1. Title Page

Title	TARGET-EU: Nivolumab plus ipilimumab with chemotherapy versus pembrolizumab with chemotherapy in patients with non-oncogenic metastatic non-small-cell lung cancer with <1% PD-L1 tumour expression
Research question & Objectives	The objective of this study is to assess overall survival of nivolumab plus ipilimumab with chemotherapy versus pembrolizumab with chemotherapy in patients with non-oncogenic metastatic NSCLC and <1% PD-L1 tumour expression
Protocol version	V1.0
Last update date	27 February 2026
Contributors	**Principal investigators:** Dr. Corina J.G. van den Hurk, corina.vd.hurk@catharinaziekenhuis.nl Dr. Ewoudt M.W. van de Garde **Contributor names:** Dr. Ronald Damhuis Prof. Ian Douglas Prof. Olaf H. Klungel Dr. Daniala L. Weir
Study registration	**Site:** https://catalogues.ema.europa.eu/node/4440/administrative-details **Identifier:** EUPAS1000000539
Sponsor	**Organization:** EU PE&PV research network **Contact:** eupepv@uu.nl
Conflict of interest	n/a

Table of contents

1. Title Page 1
2. Abstract 3
3. Amendments and updates 3
4. Milestones 4
5. Background 4
6. Research questions and objectives 5
    6.1. Primary Estimand 1 5
    6.2. Supplementary Estimand 2 6
7. Research methods 7
    7.1. Study design 7
    7.2. Study design diagram 8
    7.3 Setting 8
        7.3.1 Definition of time 0 (and other primary time anchors) for entry to the study population 9
        7.3.2 Study inclusion criteria 9
        7.3.3 Study exclusion criteria 11
    7.4 Variables 12
        7.4.1 Exposure(s) of interest 12
        7.4.2 Outcome(s) of interest 13
        7.4.3 Follow up 13
        7.4.4 Covariates (confounding variables and effect modifiers, e.g. risk factors, comorbidities, comedications) 14
    7.5 Core Emulation Table - Design Summary 16
    7.6 Data analysis 18
        7.6.1 Analysis plan 18
        7.6.2 Primary (Estimand 1) Analysis 19
        7.6.3 Supplemental (Estimand 2) Analysis 22
        7.6.4 Sensitivity analyses 24
        7.6.5 Other supplemental analyses 25
        7.6.6 Core Emulation Table -Estimation Summary 26
    7.7 Data sources 27
        7.7.1 Data source 27

7.8. Data management........31
7.9. Quality control........34
7.10. Study size........41
8. Limitation of the methods........42
9. Protection of human subjects........44
10. Reporting of adverse events........45
11. References........45

2. Abstract

Background: Indirect comparisons of clinical trial data suggest that in patients with non-oncogenic metastatic NSCLC and with <1% PD-L1 tumour expression, first line nivolumab plus ipilimumab with chemotherapy is associated with better survival compared to pembrolizumab with chemotherapy. However, head-to-head comparative trials are lacking. This study is a comparative effectiveness assessment of overall survival from these two therapeutic options in the Dutch routine care clinical setting.

Objective: The objective of this study is to compare overall survival of nivolumab plus ipilimumab with chemotherapy versus pembrolizumab with chemotherapy in patients with non-oncogenic metastatic NSCLC and <1% PD-L1 tumour expression

Methods: This study will be conducted using real-world data from the Dutch national cancer registry and include all patients who started either treatment between 1-1-2022 and 31-12-2024. The primary estimand will be estimated using an inversed probability weighted Cox proportional hazard regression with treatment group as a single covariate and using inverse probability of treatment (IPTW) weighting based on propensity score estimation. The secondary estimand will be estimated using an accelerated failure time model to estimate restricted mean survival time at 1, 2 and 3 years after start of treatment using a Weibull distribution. A sensitivity analysis will be conducted to assess the potential impact of unmeasured confounding from corticosteroids use at baseline.

3. Amendments and updates

Version date	Version number	Section of protocol	Amendment or update	Reason
27 February 2026	1.0			

4. Milestones

Table 1. Milestones

Milestone	Date
Study protocol for RWD study	8 August 2025
Preliminary results RWD study	April 2026
Final Study report	10 June 2026

5. Background

What is known about the condition: In Europe, lung cancer is the second most diagnosed type of cancer with 319,000 (12%) newly diagnosed patients in 2022. In that year, it was the most common cause of cancer death (20% of all cancer deaths) (Lung cancer prevention | Knowledge for policy, Sept 2025). Non-small-cell lung cancer (NSCLC) was the most common type, representing 85% of all lung cancers (Types of lung cancer - Lung Cancer Europe, Sept 2025).

What is known about the exposure of interest: In the Netherlands, among patients with NSCLC, about 50% are diagnosed with stage IV disease (iknl.nl, Dec 2024). Historically, treatment with chemotherapy was the standard treatment for these patients. Second-line nivolumab for NSCLC was introduced in the Netherlands in 2016. In 2017, first-line pembrolizumab for NSCLC with PD-L1 ≥50% was introduced. For patients with advanced NSCLC with lower PD-L1 expression (<50%), pembrolizumab plus chemotherapy became reimbursed as first-line treatment from 2018, based on the results from the Keynote 189 trial for non-squamous histologies (2018) and the Keynote 407 trial for squamous histology (2019) (Gandhi 2018 + Paz-Ares et al. 2018).

In 2021, the CheckMate 9LA study demonstrated that dual immunotherapy with nivolumab and ipilimumab in combination with two cycles of chemotherapy also resulted in longer overall survival (OS) compared to chemotherapy alone (Paz-Ares et al. 2021). Six-year follow-up data of the 9LA study recently confirmed a sustained OS benefit of dual immunotherapy, with the survival benefit being most pronounced in patients with no PD-L1 tumour expression (20% versus 7% alive at 6 years) compared to with PD-L1 expression (15% versus 10%) (Carbone et al. 2025).

In the Netherlands, experts in the field have suggested preference for nivolumab plus ipilimumab with two cycles of platinum-doublet in patients with good performance status, no PD-L1 tumour expression and absence of liver metastasis in 2021 (Smit et al. 2021). Since 2022, treatment with nivolumab plus ipilimumab has also become available to patients with metastatic lung cancer through reimbursement within the national care system.

A recent observational study (Verschueren et al. 2025) showed that this option indeed led to a slightly better overall response rate compared to pembrolizumab with chemotherapy.

Gaps in knowledge: at present, it is unknown if nivolumab plus ipilimumab with chemotherapy leads to improved OS over pembrolizumab with chemotherapy within the Dutch routine care clinical setting in patients with no PD-L1 tumour expression.

What is the expected contribution of this study? The findings from the present study will provide a comparative effectiveness assessment of overall survival from first line nivolumab plus ipilimumab versus pembrolizumab (both combined with induction chemotherapy) in metastatic non-oncogenic NSCLC patients with PD-L1 expression <1% within the Dutch routine care clinical setting.

