perform comparable prognostic analyses across currently available cohorts.

Biomarker analyses planned in the protocol could not be fully implemented because only positive genomic test results were available in the source data. Consequently, the denominator of tested patients could not be established, and genomic biomarkers could not be characterised consistently across sites. In addition, exact PD-L1 values above the eligibility threshold were not consistently available and were therefore not included in adjusted analyses.

Missingness and small-cell suppression limit the level of detail that can be reported for some variables and may affect the feasibility of adjusted or subgroup analyses. Suppression should not be interpreted as missingness unless explicitly stated in the source output.

Missingness patterns differed across currently available cohorts. In Rigshospitalet, findings from diagnostic assessments did not support a Missing Completely at Random (MCAR) mechanism. In HUS, there was no evidence against a MCAR mechanism, although interpretation remains constrained by the smaller cohort size and limited number of observed events. Model diagnostics identified evidence of non-proportional hazards for LIPI category 1 in adjusted OS analyses at HUS. Additional time-stratified sensitivity analyses did not materially alter interpretation of the treatment group findings. Exploratory prognostic associations involving LIPI should therefore be interpreted cautiously.

6 Conclusions

The VALO2 analyses demonstrate the potential for generating clinically meaningful RWE across the participating Nordic healthcare institutions using a federated analytical framework based on the OMOP Common Data Model and local SPEs. The Rigshospitalet, HUS and OUS analyses provide site-specific evidence on outcomes associated with first-line immune checkpoint inhibitor monotherapy among patients with PD-L1 $\geq 50\%$ metastatic NSCLC.

In the Rigshospitalet cohort, assignment to the pembrolizumab policy-period group was associated with more favourable observed OS, TTNT, and TTD outcomes than assignment to the atezolizumab policy-period group. In the HUS cohort, no statistically significant differences were observed between pembrolizumab and cemiplimab across the evaluated clinical outcomes. In the OUS cohort, pembrolizumab was associated with longer OS than cemiplimab, whereas no statistically significant differences were observed for TTNT or TTD. Because treatment allocation was not randomised, treatment comparisons varied by site, and cohort construction depended on locally available data, these findings should be interpreted as site-specific observational evidence rather than definitive comparative-effectiveness estimates.

The study confirmed the operational feasibility of conducting federated RWE studies across multiple participating Nordic institutions and SPEs, including cross-border analytical access by Danish researchers to Finnish and Norwegian health data environments without transfer of patient-level sensitive data. At the same time, differences in variable availability, OMOP-formatted data content, cohort implementation, and reporting practices were identified as important determinants of cross-site comparability. The findings highlight that successful execution of a federated analytical framework does not automatically ensure comparable study populations or directly comparable results across sites.

The VALO2 experience demonstrates that federated RWE generation is feasible across Nordic healthcare systems and provides important lessons for future multinational studies using harmonised analytical approaches, OMOP-formatted analytical datasets, and SPEs. Federated analysis was operationally feasible, but a common data model and reusable analytical code did not ensure comparable cohorts, variables, or adjustment opportunities across participating sites.

6.1 Future research

Larger multi-country cohorts will be important for improving precision, supporting subgroup analyses, and enabling more robust evaluation of prognostic variables.

Future Nordic federated studies would benefit from more consistent capture and OMOP mapping of biomarker testing, PD-L1 status, ECOG performance status, metastatic disease status, treatment-line information, and cohort attrition metrics. Future studies would also benefit from a structured cross-site feasibility assessment prior to study initiation to evaluate data availability, completeness, eligibility-criteria implementation, and analytical feasibility across participating datasets.

The VALO2 experience demonstrates that a shared analytical framework alone is insufficient to ensure comparable evidence across countries; more consistent extraction, transformation, and mapping of relevant source data to the OMOP CDM, transparent cohort construction, and harmonised reporting remain essential for meaningful cross-country RWE generation. Future studies with larger cohorts, improved capture of prognostic variables, and more complete OMOP

implementation may improve the feasibility of comparative-effectiveness analyses across Nordic healthcare systems.

References

1. Abernethy AP, Gippet J, Parulkar R, Revol C. Use of electronic health record data for quality reporting. *J Oncol Pract.* 2017;13(8):530-534.
2. Garon EB, Hellmann MD, Rizvi NA, Carcereny E, Leighl NB, Ahn MJ, et al. Five-year overall survival for patients with advanced non-small-cell lung cancer treated with pembrolizumab: results from the phase I KEYNOTE-001 study. *J Clin Oncol.* 2019;37(28):2518-2527.
3. Gridelli C, Rossi A, Carbone DP, Guarize J, Karachaliou N, Mok T, et al. Non-small-cell lung cancer. *Nat Rev Dis Primers.* 2015;1:15009.
4. Hendriks LE, Kerr KM, Menis J, Mok TS, 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. *Ann Oncol.* 2023;34(4):358-376.
5. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature.* 2018;553(7689):446-454.
6. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ Jr, Wu YL, et al. Lung cancer: current therapies and new targeted treatments. *Lancet.* 2017;389(10066):299-311.
7. Jassem J, de Marinis F, Giaccone G, Vergnenegre A, Barrios CH, Morise M, et al. Updated overall survival analysis from IMpower110: atezolizumab versus platinum-based chemotherapy in treatment-naïve programmed death-ligand 1-selected NSCLC. *J Thorac Oncol.* 2021;16(11):1872-1882.
8. Kilickap S, Baramidze A, Sezer A, Özgüroğlu M, Gümüş M, Bondarenko I, et al. Cemiplimab monotherapy for first-line treatment of patients with advanced NSCLC with PD-L1 expression of 50% or higher: five-year outcomes of EMPOWER-Lung 1. *J Thorac Oncol.* 2025;20(7):941-954.
9. Luciani A, Ghidini A, Dottorini L, Petrelli F. Safety and effectiveness of immune checkpoint inhibitors in older patients with cancer: a systematic review of 48 real-world studies. *Drugs Aging.* 2021;38(12):1055-1065.
10. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Non-Small Cell Lung Cancer. Version 4.2026.
11. Observational Health Data Sciences and Informatics. The Book of OHDSI: Observational Health Data Sciences and Informatics.
12. Observational Medical Outcomes Partnership. OMOP Common Data Model Specifications.
13. Petrelli F, Inno A, Ghidini A, Gori S, Bersanelli M. Efficacy of immune checkpoint inhibitors in elderly patients aged ≥75 years. *Cancer Immunol Immunother.* 2021;70(6):1777-1780.

14. Sezer A, Kilickap S, Gümüş M, Bondarenko I, Özgüroğlu M, Gogishvili M, et al. Cemiplimab monotherapy for first-line treatment of advanced non-small-cell lung cancer with PD-L1 of at least 50%: a multicentre, open-label, global, phase 3, randomised, controlled trial. *Lancet*. 2021;397(10274):592-604.
15. Socinski MA, Jotte RM, Cappuzzo F, Orlandi F, Stroyakovskiy D, Nogami N, et al. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N Engl J Med*. 2018;378(24):2288-2301.
16. Charlot M, Gray JE. First-line immunotherapy and clinically meaningful survival benefits for the oldest adults with lung cancer. *JAMA Oncol*. 2023;9(3):342-343.
17. 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 Oncol*. 2023;9(3):334-341.
18. Reck M, Rodríguez-Abreu D, Robinson AG, et al. *Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer*. *N Engl J Med*. 2016;375:1823-1833.
19. Kazandjian D, Gong Y, Keegan P, Pazdur R, Blumenthal GM. Prognostic Value of the Lung Immune Prognostic Index for Patients Treated for Metastatic Non-Small Cell Lung Cancer. *JAMA Oncology*. 2019;5(10):1481-1485. doi:10.1001/jamaoncol.2019.1747.
20. Banna GL, Cortellini A, Cortinovis DL, et al. The Lung Immuno-Oncology Prognostic Score (LIPS-3): a prognostic classification of patients receiving first-line pembrolizumab for PD-L1 ≥50% advanced non-small-cell lung cancer. *ESMO Open*. 2021;6(2):100078.

Appendices

Appendix A. Study population and baseline characteristics

Appendix A presents country-specific baseline characteristics by first-line treatment group. Tables are reproduced from the final analytical outputs generated within each participating site and are provided to support interpretation of cohort composition, treatment group comparability, and data availability.



Figure A1: Eligibility attrition Denmark

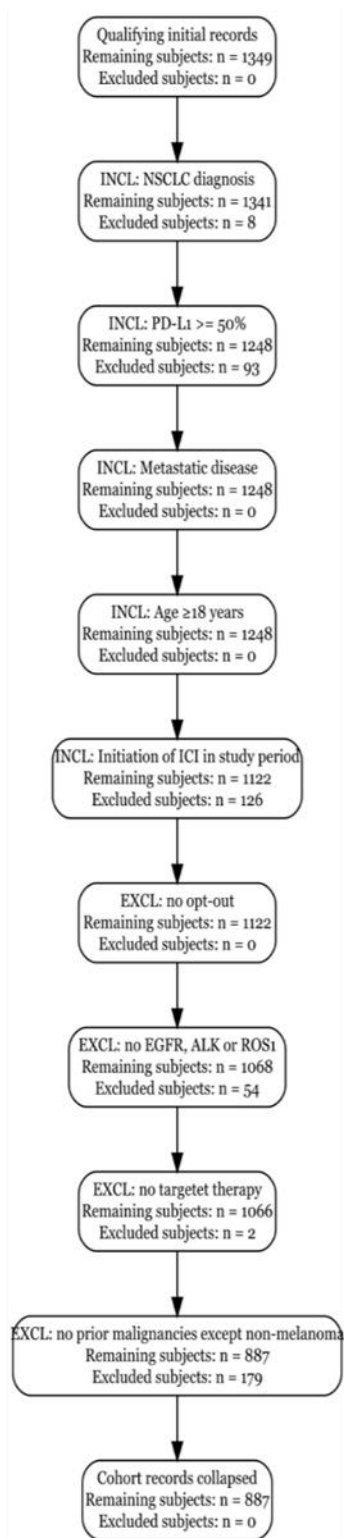




Figure A2: Eligibility attrition Finland

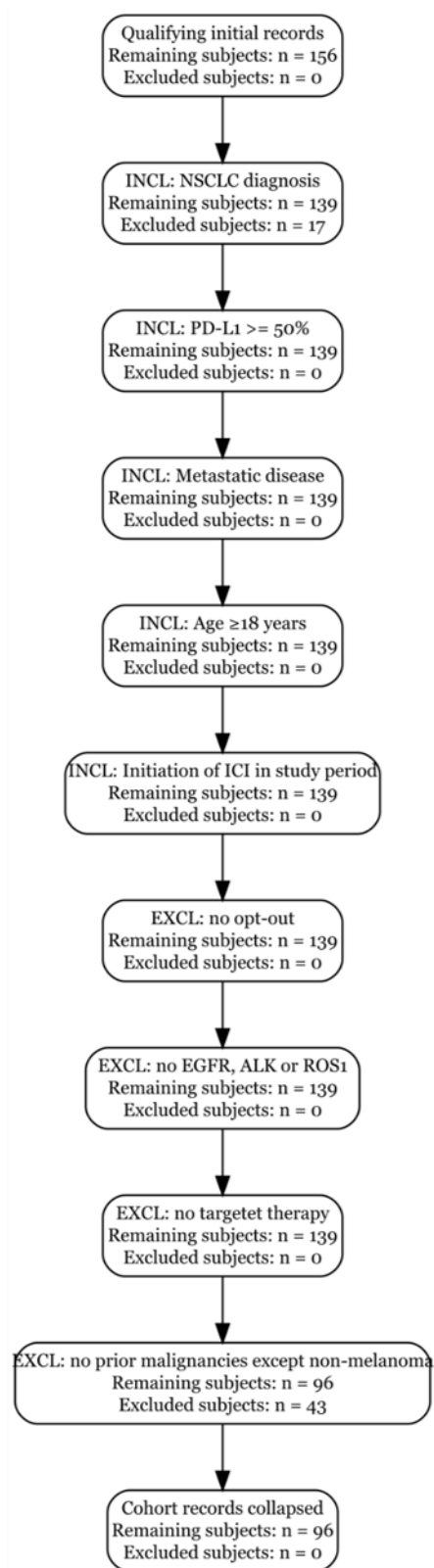


Figure A3: Eligibility attrition Norway

Patient selection flowchart for Norway, showing reduction from 6,897 initial records to 79 eligible patients