6. Research questions and objectives

The overall aim of this study is to compare overall survival between patients with primary non-oncogenic metastatic NSCLC treated with first line nivolumab plus ipilimumab with chemotherapy or first line pembrolizumab with chemotherapy.

6.1. Primary Estimand 1

The research question answered by the estimand

What is the hazard ratio (HR) of death from any cause in patients with metastatic non-oncogenic NSCLC with a PD-L1 expression of <1% receiving dual immunotherapy (nivolumab + ipilimumab)+ chemotherapy versus pembrolizumab + chemotherapy as first line of therapy, regardless of treatment discontinuation or treatment switch?

Table 2. Estimand 1

Attribute	Hypothetical Target Trial	Target Trial Emulation	Comment
Population	Patients with advanced (stage IIIb/IV) non-oncogenic NSCLC whose tumors have a PD-L1 expression of <1%	Patients with metastatic (stage IV) non-oncogenic NSCLC whose tumors have a PD-L1 expression of <1%	Stage IIIb patients were foreseen in the Hypothetical Target Trial (HTT) because the treatments under study are sometimes offered to patients with stage IIIb disease as well (off-label). Stage IIIb has a better prognosis than stage IV disease and requires a stratified analysis. In hindsight, incorporating that stratification has been overseen in the HTT design and because of knowing that the percentage of stage IIIb will be less than 10% in the study population

			together with the foreseen limited total sample size (n=200, and thus very few stage IIIb patients n=~20) we propose to focus the Target Trial Emulation (TTE) on stage IV solely. By doing so we reach a clinically relevant homogeneous population and also match with the reference pivotal 9LA clinical trial eligibility (stage IV only).
Treatment Conditions	Experimental group: nivolumab (360 mg IV/3 wk) + ipilimumab (1 mg/kg IV/ 6 wk) + 2x histology-based, platinum doublet chemotherapy (IV/ 3wk) Control group: pembrolizumab + 4x platinum doublet chemotherapy (IV/3 wk)	Index treatment: nivolumab (any dose) + ipilimumab (any dose) + ≥1x platinum doublet chemotherapy (any dose) Comparator treatment: pembrolizumab (any dose) + ≥1x platinum doublet chemotherapy (any dose)	Data source only has drug names and start dates, no dosages or total duration of treatment. It is reasonably assumed that in routine care no other dosing is used than in the label. Exposure category will be defined based on first observed drug records since date of diagnosis (new-user design). There is no information in the data source about how many cycles patients did receive.
Endpoint	Overall survival	Overall survival	
Summary Measures	Hazard Ratio	Hazard Ratio	
Intercurrent Events and Strategies to Handle Them	Treatment discontinuation: treatment policy Treatment switch: treatment policy	Treatment discontinuation: treatment policy Treatment switch: treatment policy	

*NSCLC: non-small cell lung cancer, ECOG PS: Eastern Cooperative Oncology Group Performance Status, PD-L1: Programmed death-ligand 1

Rationale for why selected strategies to handle intercurrent events are chosen

The strategy to handle intercurrent events is treatment policy. This strategy best reflects the effectiveness of the intervention under study within clinical practice, regardless of treatment discontinuation or treatment switch.

6.2. Supplementary Estimand 2

Same as estimand 1 with exception of the summary measure, which is the ‘Difference in restricted mean survival times (RMST)’.

The research question answered by the estimand:

What is the difference in RMST in patients with metastatic non-oncogenic NSCLC with a PD-L1 expression of <1% receiving dual immunotherapy (nivolumab + ipilimumab)+ chemotherapy versus pembrolizumab + chemotherapy as first line of therapy, regardless of treatment discontinuation or treatment switch?

7. Research methods

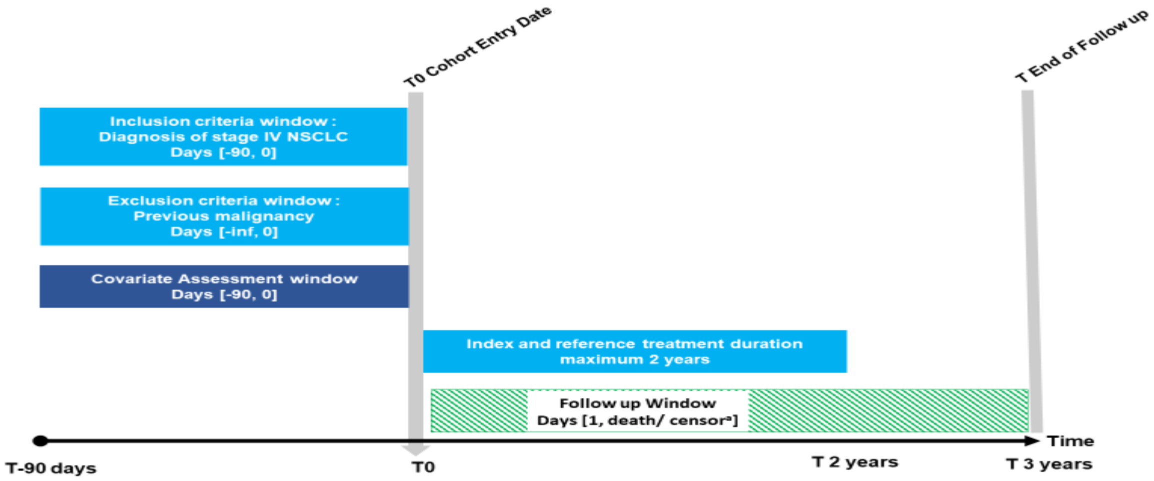
7.1. Study design

Research design (e.g. cohort, case-control, etc.): This an active comparator cohort study to investigate the effectiveness of nivolumab plus ipilimumab with chemotherapy versus pembrolizumab with chemotherapy given as first line therapy in patients with metastatic non-oncogenic NSCLC whose tumour has a PD-L1 expression <1%.

Study design choice: The active comparator design reduces risk of bias from unmeasured confounding by indication because both treatment options have the same indication (label) and can be applied for patients with metastatic non-oncogenic NSCLC.

7.2. Study design diagram

Figure 1. Study design diagram



T0 Cohort Entry Date= Initiation of index ((ipilimumab + nivolumab) or reference ((pembrolizumab) treatment
T End of treatment= maximum treatment window of 2 years

[a] Censoring criterium: end of study period

NSCLC = Non-small cell lung cancer

7.3 Setting

This study is conducted using routinely collected cancer registry data, retrieved from electronic health records (EHR) in the period 2022 – 2024. This time frame reflects the period of dual immunotherapy use in routine clinical practice in the Netherlands. The study is set in outpatient hospital care. Data are sourced from all patients diagnosed with cancer in the Netherlands, providing population-based and representative coverage of real-world clinical care.