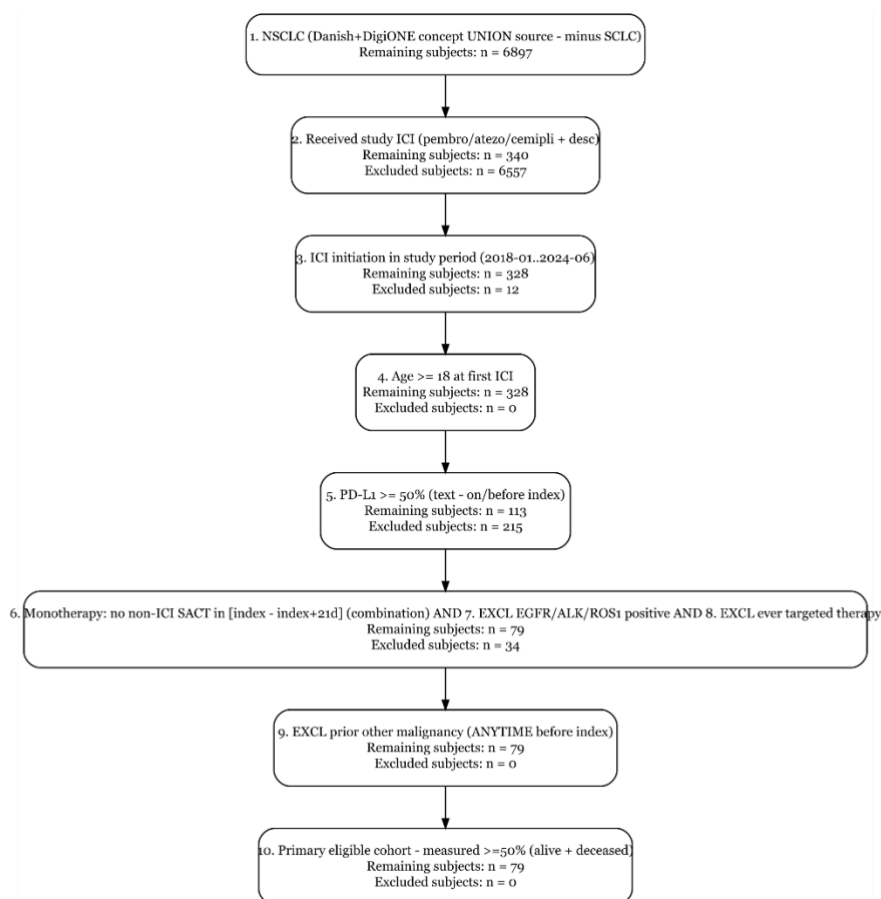


Table A1. Baseline characteristics of the Rigshospitalet analytical cohort by first-line treatment group

Characteristic	Overall N = 887 ¹	Atezolizumab N = 193 ¹	Pembrolizumab N = 694 ¹	p-value ²
Sex				0.3
Female	508 (57.3%)	104 (53.9%)	404 (58.2%)	
Male	379 (42.7%)	89 (46.1%)	290 (41.8%)	
Missing	0 (0%)	0 (0%)	0 (0%)	
Age	70 (64, 76)	69 (64, 76)	71 (64, 76)	0.5
Missing	0 (0%)	0 (0%)	0 (0%)	
Age group				0.8
<55	36 (4.1%)	7 (3.6%)	29 (4.2%)	
55–64	197 (22.2%)	42 (21.8%)	155 (22.3%)	
65–74	377 (42.5%)	88 (45.6%)	289 (41.6%)	
≥75	277 (31.2%)	56 (29.0%)	221 (31.8%)	
Missing	0 (0%)	0 (0%)	0 (0%)	
Smoking status				0.5
Ever	548 (94.5%)	129 (95.6%)	419 (94.2%)	
Never	32 (5.5%)	6 (4.4%)	26 (5.8%)	
Missing	307 (34.6%)	58 (30.1%)	249 (35.9%)	
Height, cm	169 (163, 176)	170 (164, 178)	169 (163, 176)	0.2
Missing	156 (17.6%)	44 (22.8%)	112 (16.1%)	
Weight, kg	70 (59, 81)	70 (60, 82)	69 (59, 81)	0.6
Missing	54 (6.1%)	11 (5.7%)	43 (6.2%)	
Comorbidities				
Cardiovascular	355 (40.0%)	83 (43.0%)	272 (39.2%)	0.3
COPD	129 (14.5%)	21 (10.8%)	108 (15.5%)	0.10
Diabetes	53 (6.0%)	13 (6.7%)	40 (5.8%)	0.6
Histology				0.11
Non-squamous	572 (79.4%)	134 (84.2%)	444 (78.4%)	
Squamous	147 (20.6%)	25 (15.8%)	122 (21.6%)	
Missing	162 (18.3%)	34 (17.6%)	128 (18.4%)	
ECOG-PS				0.15
0	297	63	234	
1	451	111	340	
2	65	11	54	
3	Suppressed	<5	10	
4	Suppressed	<5	<5	
Missing	Suppressed	<9	<57	
LIPI				0.6
0	408	86	322	
1	333	79	254	
2	Suppressed	<29	95	
Missing	Suppressed	<5	23	
LIPS3				0.050

Characteristic	Overall N = 887 ¹	Atezolizumab N = 193 ¹	Pembrolizumab N = 694 ¹	p-value ²
0	539	138	401	
1	188	32	156	
2	66	12	54	
3	Suppressed	<5	10	
Missing	Suppressed	<12	73	
Prior SACT	77 (8.7%)	16 (8.3%)	61 (8.8%)	0.8
Prior thoracic procedure	105 (11.8%)	25 (13.0%)	80 (11.5%)	0.6
Prior radiotherapy	180 (20.3%)	34 (17.6%)	146 (21.0%)	0.3

Values are presented as n (%) or median (Q1, Q3), unless otherwise specified. Abbreviations: COPD = chronic obstructive pulmonary disease; ECOG = Eastern Cooperative Oncology Group performance status; LIPI = Lung Immune Prognostic Index; LIPS-3 = Lung Immuno-Oncology Prognostic Score; SACT = systemic anti-cancer therapy. Small-cell suppression has been applied in accordance with local disclosure-control requirements. Suppressed values should not be interpreted as missing-data.

Table A2. Baseline characteristics of the HUS analytical cohort by first-line treatment group

Characteristic	Overall N = 91 ¹	Cemiplimab N = 17 ¹	Pembrolizumab N = 74 ¹	p-value ²
sex				0.3
Female	54 (59.3%)	12 (70.6%)	42 (56.8%)	
Male	37 (40.7%)	5 (29.4%)	32 (43.2%)	
Missing	0 (0%)	0 (0%)	0 (0%)	
Age	70 (62, 75)	72 (62, 77)	70 (63, 75)	0.5
Missing	0 (0%)	0 (0%)	0 (0%)	
Age group				0.2
<65	28 (30.8%)	5 (29.4%)	23 (31.1%)	
65-74	37 (40.7%)	5 (29.4%)	32 (43.2%)	
≥75	26 (28.6%)	7 (41.2%)	19 (25.7%)	
Missing	0 (0%)	0 (0%)	0 (0%)	
height, cm	167 (160, 174)	163 (160, 173)	167 (160, 174)	0.6
Missing	15 (16.5%)	5 (29.4%)	10 (13.5%)	
weight, kg	67 (59, 79)	64 (58, 68)	68 (59, 80)	0.2
Missing	12 (13.2%)	5 (29.4%)	7 (9.5%)	
Comorbidities				
Cardiovascular	45 (49.5%)	8 (47.1%)	37 (50.0%)	0.8
COPD	Suppressed	<5	16 (21.6%)	0.5
Diabetes	Suppressed	<5	<5	>0.9
LIPI				0.8
0	41	8	33	
1	Suppressed	<10	25	
2	Suppressed	<5	5	
Missing	Suppressed	<5	11	
Prior SACT	Suppressed	<5	<5	>0.9
Prior thoracic procedure	Suppressed	<5	5 (6.8%)	>0.9

Prior radiotherapy	22 (24.2%)	5 (29.4%)	17 (23.0%)	0.5
¹ n (%); Median (Q1, Q3)				
² Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test				

Table A3. Baseline characteristics of the OUS analytical cohort by first-line treatment group

Characteristic	Overall N = 76 ¹	Cemiplimab N = 12 ¹	Pembrolizumab N = 64 ¹	p-value ²
sex				0.2
Female	Suppressed	Suppressed	34 (53.1%)	
Male	Suppressed	Suppressed	30 (46.9%)	
Missing	0 (0%)	0 (0%)	0 (0%)	
age	71 (64, 77)	76 (67, 79)	71 (63, 76)	0.2
Missing	0 (0%)	0 (0%)	0 (0%)	
Age group				0.8
18-54	Suppressed	Suppressed	7 (10.9%)	
55-64	Suppressed	Suppressed	12 (18.8%)	
65-74	Suppressed	Suppressed	24 (37.5%)	
≥75	27 (36%)	6 (50%)	21 (32.8%)	
Missing	0 (0%)	0 (0%)	0 (0%)	
Smoking status	0	0	0	
Missing	76 (100%)	12 (100%)	64 (100%)	
Height,cm				
Missing	Suppressed	Suppressed	Suppressed	
Weight,kg				
Missing	Suppressed	Suppressed	Suppressed	
Cardiovascular	37 (48.7%)	7 (58.3%)	30 (46.9%)	0.5
COPD	Suppressed	Suppressed	10 (15.6%)	0.3
Diabetes	Suppressed	Suppressed	Suppressed	0.063
Histology				0.4
Adenocarcinoma	Suppressed	Suppressed	19 (68.0%)	
Squamous	Suppressed	Suppressed	8 (29.0%)	
Other	Suppressed	Suppressed	Suppressed	
Missing	41 (53.9%)	5 (41.6%)	Suppressed	
ECOG-PS				>0.9
0	Suppressed	Suppressed	Suppressed	
1	Suppressed	Suppressed	9	
2	Suppressed	Suppressed	10	
3	Suppressed	Suppressed	Suppressed	
4	Suppressed	Suppressed	Suppressed	
Missing	Suppressed	Suppressed	45	
LIPI	0	0	0	
Missing	76	12	64	

LIPS3	0	0	0	
Missing	76	12	64	
stage	0	0	0	
Missing	76	12	64	
Prior SACT	Suppressed	Suppressed	7 (10.9%)	>0.9
Prior thoracic procedure	Suppressed	Suppressed	9 (14.0%)	>0.9
Prior radiotherapy	0	0	0	
Missing	76 (100%)	12 (100%)	64 (100%)	

¹n (%); Median (Q1, Q3)

²Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test; NA

Appendix B. Country-specific treatment-pattern tables

Appendix B provides country-specific treatment-pattern summaries, including first-line treatment distribution, treatment sequencing, and treatment duration estimates. Tables are reproduced from the final analytical outputs generated within each participating site and are provided to support interpretation of treatment utilisation and treatment pathways across the available analytical cohorts.

Table B1. First-line treatment distribution by site

Site	Treatment	N	%
Rigshospitalet	Pembrolizumab	694	78.2
Rigshospitalet	Atezolizumab	193	21.8
HUS	Pembrolizumab	74	81.3
HUS	Cemiplimab	17	18.7
OUS	Pembrolizumab	64	84.2
OUS	Cemiplimab	12	15.8

Treatment distributions reflect the final analytical cohorts and should not be interpreted as complete descriptions of local prescribing practice.