7.3.1 Definition of time 0 (and other primary time anchors) for entry to the study population

The time 0 (T0) is the date of start of either the index or reference treatment. Incidence is assessed based on any immunotherapy record (ipilimumab + nivolumab, or pembrolizumab). Chemotherapy maximum 28 days before T0 is allowed because in clinical practice one course of chemotherapy is sometimes administered while awaiting the results of molecular diagnostics. The date of diagnosis of NSCLC should be maximum 90 days before T0 to ensure that patients will not have received any additional treatments such as radiotherapy before initiation of the treatments under study. A previous study has shown that most patients start first line systemic therapy within 90 days from diagnosis (Klarenbeek et al. 2023). All eligibility criteria (including ECOG and PD-L1 status) should be fulfilled at index date, thus measured at or before index date. The latter also holds for all other covariates.

Table 3. Operational Definition of Time 0 (index date) and other primary time anchors

Study population name(s)	Time Anchor Description (e.g. time 0)	Number of entries	Type of entry	Washout window	Care Setting	Code Type	Diagnosis position	Incident with respect to...	Measurement characteristics/ validation	Source of algorithm
Adult patients with non-oncogenic metastatic NSCLC and PD-L1 tumour expression <1% who started any of the chemo-immunotherapies of interest as first line treatment	Date of initiation of index or reference immunotherapy treatment component (T0)	Single	Incident	n/a	OP	RxNorm	n/a	Studied immunotherapies (assessed indirectly by excluding patients with past malignancy)	n/a	n/a

OP = outpatient, n/a = not applicable, RxNorm = standardized, normalized nomenclature for clinical drugs developed by the National Library of Medicine

7.3.2 Study inclusion criteria

The inclusion criteria (Table 4) will lead to a sample of patients with a specific indication (metastatic NSCLC with a PD-L1 tumour expression <1% and with good performance status (ECOG 0-1)) who started one of the two treatment options under study within a specific time period. The time period for inclusion (1-1-2022 to 31-12-2024) is based on the availability of the index treatment (reimbursed in the Netherlands from 2022) and to ensure a theoretical minimum one year and one month time of data availability for every patient to identify outcomes with the cut-off point at 1-2-2026 (data analyses planned Q1 2026). The treatment is started within 90 days from diagnosis, which matches routine care settings wherein >95% of patients with stage IV lung cancer start systemic treatment, if any (Klarenbeek et al. 2022).

In the NCR database, the ECOG performance status is missing in 15% of the subjects. Subjects with missing ECOG cannot be assessed for eligibility, and will be excluded from the study. However, for the medicines investigated in this study, ECOG information is needed to reimbursement purposes, and consequently we expect no or minimal exclusion of subjects due to missing ECOG information. Furthermore, there is no reason to expect that this will be different for the index or reference treatment, because both treatments share similar treatment indications (and thus patients characteristics and clinical work-up) thus will not introduce a risk of selection bias.

Table 4. Operational Definitions of Inclusion Criteria

Criterion	Details	Assessment window	Care Settings[1]	Code Type[2]	Diagnosis position[3]	Applied to study populations:	Measurement characteristics/validation	Source for algorithm
Lung cancer	Patients with a first diagnosis date of primary, histologically confirmed, diagnosis of non-small cell lung cancer	[-90,0]	OP	SNOMED	n/a	All	n/a	n/a
Stage IV	Defined by NCR data manager from EHR documentation	[-90, 0]	OP	TNM	n/a	All	n/a	n/a
PD-L1<1%	Lab value	[-90, 0]	OP	n/a	n/a	All		n/a
Treatment	Start of index or reference drugs (immunotherapy components: nivolumab, ipilimumab, or pembrolizumab) between 1-1-2022 and 31-12-12024	[0,0]	OP	RxNorm	n/a	All	n/a	n/a
Start chemotherapy component	Ensuring that the immunotherapy is combined with chemotherapy in the same treatment line	[-28, -0]	OP	RxNorm	n/a	All	n/a	n/a

Age ≥ 18	Defined as date time 0 – birth date	[0, 0]	OP	n/a	n/a	All	n/a	n/a
ECOG PS ≤ 1	Defined by NCR data manager from EHR documentation; In case of combination score (ECOG 0-1 or 1-2): rounded down	[-90, 0]	OP	n/a	n/a	All		n/a
No targetable oncogenic mutations or translocations#	No recorded actionable EGFR, ALK or ROS1 mutation	[-90, 0]	OP	n/a	n/a	All	n/a	n/a

OP = outpatient, n/a = not applicable, SNOMED = Systematized Nomenclature of Medicine, TNM = Classification of Malignant Tumors, NCR = Netherlands Cancer Registry, NSCLC = non-small cell lung cancer, PD-L1 = Programmed death-ligand 1, RxNorm = standardized, normalized nomenclature for clinical drugs developed by the National Library of Medicine, ECOG PS= Eastern Cooperative Oncology Group Performance status

 # because guidelines prevail targeted therapy over immunotherapy it is unlikely that patients who received the index or reference treatment within 90 days from diagnosis have an actionable mutation

7.3.3 Study exclusion criteria

This study only excludes patients with a prior diagnosis of cancer in any organ to rule out that there are patients in the sample that could have used any systemic cancer treatment before.

Table 5. Operational Definitions of Exclusion Criteria

Criterion	Details	Assessment window	Care Settings^1	Code Type^2	Diagnosis position^3	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
Previous malignancy	Any cancer diagnosis at any time before T0	[-inf, 0]	OP	SNOMED	n/a	All	n/a	n/a

OP = outpatient, n/a = not applicable, SNOMED = Systematized Nomenclature of Medicine

The flow diagram of patient selection and exclusions is displayed in Appendix Figure 1.

7.4 Variables

7.4.1 Exposure(s) of interest

Exposure to either index or reference treatment is determined based on a start date of any immunotherapy component (nivolumab plus ipilimumab, or pembrolizumab) in the determined inclusion time window. That immunotherapy is given in combination with chemotherapy is assessed through the existence of a record for any chemotherapy drug within 28 days before date of first immunotherapy drug or on the same date.

Algorithm to define duration of exposure:

Duration of exposure cannot be determined because in the data source only start date is available.

Table 6. Operational Definitions of Exposure

Exposure group name(s)	Details	Washout window	Assessment Window	Care Setting	Code Type	Diagnosis position	Applied to study populations:	Incident with respect to…	Measurement characteristics/ validation	Source of algorithm
Nivolumab + ipilimumab		n/a	[0,0]	OP	RxNorm	n/a	All	Any antineoplastic treatment (via not history of malignancy	n/a	n/a
Pembrolizumab		n/a	[0,0]	OP	RxNorm	n/a	All	Any antineoplastic treatment (via not history of malignancy	n/a	n/a
Chemotherapy		[-inf,-28]	[-28,0]	OP	RxNorm	n/a	All	Chemotherapy can be started up to 28 days before immunothe	n/a	n/a

								rapy (i.e. both exposure groups above)		

OP = outpatient, n/a = not applicable, RxNorm = standardized, normalized nomenclature for clinical drugs developed by the National Library of Medicine

**Both immunotherapies are applied in combination with chemotherapy as chemotherapy is sometimes initiated before immunotherapy while awaiting molecular diagnostics already this is the reason for allowing chemotherapy to have started up to minus 28 days before the exposure under study.*

7.4.2 Outcome(s) of interest

The rationale for choosing overall survival (OS) as endpoint for the study is that this is the most clinically meaningful assessable outcome in oncology drug research.

Table 7. Operational Definitions of Outcome

Outcome name	Details	Primary outcome	Type of outcome	Washout window	Care Settings	Code Type	Diagnosis Position	Applied to study populations:	Measurement characteristics/ validation	Source of algorithm
Overall survival	Overall survival since day 1 after start of treatment under study	Yes	Time-to-event	[-inf,0]	IP and OP	Date of death from any cause	n/a	All	n/a	Available in NCR after annual mortality update from linkage with GBA

IP = inpatient, OP = outpatient, n/a = not applicable, GBA = municipal basic administration, NCR = Netherlands Cancer Registry

7.4.3 Follow up

All patients who started either the index or comparator treatment between 1-1-2022 and 31-12-2024 will have a minimum follow-up time of one year to ensure a minimum time to identify outcomes (analyses are conducted in Q1 2026 and NCR has their annual mortality update on February 1).

Table 8. Operational Definitions of Follow Up

Follow up start	Day 1 following treatment initiation	
Follow up end[1]	**Select all that apply**	**Specify**
Date of outcome	Yes	Date of death (from any cause)
End of observation in data	Yes	Censoring: end of study period (all participants have available follow-up until 1-2-2026)
Day X following index date *(specify day)*	Yes	1126 days: Minimum of one year for every patient with a maximum of 3 years and one month
End of study period (specify date)	Yes	1-2-2026 Vital status up to February 2026 will be available in the NCR in April 2026 following the annual mortality update
End of exposure *(specify operational details, e.g. stockpiling algorithm, grace period)*	No	Follows treatment policy approach to the intercurrent event of discontinuation
Date of add to/switch from exposure *(specify algorithm)*	No	Follows treatment policy approach to the intercurrent event of therapy add on or switch
Other date *(specify)*	n/a	

[1] Follow up ends at the first occurrence of any of the selected criteria that end follow up

7.4.4 Covariates (confounding variables and effect modifiers, e.g. risk factors, comorbidities, comedications)

Covariates selected are important prognostics factors for the study outcome (death) as well as potentially associated with choosing either nivolumab + ipilimumab or pembrolizumab.

Table 9. Operational Definitions of Covariates

Characteristic	Details	Type of variable	Assessment window	Care Settings[1]	Code Type	Diagnosis Position	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
Age	In years defined by (time 0 – year of birth)/365	continuous	[0,0]	OP	n/a	n/a	All	n/a	n/a
Sex	Male, female	binary	n/a	OP	n/a	n/a	All	n/a	n/a
Histology	Adeno, squamous, or other	categorical	[-90,0]	OP	n/a	n/a	All	n/a	n/a
Liver meta	Presence of liver metastases at diagnosis	binary	[-90,0]	OP	n/a	n/a	All	n/a	n/a
CNS meta	Presence of brain metastases at diagnosis	binary	[-90,0]	OP	n/a	n/a	All	n/a	n/a
CNS meta treatment	Radiotherapy treatment for CNS metastases: none, whole brain, stereotactic	categorical	[-90,0]	OP	n/a	n/a	All	n/a	n/a
Bone meta	Presence of bone metastases at diagnosis	binary	[-90,0]	OP	n/a	n/a	All	n/a	n/a
Meta-count	Number of metastatic sites at diagnosis	numerical	[-90,0]	OP	n/a	n/a	All	n/a	n/a
KRAS mutation	Any KRAS mutation	binary	[-90,0]	OP	n/a	n/a	All	n/a	n/a
ECOG PS	0 vs 1	binary	[-90,0]	OP	n/a	n/a	All	n/a	n/a

OP = outpatient, n/a = not applicable, ECOG PS = Eastern Cooperative Oncology Group Performance Status

7.5 Core Emulation Table - Design Summary

Table 10. Comparison of Target Trial and Proposed Target Trial Emulation Design Elements (same for Estimand 1 and 2)

Attribute	Hypothetical Target Trial	Target Trial Emulation	Comment
Inclusion criteria	-histologically confirmed, non-oncogenic NSCLC -primary stage IIIb or IV disease -PD-L1 tumor expression <1% -ECOG PS 0 or 1 -age ≥18	-histologically confirmed, non-oncogenic NSCLC -primary stage IV disease -PD-L1 expression <1% -ECOG PS 0 or 1 -age ≥18 -start of index or reference drugs combined with chemotherapy (<90 days after diagnosis)	Stage IIIb patients were foreseen in the Hypothetical Target Trial (HTT) because the treatments under study are sometimes offered to patients with stage IIIb disease as well (off-label). Stage IIIb has a better prognosis than stage IV disease and requires a stratified analysis. In hindsight, incorporating that stratification has been overseen in the HTT design and because of knowing that the percentage of stage IIIb will be less than 10% in the study population together with the foreseen limited total sample size (n=200, and thus very few stage IIIb patients n=~20) we propose to focus the Target Trial Emulation (TTE) on stage IV solely. By doing so we reach a clinically relevant homogeneous population and also match with the reference pivotal 9LA clinical trial eligibility (stage IV only). The study population in the TTE is additionally restricted to those who actually initiated the study

			treatments, compared to the HTT in which all those who would be eligible to receive the interventions are included.
Exclusion criteria	A previous other malignancy	A previous other malignancy	
Setting	Multicenter specialist clinic	National cancer registry data (NCR)	Nationwide data from specialist clinics
Treatment conditions	Nivolumab + ipilimumab or pembrolizumab. Both in combination with platinum doublet chemotherapy	Nivolumab + ipilimumab or pembrolizumab in combination with chemotherapy. Chemotherapy can be started 28 days before immunotherapy.	Timing of chemotherapy reflects real world practice wherein one cycle of chemo is administered sometimes while awaiting molecular diagnostics results. If indeed non-oncogenic and PD/L1<1 immunotherapy is added. In the TTE all chemotherapies are allowed, which may introduce heterogeneity in treatments.
Treatment assignment	1:1 randomisation	Treatment received as recorded in the NCR. At the analysis stage, propensity score-based weighting to adjust for differences at baseline will be used.	Randomization cannot be directly emulated. Weighting of subjects in the statistical analysis will be used to balance confounders in absence of randomization; this analysis emulates randomised treatment assignment.
Follow up	Begins at randomisation. Ends at death, loss to follow up, withdrawal of patient consent or 5 years (end of study period)	Begins at day 1 following treatment initiation. Ends at death or at 3 years follow up (end of study period), or at end of data availability	Because of introduction of nivolumab + ipilimumab in clinical practice per 2022 actual follow-up time in the NCR data will be maximum 3 years. It is expected that 25% of patients will be alive at 3 years and therefore a larger proportion of patients will be administratively censored compared to the hypothetical trial.
Outcome	Time to death (of all causes)	Time to death (of all causes)	Date of death in the TTE will be identified using linkage with mortality records in the municipal basic administration (GBA).