Table B2. Rigshospitalet treatment sequencing

Treatment	Category	Line 1	Line 2	Line 3	Line 4
Pembro	PD-1 inhibitor, n	692	10	<5	<5
Pembro	PD-L1 inhibitor, n	<5	8	<5	<5
Pembro	Non-platinum based chemo, n	<5	79	92	25
Pembro	Other, n	<5	5	<5	<5
Pembro	Platinum-based chemo, n	<5	186	8	6
Pembro	Months on line, median (IQR)	4.88 (1.52;13.41)	2.3 (1.45;3.8)	1.97 (0.7;4.33)	1.58 (0.68;3.26)
Atezo	PD-1 inhibitor, n	38	9	<5	<5
Atezo	PD-L1 inhibitor, n	155	<5	<5	<5
Atezo	Non-platinum based chemo, n	<5	25	20	8
Atezo	Other, n	<5	<5	<5	<5
Atezo	Platinum-based chemo, n	<5	58	6	<5
Atezo	Months on line, median (IQR)	2.55 (1;7.37)	2.17 (0.92;3.2)	1.6 (0.47;2.81)	1.45 (1.39;1.61)

Small-cell suppression has been applied in accordance with local disclosure-control requirements.

Table B3. HUS treatment sequencing

Treatment	Category	Line 1	Line 2	Line 3
Pembro	PD-1 inhibitor, n	74	<5	<5
Pembro	Non-platinum based chemo, n	<5	23	5
Pembro	Other, n	<5	<5	<5
Pembro	PD-L1 inhibitor, n	<5	<5	<5
Pembro	Platinum-based chemo, n	<5	19	<5
Pembro	Months on line, median (IQR)	2.44 (0.7;6.84)	2.55 (1.68;5.26)	2.52 (0.95;4.45)
Cemipli	PD-1 inhibitor, n	17	<5	<5
Cemipli	Non-platinum based chemo, n	<5	6	<5
Cemipli	Other, n	<5	<5	<5
Cemipli	PD-L1 inhibitor, n	<5	<5	<5
Cemipli	Platinum-based chemo, n	<5	6	<5
Cemipli	Months on line, median (IQR)	3.6 (1.37;5.81)	1.93 (1.54;3.4)	1.5 (0.38;2.83)

Small-cell suppression has been applied in accordance with local disclosure-control requirements.

Table B4. OUS treatment sequencing

Treatment_group	Category	Line 1	Line 2	Line 3	Line 4
Pembro	PD-1 inhibitor, n	64	<5	<5	<5
Pembro	Non-platinum based chemo, n	<5	13	7	<5
Pembro	Platinum-based chemo, n	<5	11	<5	<5
Pembro	PD-L1 inhibitor, n	<5	<5	<5	<5
Pembro	Months on line, median (IQR)	4.18 (1.63;17.78)	2.4 (0.98;2.96)	2.79 (1.94;4.87)	2.03 (1.79;4.23)
Cemip	PD-1 inhibitor, n	12	<5		
Cemip	Non-platinum based chemo, n	<5	<5		
Cemip	Platinum-based chemo, n	<5	<5		
Cemip	PD-L1 inhibitor, n	<5	<5		
Cemip	Months on line, median (IQR)	2.74 (0.51;5.68)	2.8 (1.75;3.85)		

Small-cell suppression has been applied in accordance with local disclosure-control requirements.

Table B5. Summary of first-line treatment duration by country and treatment group

Country	Treatment	Line 1 Median Months (IQR)
Rigshospitalet	Pembrolizumab	4.88 (1.52-13.41)
Rigshospitalet	Atezolizumab	2.55 (1.00-7.37)
HUS	Pembrolizumab	2.44 (0.70-6.84)
HUS	Cemiplimab	3.60 (1.37-5.81)
OUS	Pembrolizumab	4.18 (1.63-17.78)
OUS	Cemiplimab	2.74 (0.51-5.68)

Appendix C. Site-specific survival analyses

C1. Overall survival

This section presents KM overall survival (OS) analyses for the Rigshospitalet, HUS, and OUS cohorts. KM plots are provided together with summary estimates of survival distributions and supporting event and censoring information.

Figure C1. KM overall survival by treatment group in the Rigshospitalet cohort

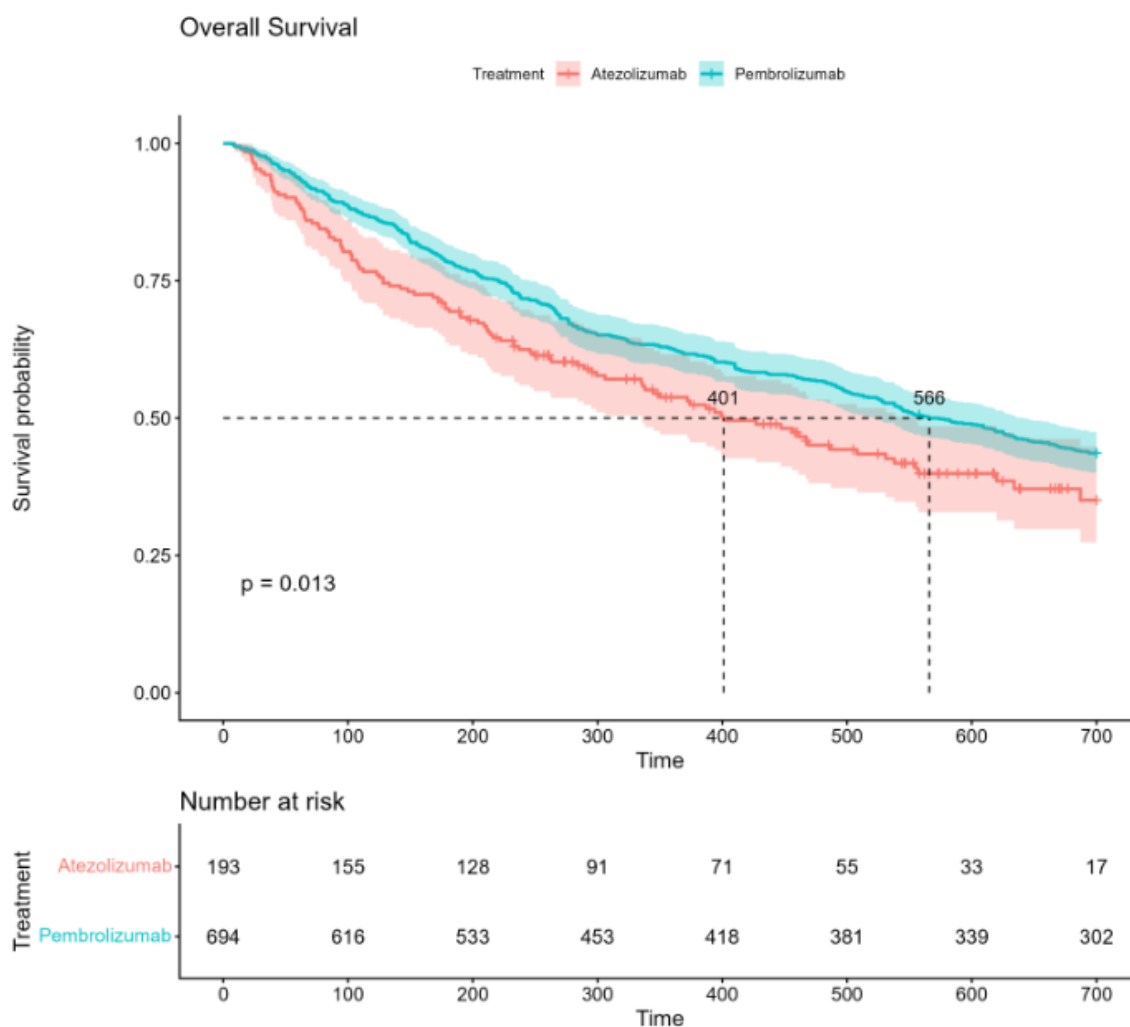


Table C1. Rigshospitalet KM survival summary

Treatment	25th percentile (95% CI)	Median OS (95% CI)	75th percentile
Atezolizumab	128 (96-196)	401 (335-555)	NR
Pembrolizumab	221 (185-246)	566 (525-641)	NR

Table C2. Rigshospitalet risk –event-censor summary

Treatment	Events	Censored	At risk at 700 days
Atezolizumab	107	86	17
Pembrolizumab	391	303	302

Figure C2. KM overall survival by treatment group in the HUS cohort

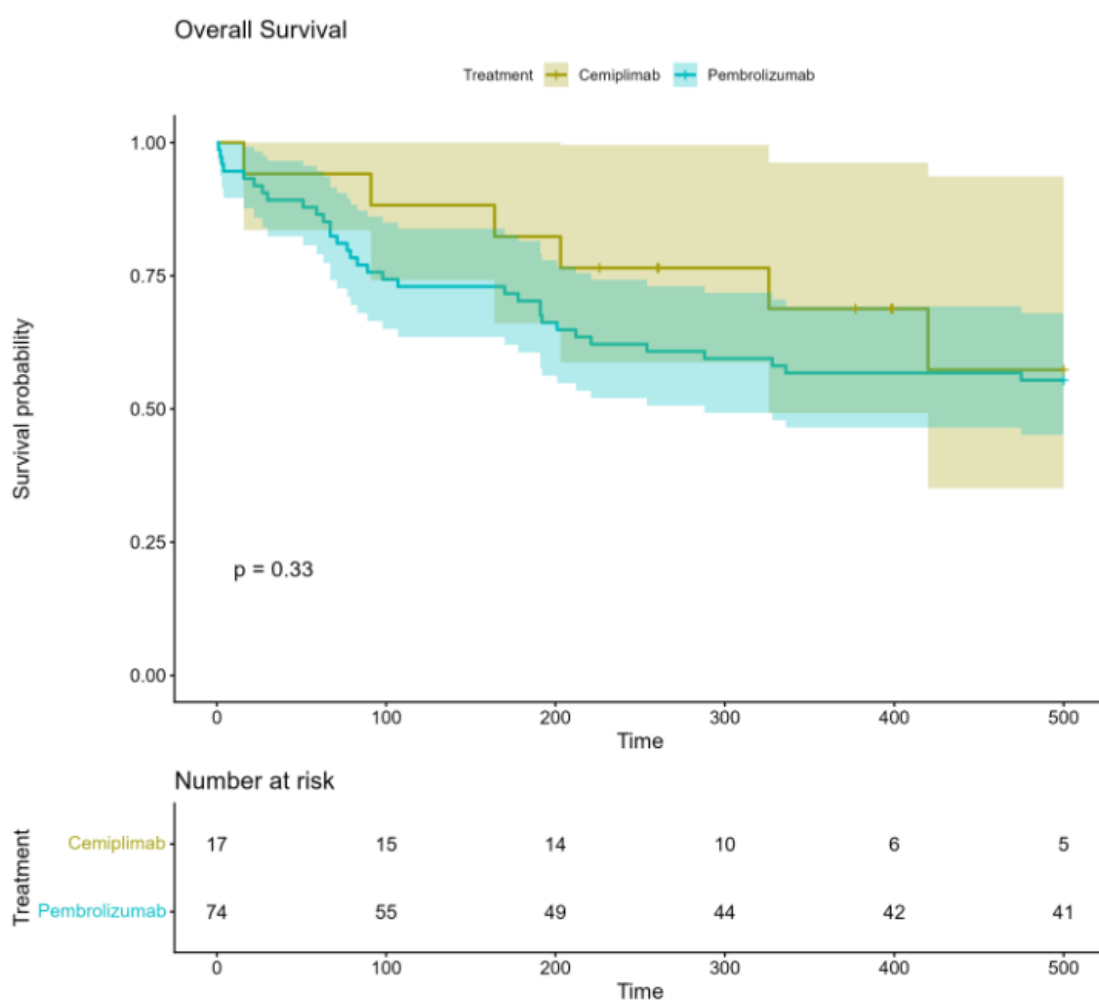


Table C3. HUS KM survival summary

Treatment	25th percentile (95% CI)	Median OS (95% CI)	75th percentile
Cemiplimab	326 (91-NA)	NR	NR
Pembrolizumab	98 (67-221)	NR	NR

Table C4. HUS risk-event-censor summary

Treatment	Events	Censored	At risk at 500 days
Cemiplimab	6	11	5
Pembrolizumab	33	41	41

* **Abbreviations:** CI = confidence interval; NR = not reached; OS = overall survival.