			See table 7.6 for quality of the NCR and the feasibility assessment. It is assumed that subjects in the NCR who are not in the mortality registry are still alive.
Intercurrent events and strategies to handle them	Treatment discontinuation: treatment policy Treatment switching: treatment policy	Treatment discontinuation: treatment policy Treatment switching: treatment policy	Whilst treatment discontinuation and switching are not recorded in the NCR, the treatment policy strategy means this is not a limitation.
Loss to follow-up	Patients who fail to return for the required study visits and his/her health condition and vital status remains unknown despite multiple attempts to contact them	End of data availability (Feb 2026), which could occur before the intended 3 year follow up is completed	Except for the end of data availability, loss to follow up in the TTE would only happen if the patient moved to another country, which is assumed to be very rare in this population and therefore inconsequential. Date of death is always available through linkage with the GBA, also if the patients is transferred to another (Dutch) hospital. So, the data is assumed to be complete.

NSCLC= non-small cell lung cancer, PD-L1: Programmed death-ligand 1, ECOG PS = Eastern Cooperative Oncology Group Performance Status, NCR= Netherlands Cancer Registry

7.6 Data analysis

7.6.1 Analysis plan

Overview

The analyses are conducted within a TTE framework.

For Estimand 1, the main estimand supporting decision making, the primary causal effect summary measure is the HR of death from any cause in patients with metastatic non-oncogenic NSCLC with a PD-L1 expression of <1% receiving dual immunotherapy (nivolumab + ipilimumab)+ chemotherapy versus pembrolizumab + chemotherapy as first line of therapy, regardless of treatment discontinuation or treatment switch. The primary analysis will be an inverse probability of treatment weighted (IPTW) Cox proportional hazard model. Sensitivity analyses will be performed to investigate the robustness of the results to assumptions made in the primary analysis. For estimand 2 the difference in RMST will be estimated using the accelerated failure time (AFT) model with Weibull distribution.

Supplementary analyses, including diagnostic and descriptive assessments to support the main analysis, will be displayed to contextualise data and results from the primary and sensitivity analyses. These cover crude and IPTW-adjusted Kaplan Meier curves, weight and propensity score distributions, covariate balance before and after weighting, censoring patterns, proportional hazards diagnostics, Schoenfeld residuals analysis, and positivity checks.

Sensitivity analyses will assess robustness of the primary findings to key assumptions, i.e. a quantitative bias analysis (QBA) for missing data on a potential confounder (details in Section 7.6.5).

7.6.2 Primary (Estimand 1) Analysis

i. Objective

Estimate whether nivolumab + ipilimumab + chemotherapy improves OS in comparison to pembrolizumab + chemotherapy

ii. Exposure contrast

Nivolumab + ipilimumab + chemotherapy versus pembrolizumab + chemotherapy

iii. Outcome

OS

iv. Analytic software:

R

v. Handling of intercurrent events

Follow-up data are not excluded after the occurrence of intercurrent events.

Handling intercurrent events according to the following strategies:

- Treatment discontinuation (immunotherapy): apply treatment policy
- Treatment switch: apply treatment policy

vi. Outcome Modelling

A Cox proportional hazards model, weighted by IPTW, will be used to estimate the effect of initiating nivolumab + ipilimumab versus pembrolizumab on time to death of all causes.

Start of follow-up: date of treatment initiation for either the index or comparator treatment.

End of follow-up: date of death or censoring, whichever comes first

Censoring:

- Administrative censoring: End of study period in the absence of death (maximum 3 years follow up) and end of data availability
- Non-administrative censoring: none (NB: the life status is assumed to be 'alive' for subjects who do not appear in the mortality registry, see Table 10).

Model covariate: treatment group (nivolumab + ipilimumab versus pembrolizumab)

Assumptions of Cox Model

- Proportional hazards: the effect of treatment is assumed to be constant over time
- Non-informative censoring: censoring is assumed to be independent of the outcome, conditional on the treatment, survival up to the time of censoring, and indirectly baseline covariates used to estimate the treatment weights

Diagnostics for Cox Model

- Proportional hazards assessed using log(-log) survival plots or Schoenfeld residuals

vii. Confounding Adjustment

Inverse Probability of Treatment Weighting (IPTW)

Randomisation cannot be emulated, but exchangeability between treatment groups can be addressed using IPWT, under certain assumptions, at the analysis stage. To adjust for baseline confounding, IPTW will be used. Propensity scores, defined as the probability of initiating nivolumab + ipilimumab versus pembrolizumab, will be estimated using logistic regression. The model will include all covariates listed in Table 9.

Stabilized weights will be calculated by dividing the marginal probability of receiving the treatment actually received (i.e., the overall proportion treated in the study population) by the individual's estimated propensity score (i.e., the conditional probability of receiving their observed treatment). Weights will be truncated at the 1st and 99th percentiles to limit the influence of extreme values.

Weight truncation reduces the influence of individuals with highly improbable treatment assignments but does not resolve propensity score non-overlap. Therefore, if regions of the propensity score distribution show insufficient overlap, we plan to restrict analyses to the overlapping region (trimming) or apply overlap weights. This may lead to the exclusion of subjects from the analyses, which may mean the final study population deviates from the original intended study population.

Truncated stabilized IPTW weights will then be applied in the Cox proportional hazards model (weighted likelihood) to estimate the marginal treatment effect (nivolumab + ipilimumab versus pembrolizumab) on time to death of any cause. Standard errors will be estimated using robust (sandwich) variance estimators.

Assumptions underlying IPTW

- No residual unmeasured confounding: all relevant individuals characteristics are included in the propensity score model
- Positivity: each individual has a non-zero probability of receiving either treatment
- Correct model specification: the propensity score model is correctly specified [functional form, covariate inclusion]
- Consistency: each individual's potential outcome under the observed treatment equals their actual outcome

Diagnostics for IPTW

- Positivity check: Plot a histogram showing the distribution of inverse probability of treatment weights by treatment group, (before and after truncation)
- Covariate balance: check that characteristics are balanced across treatment groups after weighting
  - Evaluate standardized mean differences (SMDs): SMDs < 0.1 will be considered acceptable

viii. Missing Data handling

Because of the high quality of the data source (NCR) and assumptions made, we don’t expect missing data on the outcome (all records can be linked to the municipal basic administration (GBA) for the annual mortality update and it is assumed that no record of death means the subject is alive), neither on the covariates in the subjects included in the study population. Therefore, missing data imputation methods are not applicable.

ix. Subgroup analyses

- Histology squamous versus adenocarcinoma versus other
- With or without CNS metastases at baseline (and treated with radiotherapy yes/no)

7.6.3 Supplemental (Estimand 2) Analysis

i. Objective

Estimate whether there is a difference in RMST between nivolumab+ipilimumab and pembrolizumab

ii. Exposure contrast

Nivolumab+ipilimumab + chemotherapy versus mono pembrolizumab + chemotherapy

iii. Outcome

OS

iv. Software:

R

v. Handling of intercurrent events:

Same as 7.6.2

vi. Outcome Model

A Weibull Accelerated Failure Time (AFT) model will be used, followed by estimation of the RMST at fixed time points (1, 2 and 3 years after start of treatment).

Model covariate: treatment group (nivolumab + ipilimumab versus pembrolizumab)

Start of follow-up: date of treatment initiation for either the index or comparator treatment

End of follow-up: Date of death

Censoring:

- Administrative censoring: End of study period in the absence of death (maximum 3 years follow up) and end of data availability
- Non-administrative censoring: None

Model Assumptions

- Survival times follow a Weibull distribution
- Non-informative censoring (conditional on included covariates and survival up to time t)
- Log-linear relationship between covariates and log survival time

Diagnostics

- Log(-log(S(t))) vs log(t) should be linear Q-Q plot of residuals

To estimate the RMST at 1, 2 and 3 years from the Weibull AFT model, we first use the model to obtain the predicted survival curve for each treatment group. The RMST is then calculated as the average survival time up to a fixed time point, which corresponds to the area under the survival curve between time zero and the chosen time horizon (1, 2 or 3 years).

Steps to take:

- Fit the Weibull AFT model, which gives the shape and scale of the survival curve for each group.
- From this model, generate the predicted survival probability at each time.
- Integrate (i.e., add up) the survival probabilities from time 0 to 1 year and separately from time 0 to 2 years and from time 0 to 3 years. The result is the expected survival time lived within those windows.
- Compare the RMST values between treatment groups to obtain the difference in average survival time over 1, 2 and 3 years.

vii. Confounding adjustment

Same as 7.6.2; IPTW will be incorporated into the AFT model to obtain estimates of RMST at year 1, 2 and 3 years follow up.

viii. Missing data handling:

Same as 7.6.2

ix. Subgroup analyses:

Same as 7.6.2

7.6.4 Sensitivity analyses

Table 11. Sensitivity analyses – rationale, strengths and limitations

	Quantitative Bias Analysis (QBA) for missing data on a potential confounder (corticosteroid use)
Analysis method	A simple bias analysis formula will be applied to the adjusted HR to determine the possible impact of a specific potential unmeasured confounder (Brown et al 2024). By applying this analysis to the adjusted HR we are assuming that the confounding structure remains as specified, even after adjustment for other covariates.
Assumptions	It is assumed that baseline corticosteroid use could be a confounder of the treatment-outcome relationship, potentially affecting both treatment assignment and the outcome (death). Corticosteroid use can be a reason to prefer mono immunotherapy over dual immunotherapy. As corticosteroids are administered to manage symptoms, and thus are related to more symptomatic disease, these patients possibly have a worse prognosis. Clinical experience with the patient population for this study suggests corticosteroid prevalence amongst those treated with nivolumab + ipilimumab to be ~10%. The association between corticosteroid use and mortality was measured as 1.52 in a study of a similar patient population in Italy (Inno et al, 2025)
What is Being Varied?	1. Hypothesized exposure to corticosteroids at baseline: Scenario 1: Prevalence of exposure to corticosteroids is 10% for nivolumab + ipilimumab, 20% for pembrolizumab Scenario 2: Prevalence of exposure to corticosteroids is 10% for nivolumab + ipilimumab, 30% for pembrolizumab Scenario 3: Prevalence of exposure to corticosteroids is 20% for nivolumab + ipilimumab, 10% for pembrolizumab 2. Hypothesized association between corticosteroid use and mortality Scenario 1: Assumed hazard ratio of 1.5 (Inno et al 2025) Scenario 2: Assumed hazard ratio of 1.25

	Scenario 3: Assumed hazard ratio of 1.75
Why (Objective)	The data source does not contain information about co-morbidities and use of non-cancer medications. It is known that patients using corticosteroids for lung cancer symptom control have a worse prognosis (Inno et al. 2025) and the use of immunosuppressants (e.g. corticosteroids) either for co-morbities or lung cancer symptom control might be related to preferring single immunotherapy (pembrolizumab) over dual immunotherapy (niviolumab+ipilimumab). This sensitivity analysis estimates the potential impact on missing information about corticosteroids use on the study findings.
Strengths Compared to Primary Analysis	• Addressing unmeasured confounding • Improve internal validity • Quantifying uncertainty
Limitations Compared to Primary Analysis	• Subjectivity of setting parameters

7.6.5 Other supplemental analyses

Baseline characteristics will be presented stratified by treatment group, before and after weighting (Appendix Table 1). Categorical and binary variables will be summarized as counts and percentages, while the continuous variable of age at index will be reported using means and standard deviations or medians and interquartile ranges, as appropriate.

The crude HRs for Estimand 1 (Appendix Table 2) and Estimand 2 (Appendix Table 3) at pre-specified time points (1, 2 and 3 years) will be calculated for each treatment group, with 95% confidence intervals.

Kaplan–Meier methods will be used to compare the time-to-event distribution of death between patients treated with nivolumab + ipilimumab + chemotherapy and pembrolizumab + chemotherapy (Appendix Figure 3) Crude Kaplan–Meier cumulative incidence curves will be estimated separately for patients of both groups. The cumulative incidence (absolute risk) of death will be estimated from the Kaplan–Meier curves at pre-specified time points of 1, 2 and 3 years for each treatment group, together with 95% confidence intervals. IPTW Kaplan–Meier curves will also be estimated. Time will be measured from the index date until the death or censoring, whatever comes first.

We will also conduct descriptive analyses to characterize censoring patterns overall and across treatment groups (Appendix Table 5). This will include median (IQR) time to censoring overall and according to the reason for censoring. This will be estimated separately for the overall study population and by treatment group.