Figure C3. KM overall survival by treatment group in the OUS cohort

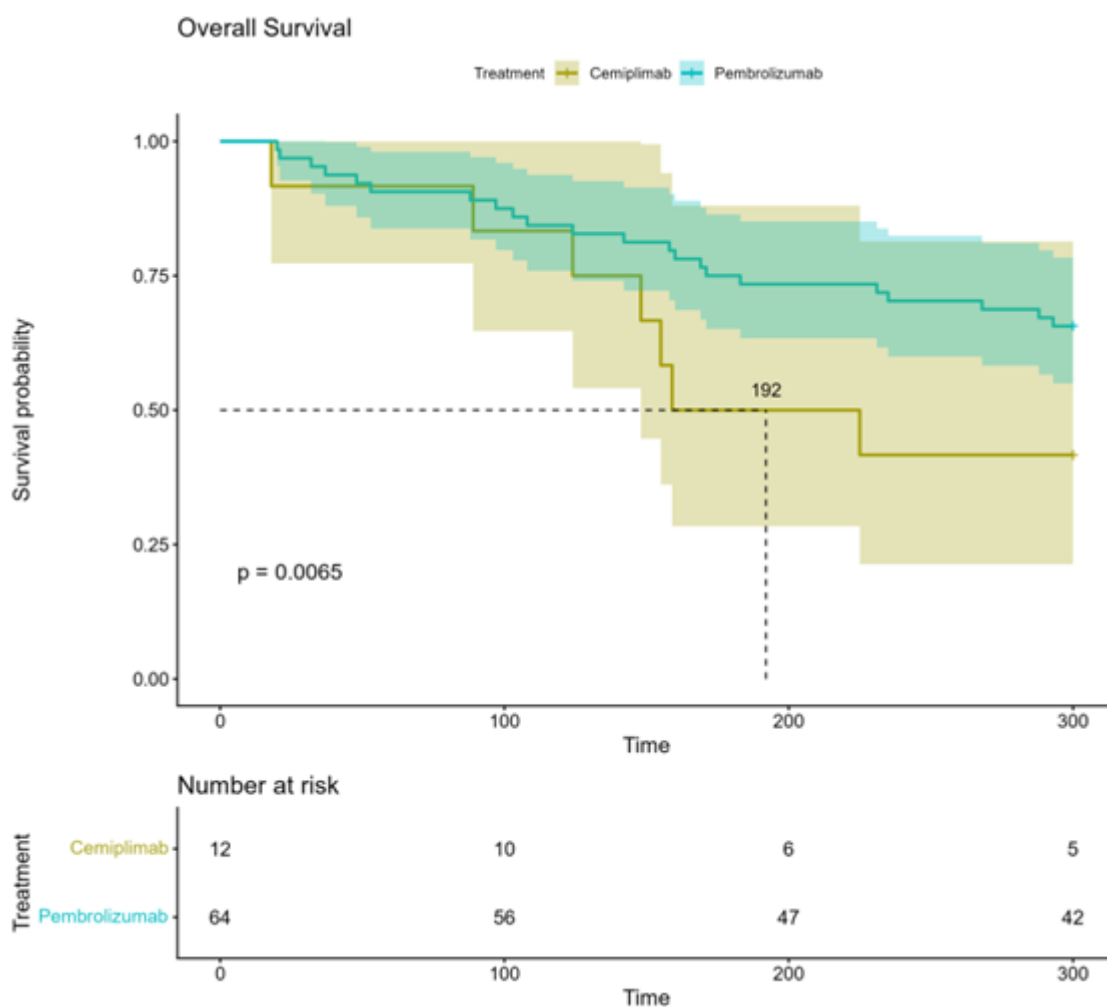


Table C5. OUS KM survival summary

Treatment	25th percentile (95% CI)	50th percentile (95% CI)	75th percentile (95% CI)
Cemiplimab	136 (89-NA)	192 (148-NA)	NA (225-NA)
Pembrolizumab	177 (124-NA)	NA (NA-NA)	NA (NA-NA)

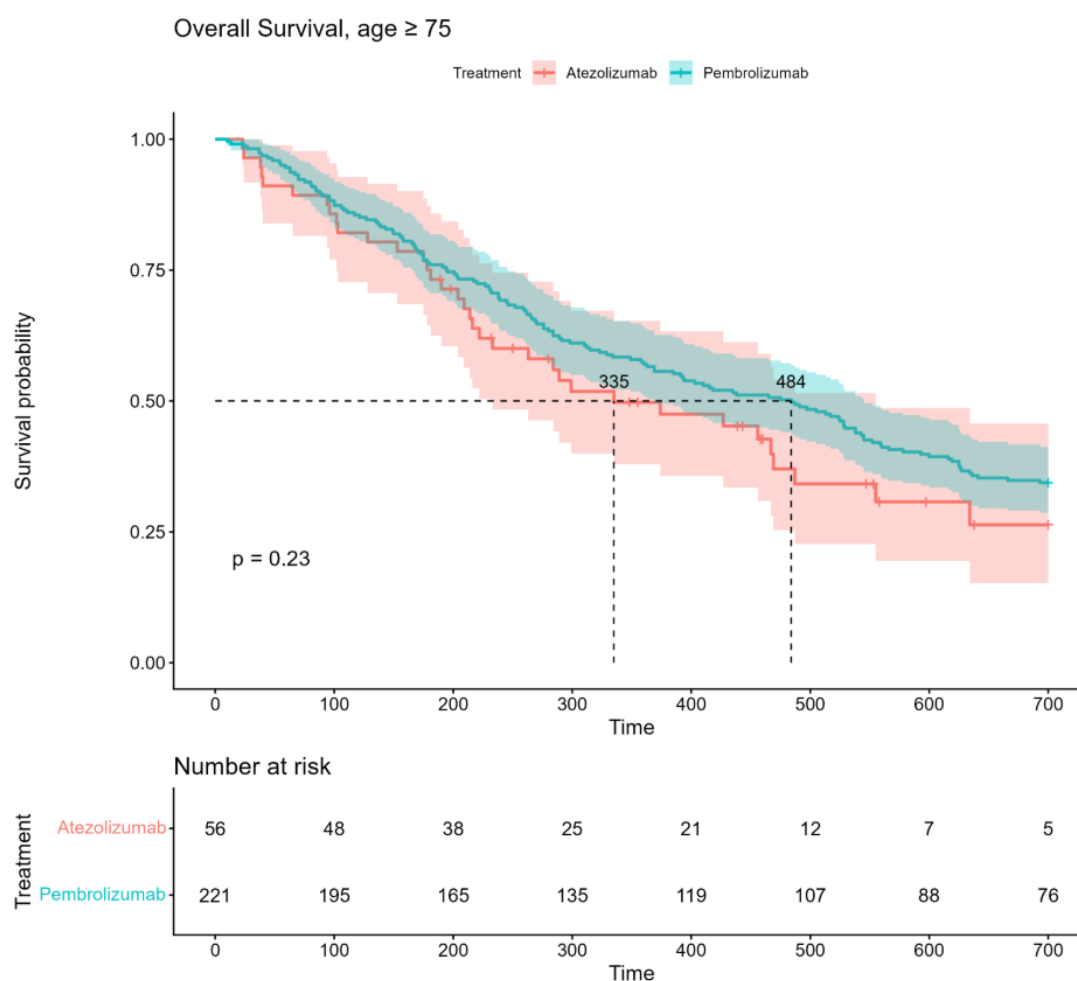
Table C6. OUS risk-event-censor summary

Treatment	Event	Censor	At risk at 300 days
Cemiplimab	7	5	5
Pembrolizumab	22	42	42

C2. Adjusted and age-stratified overall survival analyses

Age-stratified KM analyses were performed to explore overall survival among patients aged ≥ 75 years and < 75 years within each country-specific cohort. These analyses should be interpreted descriptively given reduced sample-size and event counts within subgroup strata. Age-stratified analyses were not reported for OUS in the available analytical output.

Figure C4. KM overall survival by treatment group among patients aged ≥ 75 years in the Rigshospitalet cohort.



KM estimates of overall survival among patients aged ≥ 75 years in the Rigshospitalet cohort, stratified by first-line treatment group. Tick marks indicate censored observations. Abbreviations: OS = overall survival

Figure C5. KM overall survival by treatment group among patients aged <75 years in the Rigshospitalet cohort. KM estimates of overall survival among patients aged <75 years in the Rigshospitalet cohort, stratified by first-line treatment group. Tick marks indicate censored observations. Abbreviations: OS = overall survival.

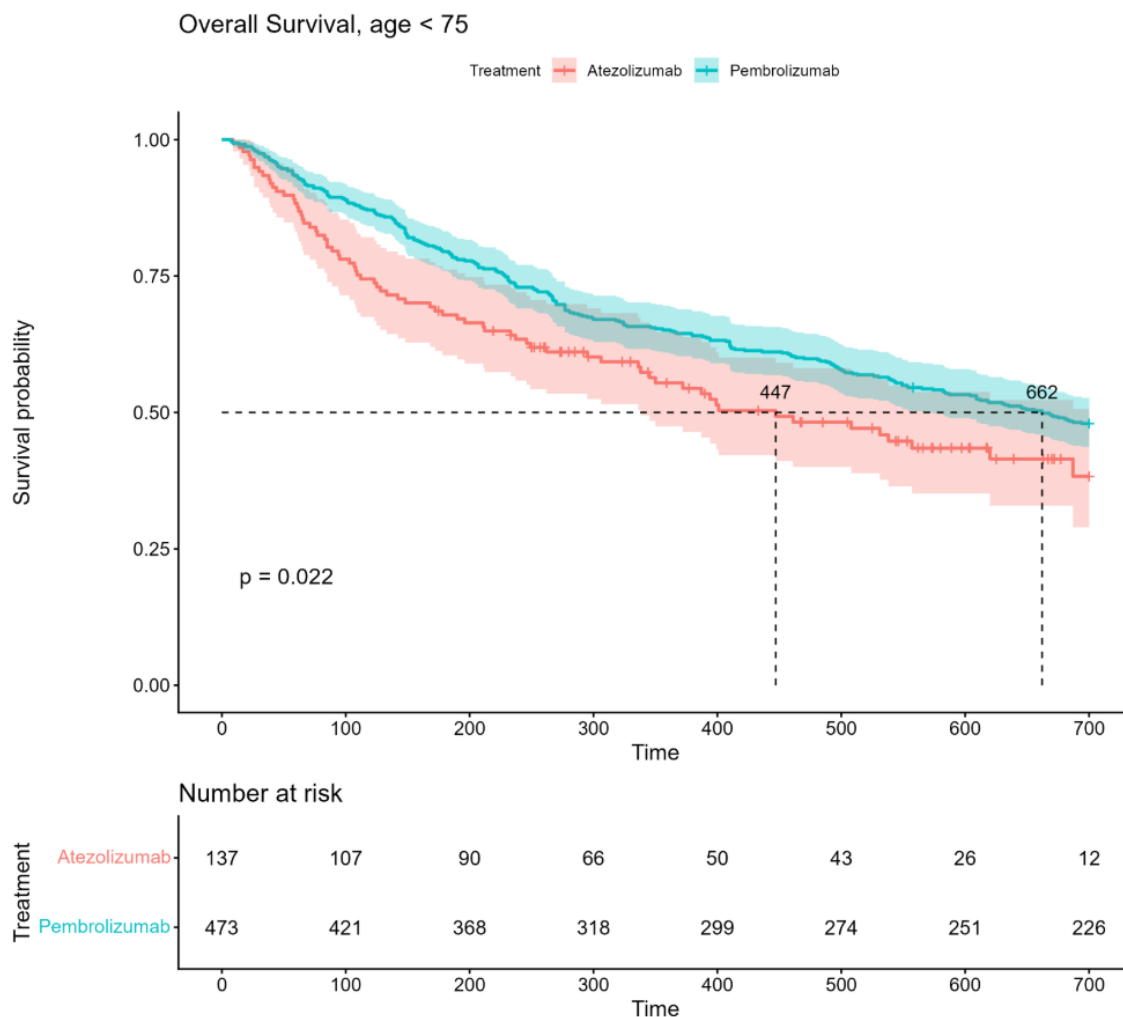


Figure C6. KM overall survival by treatment group among patients aged ≥ 75 years in the HUS cohort. KM estimates of overall survival among patients aged ≥ 75 years in the HUS cohort, stratified by first-line treatment group. Tick marks indicate censored observations. Abbreviations: OS = overall survival.