7.6.6 Core Emulation Table -Estimation Summary

Table 12. Core Emulation Table: Estimation Summary

Estimand 1

Attribute	Hypothetical Target Trial	Target Trial Emulation	Comment
Attribute analysis Method	Cox proportional hazards model to estimate the hazard ratio for time to death, with treatment group as the only covariate.	Cox proportional-hazards model to estimate the hazard ratio for time to death, with treatment group as the only covariate and using inverse probability of treatment weighting (IPTW) based on propensity score estimation.	IPTW used to “emulate” randomization in observational data. In the event that subjects are excluded due to non-overlapping PS distributions between treatment groups, this may lead to selection bias due to a change in the originally specified study population which would not occur in the target trial.
Missing Data Assumptions and Methods	No missing data because of prospective CRF filling	Population selection and analyses based on variables with no missing entries in the data source.	n.a.
Statistical Model Assumptions	Proportional hazards assumption for Cox model, assessed using model diagnostics such as Schoenfeld residuals and log(-log) survival plots	Proportional hazards assumption for Cox model, assessed using model diagnostics such as Schoenfeld residuals and log(-log) survival plots IPTW Assumptions: no unmeasured confounding, positivity, correct model specification, consistency	Diagnostics confirm appropriateness of the Cox model; potential violations addressed in estimand 2 (RMST) Some assumptions for IPTW are difficult to verify (e.g., unmeasured confounding).
Sensitivity Analyses	Sensitivity analysis under the Censoring Not At Random (CNAR) assumption will be carried out	QBA assesses potential bias due to unmeasured confounding related to baseline corticosteroid use, which may influence both the treatment assignment and the outcome	CNAR is not applicable anymore in the TTE because of administrative censoring only. In the HTT, baseline characteristics of patients who were lost to follow up could be analysed. We have added sensitivity analyses to assess potential unmeasured confounding, which is n.a. for a RCT.

QBA: quantitative bias analysis, RCT: randomised controlled trial

Estimand 2

Attribute	Hypothetical Target Trial	Target Trial Emulation	Comment
Attribute analysis Method	Accelerated failure time (AFT) model with Weibull distribution	AFT model with Weibull distribution	
Missing Data Assumptions and Methods	No missing data assumed	Same assumption	
Statistical Model Assumptions	Survival times follow Weibull distribution and non-informative censoring. Assumption of log-linear relationship between covariates and log-linear survival time, assessed using normally distributed residuals in Q-Q plot confirm appropriateness of AFT model with Weibull distribution	Survival times follow Weibull distribution and non-informative censoring. Assumption of log-linear relationship between covariates and log-linear survival time, assessed using normally distributed residuals in Q-Q plot confirm appropriateness of AFT model with Weibull distribution	
Sensitivity Analyses	Sensitivity analysis under the Censoring Not At Random assumption will be carried out	n.a.	Sensitivity analyses only used for primary Cox model analysis, not for AFT-based Estimand 2

7.7 Data sources

7.7.1 Data source

Selection and feasibility:

The NCR collects data with the primary aim of enabling several stakeholders to reflect on and improve oncological and palliative care (https://iknl.nl/en/about-iknl). The NCR compiles clinical data (i.e., hospital inpatient and outpatient data) of all individuals newly diagnosed with cancer in the Netherlands. A record in the NCR is triggered by biopsies from the national pathology database (PALGA) or the registration of a cancer diagnosis in the national registration of hospital care (Landelijke Basisregistratie Ziekenhuiszorg; LBZ).

A group of data managers register data of each cancer patient in the NCR. The relevant data for each tumour (and the corresponding patient and treatment) is registered after a set amount of time, typically 6-12 months after the incidence date. The vital status of patients is updated once a year through linkage with the municipal basic administration (Gemeentelijke Basis Administratie, GBA). The data managers are often specialized in a specific type of cancer: lung cancer in this case study. The data managers are well-trained and are using coding manuals. The data entry application performs checks on the data that is entered, automatic checks are done on the database, as well as manual checks of random samples. A group of data managers is responsible for data quality and researchers in the organization can flag potential quality issues.

The NCR will be used to conduct this target trial emulation because it is the largest data set available to the study consortium on the required population, critical data elements are readily available and fairly reliable. Data is expected to be received within 2 months after the data request. A data recency of approximately 12 months is reasonable for the research question and updates of death data can be delivered upon request (IKNL/GBA). A sample size of about 200 patients for nivolumab + ipilimumab and over 1000 for pembrolizumab is available for the years 2022-2024 (extrapolated from Verschueren et al. 2024 and personal communication with 12 clinics within the NVALT network). Information on the quality of the data source is collected to inform the feasibility assessment of this TTE. The data quality report is displayed in Appendix 2.

Strengths of data source(s):

The strength of the NCR lies in the coverage of the whole of the Netherlands and of critical variables. For instance, information on variables such as lung cancer diagnosis, age, and date of death is available for 100% of the patients (not moved abroad). The likelihood of moving abroad is considered extremely unlikely in this patient population and therefore ignorable. The death date for this study will be up to date for Q1 2026 and at that time registration for the year 2024 will be completed as well. Treatment data is available for the first-line treatment.

Limitations with potential impact in the study results:

Potential limitations of using the NCR as a data source are the following: First, information on second line (2L) treatments is only available in the NCR if started at time of registration (6-12 months following diagnosis), so is incomplete and not feasible to use in this study. This may prevent the correct identification of treatment switch. But this is not relevant in this case, because this intercurrent event has been handled with a treatment policy strategy.

Data quality: reliability and relevance:

Data source provenance can be accessed through the following links:

- https://iknl.nl/en/about-iknl
- DARWIN:”2024_IKNL_xlsx (for the EMA catalogue)

- DARWIN.pdf (for the onboarding as a data partner)
- 20240918-itemset-long.pdf
- https://iknl.nl/en/ncr-data

In addition, Table 13 displays the data quality of the data source. It is concluded that the study is feasible. Although some data elements in the NCR have missing values (e.g. staging and ECOG performance status overall are 6 and 15% missing, respectively) this will not affect the conduct of the study because these items are applied as inclusion criteria for the study and subsequently, this won’t be missing in the analyses. Also, there is no reason to assume that missingness will be differential for the index or comparator treatment and the percentage of missing is expected to be even lower in patients that started first line systemic treatment What remains is the lag time in registration of patients in the NCR 6-12 months after diagnosis. This policy prevents the inclusion of very recently diagnosed patients (e.g. in 2025) but also that is not considered a concern because of the upside of a median longer follow-up time of the cohort.

Table 13. Data quality analyses data source (NKR) (WP1)