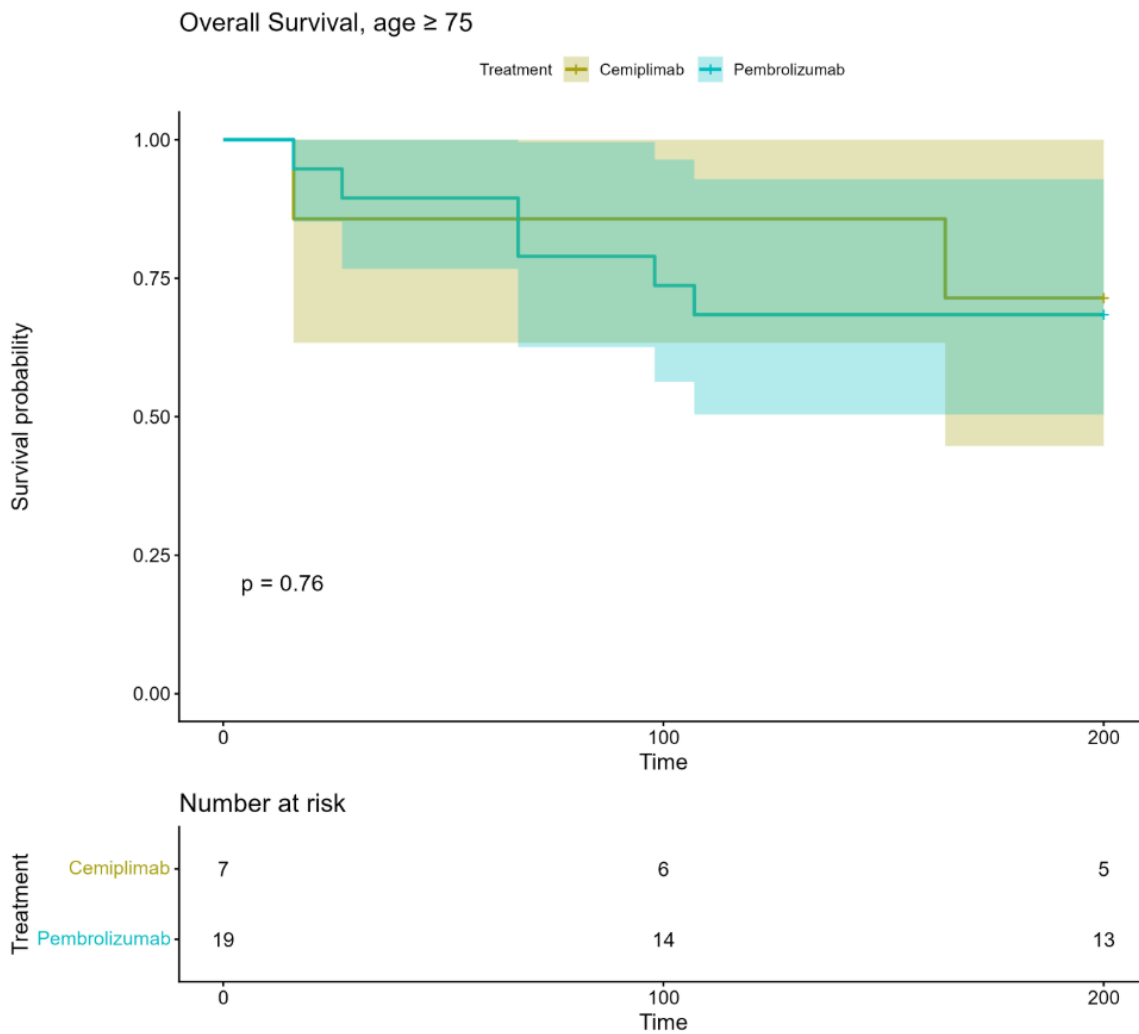


Figure C7. KM overall survival by treatment group among patients aged <75 years in the HUS cohort. KM estimates of overall survival among patients aged <75 years in the HUS cohort, stratified by first-line treatment group. Tick marks indicate censored observations. Abbreviations: OS = overall survival.

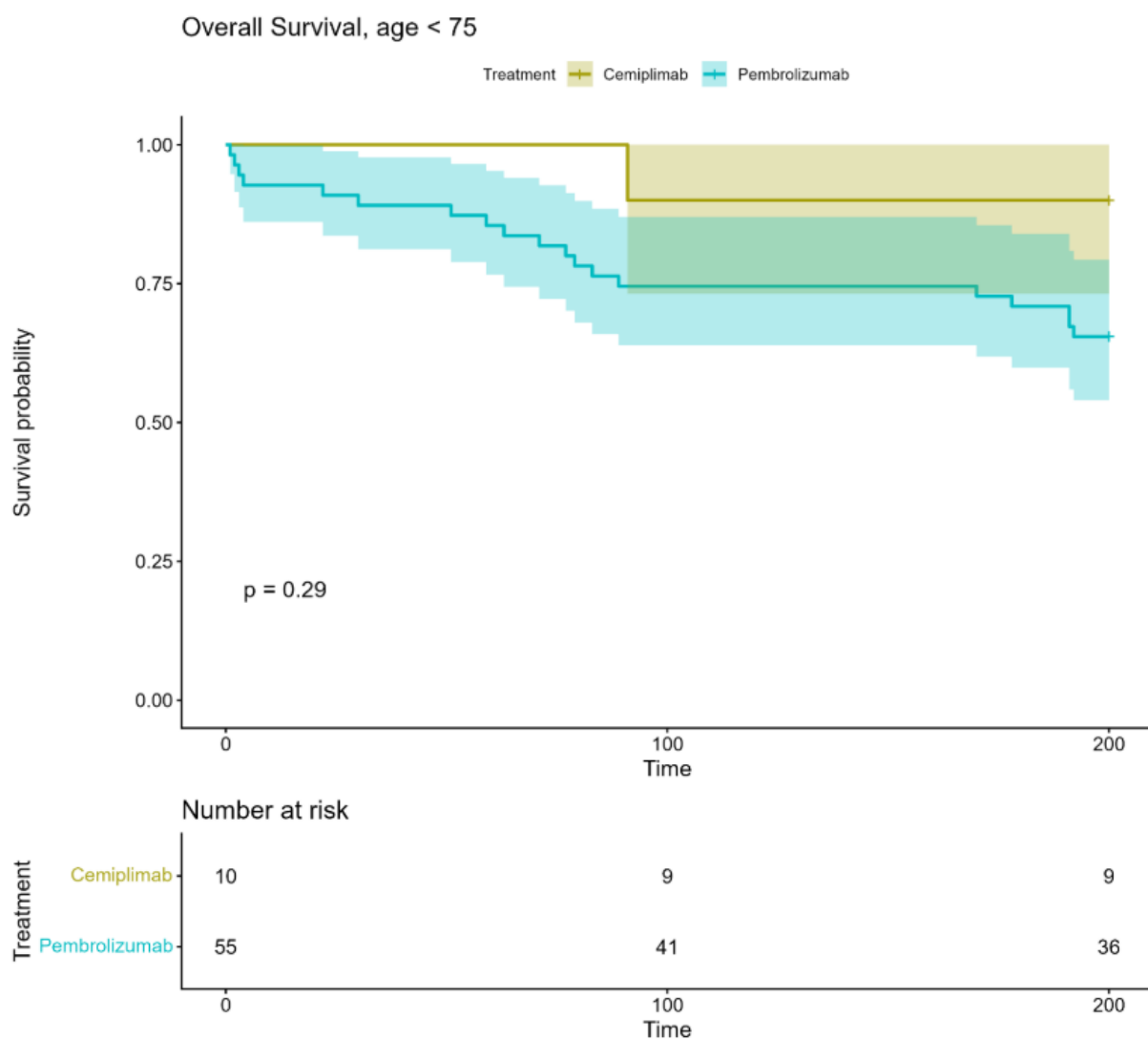


Table C7. Age-stratified adjusted Cox models for OS

Country	Age group	HR (95% CI)	p-value
Rigshospitalet	≥75 years	0.72 (0.49-1.07)	0.104
Rigshospitalet	<75 years	0.74 (0.56-0.97)	0.031
HUS	≥75 years	0.72 (0.10-5.24)	0.698
HUS	<75 years	2.22 (0.60-8.18)	0.221

Treatment-effect estimates were generally consistent across age strata in the Rigshospitalet cohort. In the HUS cohort, subgroup estimates were associated with substantial uncertainty owing to limited sample-size and event counts.

Primary overall survival analyses were performed using multivariable Cox proportional hazards regression models adjusted for baseline covariates and pooled across multiple imputed datasets. Treatment-effect estimates from the primary adjusted models are presented in Table C8.

In the Rigshospitalet cohort, the estimated association favoured the pembrolizumab policy-period group in both age strata, although the association was statistically significant only among patients aged <75 years. In the HUS cohort, estimates were associated with substantial uncertainty because of the limited subgroup sizes and event counts. Age-stratified analyses were not reported for OUS.

Table C8. Site-specific multivariable Cox models for overall survival

Site	Comparison	Analysis	HR (95% CI)	p-value
Rigshospitalet	Pembrolizumab versus atezolizumab policy-period group	Multivariable model after multiple-imputation	0.72 (0.58–0.90)	0.004
HUS	Pembrolizumab versus cemiplimab	Multivariable model after multiple-imputation	1.48 (0.60–3.67)	0.391
OUS	Pembrolizumab versus cemiplimab	Available-case multivariable model before multiple-imputation	0.42 (0.20–0.88)	0.022

* **Abbreviations:** CI = confidence interval; HR = hazard ratio; HUS= Helsinki University Hospital; OS = overall survival; OUS = Oslo University Hospital; SACT = systemic anti-cancer therapy.

Note: Exposure groups at Rigshospitalet were assigned according to predefined treatment-policy periods. HUS and OUS exposure groups reflected the PD-(L)1 inhibitor initiated in clinical practice. Multiple-imputation was not performed for OUS. The OUS model included treatment group, age, sex, prior SACT, and prior thoracic procedure. Because model specification and missing-data handling differed across sites, effect estimates should not be regarded as directly comparable between countries.

C3. Sensitivity analyses

Sensitivity analyses were performed to evaluate the robustness of the primary overall survival findings to alternative modelling approaches. For Rigshospitalet and HUS, these analyses included unadjusted Cox proportional hazards models and IPTW-adjusted Cox models. For OUS, only an unadjusted Cox model was available because multiple-imputation and IPTW were not performed. Across participating sites, sensitivity analyses did not materially alter the interpretation of the primary findings.

Table C9. Sensitivity analyses for overall survival

Country	Analysis	HR (95% CI)	p-value
Rigshospitalet	Unadjusted Cox	0.76 (0.61-0.94)	0.013
Rigshospitalet	IPTW-adjusted Cox	0.76 (0.60-0.95)	0.016
HUS	Unadjusted Cox	1.53 (0.65-3.62)	0.329
HUS	IPTW-adjusted Cox	1.41 (0.56-3.55)	0.469
OUS	Unadjusted Cox	0.38 (0.18–0.78)	0.009
OUS	IPTW-adjusted Cox	Not performed	-

Sensitivity analyses did not materially alter the interpretation of the primary overall survival analyses. IPTW-adjusted analyses were not performed for OUS because multiple-imputation was not conducted.

C4. Competing-risk analyses

Competing-risk analyses were performed to evaluate treatment effects for time to next treatment or death (TTNT) and time to treatment discontinuation or death (TTD). Fine–Gray subdistribution hazard models were used to account for competing events and to provide estimates of relative treatment effects across the country-specific cohorts.

Table C10. Treatment effects in competing-risk analyses

Country	Endpoint	Estimate
Rigshospitalet	TTNT	sHR 0.71 (0.57-0.87)
Rigshospitalet	TTD	sHR 0.79 (0.65-0.95)
HUS	TTNT	sHR 1.45 (0.67-3.14)
HUS	TTD	sHR 1.06 (0.63-1.79)
OUS	TTNT	sHR 0.65 (0.26–1.62)
OUS	TTD	sHR 0.97 (0.41–2.31)

In the Rigshospitalet cohort, assignment to the pembrolizumab policy-period group was associated with lower subdistribution hazards of both TTNT and TTD than assignment to the atezolizumab policy-period group. In the HUS cohort, no statistically significant differences were observed between pembrolizumab and cemiplimab. Similarly, no statistically significant differences were observed for TTNT or TTD in the OUS cohort. These findings should be interpreted in conjunction with the overall survival analyses and within the context of differences in sample-size, treatment comparisons, and event counts across participating sites.

C5. Exploratory prognostic analyses

LIPI and LIPS-3 analyses were performed as exploratory assessments of prognostic value within the available country-specific datasets. Because these analyses were not primary study objectives and were subject to variable availability and sample-size limitations, findings should be considered hypothesis-generating rather than confirmatory. LIPI and LIPS-3 analyses were not performed for OUS because the variables required to derive these scores were unavailable in the analytical output.

Table C11. Selected exploratory prognostic findings

Country	Endpoint	Variable	Estimate
Rigshospitalet	OS	LIPI 1	HR 1.44 (1.19-1.73)
Rigshospitalet	OS	LIPI 2	HR 1.88 (1.40-2.51)
Rigshospitalet	TTD	LIPI 1	sHR 1.24 (1.06-1.46)
Rigshospitalet	TTD	LIPI 2	sHR 1.65 (1.23-2.22)
HUS	TTD	LIPI 2	sHR 2.50 (1.12-5.56)

Table C12. Summary of treatment effects from competing-risk analyses

Country	Endpoint	Treatment-Effect
Rigshospitalet	TTNT	sHR 0.71 (0.57–0.87)
Rigshospitalet	TTD	sHR 0.79 (0.65–0.95)
HUS	TTNT	sHR 1.45 (0.67–3.14)
HUS	TTD	sHR 1.06 (0.63–1.79)
OUS	TTNT	sHR 0.65 (0.26–1.62)
OUS	TTD	sHR 0.97 (0.41–2.31)

Appendix D. Model diagnostics and assumptions

Appendix D summarises the diagnostic checks and modelling assumptions relevant to the country-specific survival and competing-risk analyses. These materials are intended to support transparency around model specification, covariate selection, proportional hazards considerations, imputation approaches, sparse-data constraints, and the interpretation of exploratory analyses.

D1. Missing-data assessment

Missing-data handling and imputation assumptions. Multiple-imputation approaches were used where appropriate for baseline covariates with missingness. Multiple-imputation was performed for the Rigshospitalet and HUS analyses but was not performed for OUS because of the smaller cohort size and reduced set of adjustment variables. Treatment exposure, outcome status, and derived event times were not imputed.

Table D1. Little's MCAR test results

Country	Statistic	Degrees of Freedom	p-value
Rigshospitalet	460.69	361	<0.001
HUS	21.91	26	0.694

Little's MCAR test indicated that the MCAR assumption was not supported in Rigshospitalet ($p < 0.001$) but was not contradicted in HUS ($p = 0.694$). These findings informed the multiple-imputation strategy applied in those analyses. No corresponding assessment was performed for OUS because multiple-imputation was not conducted.

D2. Cox proportional hazards assessment

Cox proportional hazards diagnostics. Overall Survival models were evaluated with attention to proportional hazards assumptions, influential observations, event counts, and the stability of treatment-effect estimates after covariate adjustment.

Where proportional hazards assumptions could not be fully assessed because of sparse events or limited follow-up, results should be interpreted descriptively and in conjunction with KM curves, confidence intervals, and sensitivity analyses.

Table D2. Summary of proportional hazards diagnostic results

Country	Variable	p-value
Rigshospitalet	Treatment group	0.078
Rigshospitalet	LIPi category 2	<0.001
Rigshospitalet	LIPS-3 category 2	0.020
Rigshospitalet	Global test	<0.001
HUS	LIPi category 1	0.007
HUS	Global test	0.140
OUS	Treatment group	0.388
OUS	Global test	0.497

For OUS, Schoenfeld residual assessments did not indicate evidence of violation of the proportional hazards assumption for the treatment group variable ($p = 0.388$) or the global model ($p = 0.497$).

D3. Multicollinearity assessment

Potential multicollinearity among covariates was assessed using variance inflation factors (VIFs). VIF estimates were reviewed to evaluate model stability and the potential impact of correlated covariates on regression estimates.

Table D3. Summary of multicollinearity assessment results

Country	Maximum VIF*	Interpretation
Rigshospitalet	1.55	No evidence of problematic multicollinearity
HUS	1.94	No evidence of problematic multicollinearity
OUS	1.62	No evidence of problematic multicollinearity

(using the highest Norwegian VIF: sex = 1.624650)

D4. IPTW balance assessment

All adjusted standardised differences remained below commonly used thresholds indicative of major imbalance in analyses where IPTW was performed. IPTW analyses were not conducted for OUS.

Statistical Annex

VALO 2 NSCLC pilot study – Value from Nordic Health Data

Statistical Annex based on Protocol v1.0 dated 29 January 2026

The Finnish Innovation Fund Sitra (on behalf of the Nordic Council of Ministers)

Itämerenkatu 11-13

00181 Helsinki

Finland

12 March 2026

Authors

Role Principal Investigator

Name: Andreas Bjerrum

Role: IQVIA Biostatistician

Name: Doneal Thomas

Role: IQVIA Senior Biostatistical Reviewer

Name: Gerald Lebovic

Table of Contents

List of Abbreviations.....	4
Executive summary.....	6
1 Introduction	7
1.1 Disease Burden and Epidemiology	7
1.2 Molecular Landscape and biomarker-driven treatment	7
1.3 First-line Immunotherapy for Patients with PD-L1 $\geq 50\%$ mNSCLC.....	8
1.4 Treatment Challenges and Patient Heterogeneity	8
1.5 Study Rationale.....	9
2 Objectives	11
2.1 Study Aim.....	11
2.2 Primary Objective.....	11
2.3 Secondary Objectives.....	11
2.4 Exploratory Objectives.....	11
2.5 Infrastructure Objective.....	12
3 Methods.....	13
3.1 Study Design.....	13
3.2 Data Sources	13
3.3 Federated Analysis Framework.....	14
3.4 Study Population.....	14
3.4.1 Inclusion Criteria	14
3.4.2 Exclusion Criteria	15
3.4.3 Site-specific Cohort Implementation.....	15
3.5 Cohort Definition & Index Date	16
3.6 Study Variables.....	16
3.6.1 Variable Availability and Implementation	16
3.6.2 Prognostic Scores.....	17
3.7 Treatment Variables	18
3.8 Study Outcomes	18
3.8.1 Primary Outcome	18
3.8.2 Secondary Outcomes	18
3.8.3 Additional Outcomes.....	19
3.8.4 Censoring and Competing-risk Rules.....	19
3.9 Statistical Analysis	19
3.9.1 Primary Objective: Overall survival	20
3.9.2 Secondary Objectives: Treatment Patterns and Sequences.....	20
3.9.3 Secondary Objectives: TTNT and TTD	21
3.9.4 Exploratory Prognostic Analyses.....	21
3.9.5 Subgroup Analyses.....	21
3.9.6 Handling of Missing-data and Partial Dates	22
3.9.7 Model Diagnostics and Sensitivity Analyses	22
3.9.8 Analytical Considerations	23
3.10 Data Quality Assessment	23
3.11 Ethical and Governance Considerations.....	24
4 Results	25
4.1 Study Population and Eligibility Attrition	25
4.1.1 Analytical Cohorts and Treatment Groups Represented	25
4.1.2 Eligibility Attrition.....	26
4.1.3 Final Analytical Cohort by Site and Treatment.....	27
4.2 Characteristics of the Study Population	27

4.2.1 Demographic Characteristics	27
4.2.2 <i>Clinical characteristics and prior treatment</i>	28
4.2.3 Data Availability and Completeness	31
4.3 Overall Survival (Primary Objective)	31
4.3.1 Overall Survival	32
4.3.2 Age-stratified Overall Survival	35
4.3.3 Sensitivity Analyses	36
4.4 Treatment Patterns and Treatment Sequences (Secondary Objective)	36
4.4.1 First-Line Treatment Distribution	36
4.4.2 Subsequent Lines of Therapy	36
4.4.3 Treatment Duration	37
4.5 Other Clinical Outcomes: TTNT and TTD (Secondary Objective)	38
4.5.1 Time to Next Treatment or Death (TTNT)	38
4.5.2 Time to Treatment Discontinuation or Death (TTD)	41
4.6 Exploratory Prognostic Analyses (Exploratory Objective)	44
4.7 Infrastructure Feasibility Findings (Infrastructure Objective)	45
4.7.1 Infrastructure Implementation	45
4.7.2 Variable Availability in OMOP Formatted Data	45
4.7.3 Cohort Implementation Considerations	46
5 Discussion	47
5.1 Main Findings	47
5.2 Primary Objective: Overall Survival	47
5.3 Secondary Objectives: Treatment Patterns, TTNT, and TTD	48
5.3.1 Treatment Patterns	48
5.4 Exploratory Objective: Prognostic Analyses	50
5.5 Limitations	51
5.5.1 Cohort Construction and Attrition	51
5.5.2 Cross-Country Interpretability	52
5.5.3 Observational Design and Residual Confounding	52
5.5.4 Missingness, Suppression, and Biomarker Limitations	53
6 Conclusions	54
6.1 Future Research	54
References (need to be updated in 2 nd version)	55
Appendices	57
Appendix A. Study population and baseline characteristics	58
6.2 Appendix B. Country-specific Treatment-pattern Tables	65
6.3 Appendix C. Site-specific Survival Analyses	67
6.4 Appendix D. Model Diagnostics and Assumptions	77
Statistical Annex	80
Authors	81
7 Statistical annex	85
7.1 Treatment Patterns	85
7.1.1 Next Line-of-Therapy	85
7.1.2 Treatment Discontinuation	86
7.2 Time-to-Event Analysis	87
7.3 Average Treatment-Effect	89
7.3.1 Inverse Probability of Treatment Weighting	89
7.4 Sensitivity Analysis: Inverse Probability of Censoring Weighting	90
7.5 Handling Missing-Data	91
7.5.1 Descriptive Analyses	91

7.5.2 Primary Modelling Approach	91
7.5.3 Assessment of Missing-Data Mechanisms	92
7.5.4 Multiple-Imputation	92
7.5.5 Missing-Data in Propensity Score and Weighted Analyses	92
7.5.6 Variables Not Subject to Imputation	92
7.5.7 Exploratory Analyses for Missing Not at Random (MNAR)	93
7.5.8 Partial and Missing Dates	93
7.6 Pooled Analysis	94
7.6.1 Meta-analysis Framework	94
7.6.2 Heterogeneity Assessment.....	94
7.7 References.....	95

7 Statistical annex

7.1 Treatment patterns

7.1.1 Next line of therapy

The time to next line of therapy (TTNT) as a composite endpoint is defined as the time from index (i.e., start of 1L therapy) to the initiation date of the next line of anti-cancer therapy to which the patient had **no prior exposure** or death from any cause. A next LOT is triggered if the earliest of the following is observed in the treatment episode and not previously observed in the treatment history:

1. Switching to a new monotherapy ICI (e.g., switching from pembrolizumab → atezolizumab) or combination, not previously used. Note that the first ICI should not be re-administered and the start date for the next LOT (i.e., the switch date) will trigger the TTNT event date (see **Patient 1** in Figure 1).
 - a. **Patient 1** initiates therapy B **>60 days** of the last administration of therapy A, provided therapy A is not re-administered after initiation of therapy B. Therapy B initiated a new LOT.
2. Start chemotherapy (not included in Figure)
3. Add-on another therapy
 - a. To distinguish between a brief delay in dispensation pattern and an add-on or change in LOT, a minimum duration of **3 weeks (21 days)** on the prior therapy, therapy A, will be required (see **Patient 2** in **Error! Reference source not found.**). Accordingly, initiation of therapy B for **Patient 2** will be classified as a new LOT.
 - b. For **Patient 3** in Figure 1, the initiation of therapy B will be categorised as part of therapy A, rather than being considered a new LOT.

Patients without a qualifying next LOT will be censored at the earliest of last documented contact date (i.e., last known date alive date or loss to follow-up) or end of the study period.

TTNT will be calculated as: $TTNT = (Event\ date - Index\ date) + 1$.

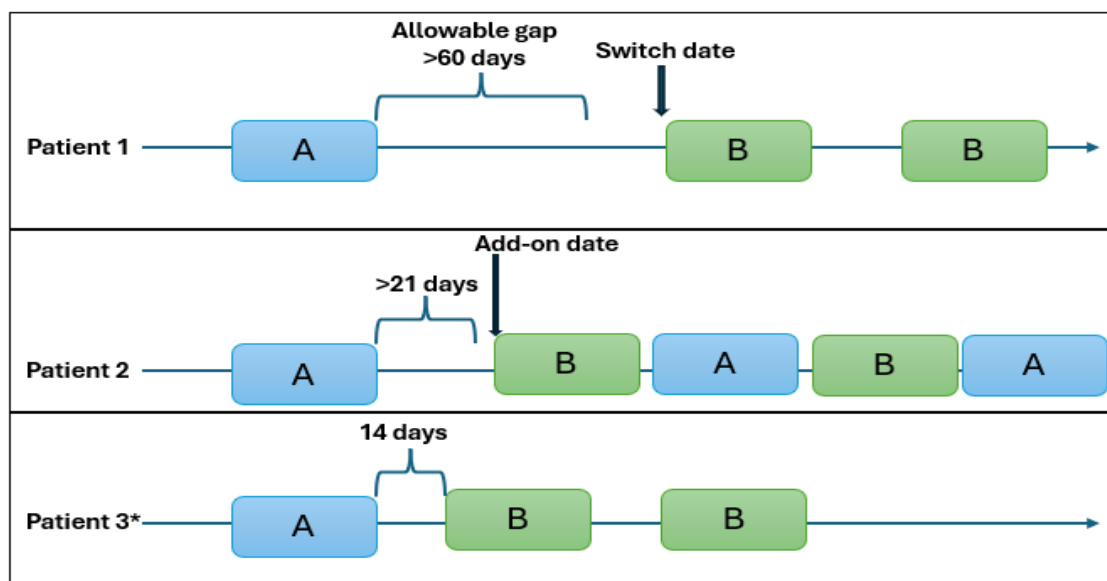


Figure 1: Schematic of Treatment Switching and Add-on Rules to Define Lines of Therapy

7.1.2 Treatment discontinuation

The time to treatment discontinuation (TTD) as a composite endpoint is defined as the time from treatment initiation to the earliest of:

1. Permanent discontinuation of index monotherapy (last administration date + allowable window of 60 days, see Figure)
2. Switch to another anti-cancer therapy
3. Add-on therapy
4. Death from any cause

Patients without TTD event will be censored at the last date known to be receiving treatment, last documented contact date or at the end of the study period. TTD will be calculated as: $TTD = (Event\ date - Index\ date) + 1$.

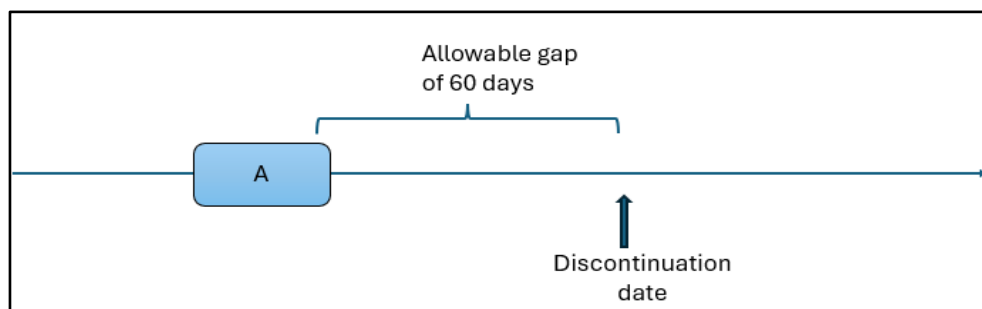


Figure 2: Discontinuation - occurs when the administration date + 60-day allowable gap pass without another therapy being administered or the first therapy being re-administered.

7.2 Time to event analysis

For time-to-event analyses without competing risks, outcomes will be summarised using the Kaplan–Meier (KM) method, and covariate effects will be assessed using Cox proportional hazards (PH) models. For endpoints such as TTNT and TTD, for which death or new malignancies may act as competing risks, cumulative-incidence functions will be estimated using the Aalen–Johansen estimator. In accordance with best practice [1], Fine–Gray subdistribution hazard models will be used to estimate relative covariate effects on the cumulative-incidence, as this approach directly targets the probability of experiencing the event of interest in the presence of competing risks, whereas KM and cause-specific Cox methods may overestimate event risk when competing events are treated as independent censoring (i.e., remove individuals who experience a competing event from the risk set). Correspondence of analysis method and endpoints can be found in **Table** .

Age-stratified analyses will be performed when ≥ 10 events are observed per category, otherwise presentations will be eliminated

Table 1: Time to event analysis

Endpoint	Primary method	Sensitivity method	Event	Censoring rules
Time to death from any cause (OS)	Kaplan-Meier and Cox PH	IPTW-weighted Kaplan-Meier and Cox PH	Death from any cause	Whichever occurs first: last documented contact date, end of study period
Composite: Time to next line of therapy or time to death (TTNT)	Aalen-Johansen	Fine-Gray	Initiation of a next line of therapy or death from any cause	Whichever occurs first: last documented contact date, end of study period
Composite: Time to treatment discontinuation or time to death (TTD)	Aalen-Johansen	Fine-Gray and IPCW-weighted Cox	Index monotherapy discontinuation or death from any cause	Whichever occurs first: last documented contact date, end of study period if ongoing treatment

When fitting the Cox PH model with multiple predictor variables, the following will be evaluated:

1. Model fit using the Cox-Snell residuals plot. Should the plot indicate a poor fit, some covariates of the model may have to be transformed or removed. This decision will be made with the study team.
2. PH assumption using the Schoenfeld residuals test and Schoenfeld residuals plots (for continuous covariates).
 - a. If the PH assumption holds, then no trend would be expected from the residuals plot against time (one plot for each covariate included).
 - b. If the PH assumption does not hold, there are options for improving the non-proportionality in the model. For example (but not limited to), excluding the problematic covariate from the model (if not of inferential interest), including new covariate(s), adding non-linear terms for existing covariates, adding interactions with time among covariates or stratification on problematic variable.
3. Influential observations using the deviance residuals plot. These observations will be checked for accuracy. If accurate, a decision will be made in conjunction with the study team.
4. Multicollinearity using the correlation matrix and variance inflation factor greater than 5.
5. The 1 in 10 rule of thumb that approximately includes one candidate predictor in the model for every 10 events. If low correlation exists between the covariates in Step 4, then this criterion may be relaxed slightly.

7.3 Average treatment effect

The primary analysis will focus on estimating the average treatment-effect (ATE) for pairwise comparison and will use inverse probability of treatment weighting (IPTW) to generate propensity scores (PS) for each patient in the study [2]. The following causal methods will be applied only if baseline imbalances or censoring patterns threaten validity of conventional regression models.

7.3.1 Inverse probability of treatment weighting

The aim of the IPTW approach is to control for potential confounding by constructing a pseudo-population, in which the intention-to-treat (ITT) exposure variable, i.e., initiation of 1L monotherapy pembrolizumab or atezolizumab, or cemiplimab, is independent of selected measured confounders (listed below).

The following algorithm will be employed to generate and assess PS weights:

1. The generalised PS, that is, $e_j(\mathbf{X}) = P(Z = j | \mathbf{X})$, will be estimated using a multinomial logistic regression model, where $j \in \{1, 2, 3\}$ is the treatment/exposure assignment indicator and $\mathbf{X}$ is the vector of baseline covariates. In this model, the ITT exposure variable (i.e. study treatment of interest) will be treated as the outcome variable. PS will be computed at the index date as the probability to receive 1L monotherapy pembrolizumab, atezolizumab, or cemiplimab conditional on observed baseline covariates as fixed effect. Variables included in the PS model will include:
 - Age at index
 - Sex
 - ECOG-PS
 - Charlson's CI
 - Smoking status
 - Height/weight
 - Stage at diagnosis
 - Prior SACT
 - Prior cancer-related surgery
 - Prior radiotherapy
 - Index year
 - NSCLC Histology (squamous, non-squamous)
 - Lung Immune Prognostic Index (LIPI)
2. For each patient i , non-stabilised PS weights will be calculated as the multiplicative inverse of the PS, that is, $w_i = 1/e_{Z_i}(\mathbf{X}_i)$, where e_{Z_i} denotes the estimated PS for patient i receiving the actual treatment $Z_i \in \{1, 2, 3\}$.
3. Stabilised weights will be calculated using the non-stabilised PS weights derived in Step 2, by multiplying the non-stabilised weights by the marginal probability of receiving the actual treatment assignment (i.e., without considering covariates). That is, $w_i^s = P(Z = Z_i)/e_{Z_i}(\mathbf{X}_i)$ [3].
4. To assess the positivity assumption (i.e., each patient has a nonzero probability of receiving each treatment), the stabilised weight distribution will be described by outputting numeric summary (i.e., min, max, mean and various percentiles). Note that the mean of stabilised weights ≈ 1 . Extreme weights will be trimmed at the 99th percentile after assessment with the project team.
5. To assess the balance of the generalised PS model, group vs combined standardised mean differences (SMDs) [4, 5] and overlapping density plots will be displayed. Balance will be

deemed acceptable when, for each covariate, the maximum group vs combined SMD across the three exposure groups is < 0.10 . Any covariate with $SMD \geq 0.1$ will be adjusted for in the outcome model. As a secondary check, pairwise SMDs (see details below) will be reported [6].

a. Group vs Combined SMD

For continuous covariate X and exposure group $j \in \{1, 2, 3\}$:

$$SMD_{j,comb}(X) = \frac{|\bar{X}_j^w - \bar{X}_{comb}^w|}{s_{comb}^w}$$

Where $\bar{X}_{comb}^w$ and s_{comb}^w are weighted mean and standard difference (SD) from the combined population. Note that it is required that all $SMD_{j,comb}(X) < 0.10$ to check whether each exposure group is balanced to the combined population created by IPTW [4, 5].

For categorical covariates, SMDs will be computed for each level's proportion [4].

b. Pairwise SMDs (secondary check)

For exposure groups g and h ,

$$SMD_{g,h}(X) = \frac{|\bar{X}_g^w - \bar{X}_h^w|}{\sqrt{(s_g^{2(w)} + s_h^{2(w)})/2}}$$

Where $\bar{X}_g^w$ and $s_g^{2(w)}$ are the weighted mean and variance from exposure group g . To check balance across pairs, $\max\{SMD_{1,2}, SMD_{1,3}, SMD_{2,3}\} < 0.10$ [6].

This approach will be implemented, if necessary, when modelling is performed, as described in section **Error! Reference source not found.**

7.4 Sensitivity analysis: inverse probability of censoring weighting

A sensitivity analysis may be performed for the time to treatment discontinuation (TTD) composite endpoint assuming potential informative censoring (i.e., factors that may influence TTD) to be modelled by Inverse Probability of Censoring Weighting (IPCW) within the marginal structural model (MSM) framework [7, 8]. A similar IPCW-based MSM sensitivity analysis may be conducted for TTNT, pending adequate event counts. IPCW estimators will correct for informative censoring by allocating additional weights to patients who are not censored. The following steps will be used to calculate the censoring weights using the approach described by Cole and Hernán [7]:

1. Fit a Cox proportional hazard model treating censoring (i.e., administrative censoring or last documented contact date) as the event of interest with the time origin at treatment start. Patients who experience discontinuation or death will be treated as censored in this censoring model [9]. The list of covariates will include baseline predictors included in Step 1 of the IPTW PS model above. If proportional hazards for the censoring model are violated, flexible time terms (e.g., natural splines) or covariate–time interactions will be considered to improve model fit, subject to discussion with the study team.
2. For each patient i at time t , estimate the probability of remaining uncensored up to time t using the Cox PH model fitted in Step 1, by calculating predicted survival probabilities that incorporate both baseline and any time-varying covariates. The equation of IPCW is as follow:

$$w_i^{IPCW}(t) = \hat{S}_{uncen}(t|X_i) = \frac{1}{\exp(\lambda_i(t))}$$

3. The weights derived in Step 2 can become very large, which could lead to instability in parameter estimation. A stabilised weight will be used by incorporating a term to the numerator of each $w_i^{IPCW}(t)$. That is, the unconditional probability of a patient i remaining uncensored at their observed time t (i.e., fitting a Cox PH model that includes only baseline covariates, excluding possible time-varying terms). The numerator term will be denoted by $\hat{S}_{uncen}(t|X_i^{baseline})$ and the final stabilised weights will be calculated as:

$$sw_i^{IPCW}(t) = \frac{\hat{S}_{uncen}(t|X_i^{baseline})}{\hat{S}_{uncen}(t|X_i)}$$

Note that the distribution of the stabilised weights will be inspected for extreme weights and if detected, truncation at the 99th percentile may be applied after consultation with the study team. Also verify that the mean stabilised IPCW is approximately 1.

4. Let sw_i^{IPTW} be the stabilised IPTW from the multinomial model as described in Section 7.3.1 Inverse Propability of Treatment Weighing and $sw_i^{IPCW}(t)$ the stabilised IPCW in Step 3, the final analysis weight to be used in all weighted survival estimators (i.e., KM and Cox PH) for TTD is:

$$w_i(t) = sw_i^{IPTW} \times sw_i^{IPCW}(t) \text{ [7]}$$

7.5 Handling missing data

This study is based on routinely collected real-world data from multiple Nordic sites, mapped to the OMOP Common Data Model. The extent and pattern of missing data are expected to vary across sites and variables. Missing data will be handled using a principled, endpoint-specific approach that prioritises transparency, avoids unnecessary modelling assumptions, and aligns with the study's observational design.

No imputation will be applied to **outcome variables** or to **exposure assignment**.

7.5.1 Descriptive analyses

For all baseline covariates and clinical characteristics:

- Missing data will be summarised explicitly as counts and percentages.
- A separate “Missing” category will be reported for categorical variables, where appropriate.
- All summaries will clearly state the denominator used (total cohort vs. non-missing observations).

7.5.2 Primary modelling approach

Complete-Case Analysis

For primary regression analyses:

- Complete-case analysis will be used when the proportion of missing data for all covariates included in each model is <5% [10].
- This approach is considered appropriate under assumptions of Missing Completely at Random (MCAR) or low-level (i.e., ≤40%) Missing at Random (MAR), given minimal information loss [11].

7.5.3 Assessment of missing data mechanisms

Prior to any imputation, the extent and structure of missing data will be evaluated using:

- Descriptive summaries of missingness by site, exposure group, and outcome.
- Exploratory analyses to evaluate associations between missingness indicators and observed covariates.
- Where feasible, Little's MCAR test [12] to assess whether data are Missing Completely at Random.

7.5.4 Multiple imputation

When to use

Multiple imputation will be considered only when all the following conditions are met:

1. A covariate is deemed clinically important for confounding adjustment.
2. Missingness exceeds **5%** but is not excessive (generally **≤40%**).
3. Data are plausibly Missing at Random (MAR), that is, missing values may be conditional on observed variables.

Covariates with >40% missingness will not be imputed and will be excluded from multivariable adjustment models.

Imputation method

When applied, missing covariate data will be handled using **Multiple Imputation by Chained Equations (MICE)**:

The imputation model will include:

- All covariates used in the analysis model.
- The exposure variable.
- Outcome indicators (without imputing the outcome itself).

The number of imputations will approximately match the percentage of incomplete cases (e.g., 20 imputations for ~20% missingness) [13].

Imputation will be performed within site, consistent with the federated analysis framework.

Estimates across imputed datasets will be combined using **Rubin's rules** [14].

7.5.5 Missing data in propensity-score and weighted analyses

For analyses using inverse probability weighting (IPTW, IPCW):

- Missingness in baseline covariates required for propensity-score estimation will be addressed using:
 - Multiple imputation (as described above), or
 - Complete-case analysis if missingness is <5%.

Balance diagnostics (standardised mean differences) will be performed post-imputation and post-weighting to confirm adequacy of confounding control.

7.5.6 Variables Not Subject to Imputation

The following variables will not be imputed under any circumstances:

- Exposure variables (type of 1L ICI monotherapy).
- Time-to-event outcomes (OS, TTNT, TTD).
- Event indicators (death, treatment discontinuation, next line of therapy).

Patients with missing outcome data will be censored according to definitions used in analysis.

7.5.7 Exploratory analyses for Missing Not at Random (MNAR)

If there is concern that missing data are Missing Not at Random (MNAR) and multiple imputation is not appropriate:

- Sensitivity analyses will be conducted using **extreme-value** or **tipping-point scenarios**, where missing values are imputed with clinically plausible upper and lower bounds upon consultation with the study team.

Results will be compared to the primary analysis to assess robustness.

7.5.8 Partial and missing dates

General imputation rules

The following general rules for missing or partial dates will be applied:

1. If a date is missing the day portion (e.g., July 2015), the incomplete date will be imputed by using the fifteenth day of the month (e.g., 15 July 2015).
2. If a date is missing the month portion (e.g., 30 2015), the incomplete date will be imputed as the month of June (e.g., 30 June 2015).
3. If a date is missing the day and month portion (e.g., 2015), the incomplete date will be imputed as the thirtieth day of June (e.g., 30 June 2015).
4. For all imputation rules above, an imputed date cannot conflict with any known recorded date.
 - a. For example, if a missing start date is imputed, a lower limit of the patient's birth date (if available) will be used, such that if the imputed date is before the patient's birth date, the imputed date will be replaced with the patient's birth date.
 - b. Similarly, for end dates, if the imputed date is after the patient's death date, the imputed date will be replaced with the patient's death date.
5. No imputation will be performed for a date where day, month and year are missing.
6. In case there is a conflict between start and end dates after the imputation rules are applied (e.g., same start and end date) the dates will be set to be missing.

Specific imputation rules for time to event

The following imputation rule for partial or missing event dates will be used for time-to-event analysis:

1. If year and month are available, but day is missing, the date is set to the fifteenth day of the month.
2. If the month is missing or the date is completely missing, the TTE will not be calculated.

The following imputation rule for partial or missing censoring dates will be used for TTE analysis:

1. For partial or missing censoring dates the analogous rule applies, with the censoring date needing to have at least the month and year available.
2. Otherwise, the last available date before the missing censoring date is imputed as the censoring date.

Imputation rules for partial or missing start dates for TTE variables:

1. If the start day for the calculation is missing, this day will be set to the 15th day of the month.
2. If the month is missing or the date is completely missing, the TTE will not be calculated.

7.6 Pooled analysis

The primary analyses will be performed separately within each country and pooled effect estimates will be generated to improve statistical precision and allow inference across the Nordic region. Country-specific effect estimates will be produced for OS, TTD, TTNT.

7.6.1 Meta-analysis framework

Pooled estimates will be generated using meta-analysis of country-specific effect estimates, rather than pooling patient-level data directly, to respect data localisation constraints.

For each endpoint, country-specific KM estimates, hazard ratios (HRs) and corresponding standard errors will be combined using:

- Fixed-effects meta-analysis, assuming a common underlying treatment effect across countries, or
- Random-effects meta-analysis (DerSimonian–Laird or restricted maximum likelihood), if between-country heterogeneity is observed.

The choice between fixed- and random-effects models will be guided by heterogeneity assessments described below.

7.6.2 Heterogeneity assessment

Prior to conducting the pooled analysis, heterogeneity across the study countries will be assessed using:

- Cochran’s Q statistic
- I² statistic and its 95% CI, representing the proportion of total variability attributable to heterogeneity rather than sampling error.

Consistent with established guidance, I² values of approximately 25%, 50%, and 75% will be interpreted as indicating low, moderate, and high heterogeneity, respectively [15]. If substantial heterogeneity is identified (e.g., I² > 75%), pooled estimates will be interpreted cautiously, and country-specific results will be emphasised.

7.7 References

1. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol*. 2012;41(3):861–870. doi:10.1093/ije/dyr213.
2. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015 Dec 10;34(28):3661–79.
3. Pezzi A, Cavo M, Biggeri A, et al. Inverse probability weighting to estimate causal effect of a singular phase in a multiphase randomized clinical trial for multiple myeloma. *BMC Med Res Methodol*. 2016;16:150. doi: 10.1186/s12874-016-0253-9
4. Li F, Morgan KL, Zaslavsky AM. Balancing Covariates via Propensity Score Weighting. *JASA*. 2018;113(521):390–400.
5. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*. 2011;46:399–424.
6. McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Stat Med*. 2013;32:3388–3414. <https://doi.org/10.1002/sim.5753>
7. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168:656–664. doi:10.1093/aje/kwn164
8. Austin PC, Fine JP. Inverse Probability of Treatment Weighting Using the Propensity Score With Competing Risks in Survival Analysis. *Stat Med*. 2025 Feb 28;44(5):e70009. doi: 10.1002/sim.70009. PMID: 39915951; PMCID: PMC11803134.
9. Blumenthal GM, et al. Analysis of time-to-treatment discontinuation of targeted therapy, immunotherapy, and chemotherapy in clinical trials of patients with non–small-cell lung cancer. *Ann Oncol*. 2019;30:830–838.
10. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis [Internet]. New York, NY: Springer; 2001 [cited 2023 Aug 8]. (Springer Series in Statistics). Available from: <http://link.springer.com/10.1007/978-1-4757-3462-1>
11. Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York: John Wiley & Sons; 1987. 304 p.
12. Little RJA. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*. 1988;83(404):1198–202.
13. Von Hippel PT. How to Impute Interactions, Squares, and other Transformed Variables. *Sociological Methodology*. 2009;39(1):265–91.
14. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Medical Research Methodology*. 2009;9(1):57.
15. Higgins, J.P., S.G. Thompson, J.J. Deeks, and D.G. Altman, Measuring inconsistency in meta-analyses. *Bmj*, 2003. 327(7414): p. 557–560