Sample size estimation from the hypothetical trial protocol	Feasibility assessment *(yes/yes, with limitations/no)*	Rationale for the feasibility assessment	Limitations identified during the feasibility assessment and categorisation	Description of potential impact of the identified limitations on the study results
The estimated sample consists of approximately 244 participants (1:1 ratio of nivolumab + ipilimumab + chemotherapy and pembrolizumab + chemotherapy, thus 122 per group). NCR includes 5,000 patients with stage IV NSCLC and 1,000 with stage III NSCLC. Since 2021, ~100 patients have been treated with nivolumab + ipilimumab, compared to 3,000 patients receiving pembrolizumab. The sample size for pembrolizumab is adequate, while the size for nivolumab + ipilimumab could be limited.	Yes	Elements with high criticality are in their majority available, but some of them have limitations. The time elapsed from when a user requests the data to when they actually receive it is 2 months. Data recency is ~12 months before extraction, reasonably enough for the research question. Sample size is achievable.	·Potentially major: The median length of follow-up per patient is approximately 9 months ·Potentially major: 15% ECOG missing ·Potentially major: Difficult to detect previous systemic anti-cancer treatment, autoimmune disease or severe infectious disease (e.g., HIV) ·Minor: Only prescription of first line of treatment (if stage changes a treatment is considered as a new first line); it won't be possible to differentiate discontinuation from switch. ·Minor: TNM reliable, but 6% have missing the specific stage ·Minor: Some cancer patients do not have a biopsy and pathology, but might be picked by diagnostic code ·Minor: Data is registered 6-12 months after diagnosis so there is a lag	Although the median follow-up time in the NCR is 9 months, this includes patients with all types of cancer with different survival durations. However, this variation is likely non-differential, meaning it is not expected to bias the results in favour of or against any particular cancer group. If the patients included in the study have a longer survival time, the registry will allow for the follow-up required by protocol. Missing ECOG data may prevent us from including some subjects. Previous cancer or anti-cancer treatments can be detected from the patient's previous records in the registry. The history of autoimmune disease or severe infections cannot be detected, but we believe that this fact is already implicit in the physician's decision to treat the patient. This should be taken into account in the interpretation of the results.

Table 14. Metadata about data sources and software

	Data
Data Source(s):	NCR
Study Period:	1-1-2022 to 01-02-2026
Eligible Cohort Entry Period:	1-1-2022 to 31-12-2024
Data Version (or date of last update):	1-2-2026 for vital status
Data sampling/extraction criteria:	Patients with primary, non-oncogenic, histologically confirmed, stage IV NSCLC and a PD-L1 expression of <1% treated with either the index or reference treatment
Type(s) of data:	Hospital data from outpatient clinic setting, administration data from individual patients' electronic health records and death records
Data linkage:	NCR is linked to the GBA for data regarding deaths. The last linkage will be performed at 1-02-2026
Conversion to CDM*:	In the NCR data is registered in OMOP CDM and will be converted to the concePTION CDM by WP2
Software for data management:	RANK, developed and maintained by the in-house Software Development Department of IKNL. The changes in the database are loaded into a datawarehouse (DWH) every night.

*CDM = Common Data Model, NCR = Netherlands Cancer Registry, GBA = municipal basic administration

Further meta data are available in the EMA catalogue of data sources: Netherlands Cancer Registry | HMA-EMA Catalogues of real-world data sources and studies

7.8. Data management

The study will be conducted in a distributed manner using the UMCU, ARS Toscana and VAC4EU tools, procedures, and pipeline. Figure 2 specifies the data sets (D) and transformation processes (T), programming follows this pipeline, with involvement of different types of experts.

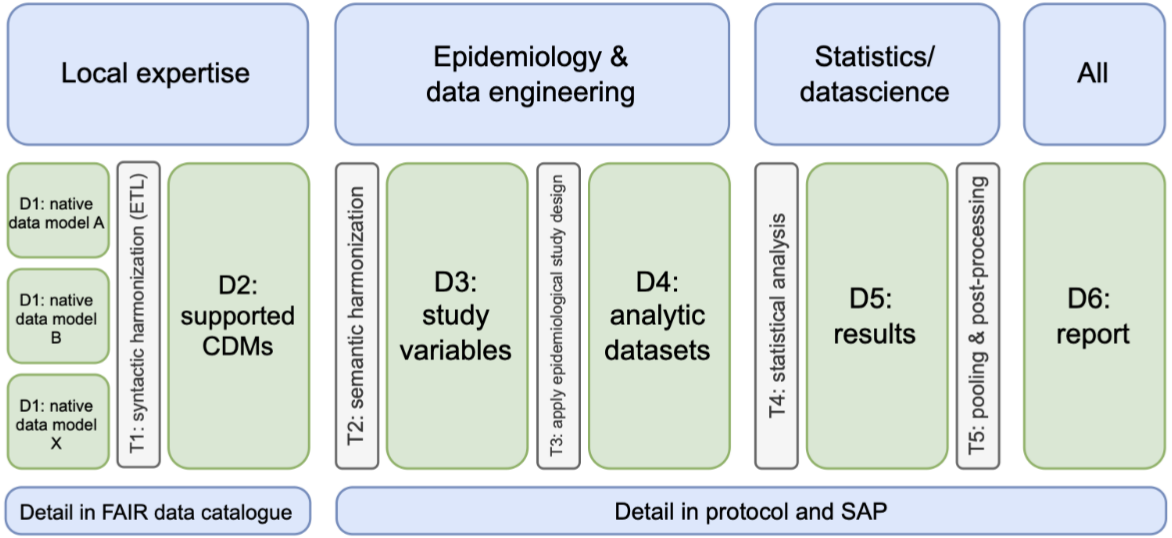


Figure 2. Data Management from the data transformation perspective

D1: Original data can be in any native format

The RWD-RWE pipeline used by VAC4EU starts with data banks that are controlled by the Data Expert and Access Partner (DEAP) and can be in any format. Data stays local and does not leave the secure environments of the DEAP. The ETL (extract, transform, load, see below for more details under 'T1') design is shared in a searchable FAIR VAC4EU catalogue. The VAC4EU FAIR Molgenis data catalogue is a meta-data management tool designed to contain searchable meta-data describing organisations that can provide access to specific data sources.

T1: Syntactic harmonisation (ETL)

T1: Syntactic harmonisation is conducted through an extraction, transformation, and loading (ETL) process of native data into the ConcePTION common data model (CDM) (see section 'D2: Common data model'). To harmonise the structure of the data sets stored and maintained by each data partner, a shared syntactic foundation is used. The ETL process has various structured steps as described by Thurin et al [18]:

- The DEAP is asked to share the data dictionaries of their data banks (selected tables and variable names/structure)
- Metadata (descriptive data about the data sources and databanks) & data dictionaries, are uploaded in FAIR data catalogue (Molgenis).

D2: Common data model

For this project, the CDM (D2) is the OMOP common data model. The CDM version that is used is v5.4, which is available as an open-source CDM here: https://ohdsi.github.io/CommonDataModel/cdm54.html.

T2: Semantic harmonisation

During the T2 step, many data transformations occur related to the completion of missing features in the data. Based on the relevant diagnostic medical codes and keywords, as well as other relevant concepts (e.g., medications), one or more phenotype algorithms are constructed (typically one sensitive, or broad, algorithm and one specific, or narrow, algorithm) to operationalise the identification and measurement of each event. In this step we conduct time anchoring (observation periods, look back periods), clean the data such as sort on record level, aggregate across multiple records, and combine concepts for implantation of algorithms, and rule-based creation of study variables.

In this phase of the creation of study variables, semantic mapping is conducted. This semantic mapping across different vocabularies is conducted as part of the R-study script using different functionalities. To reconcile differences between different terminologies and native data availability, machine-readable code lists are used that comprise the terminologies that are used in the network (e.g. ICD-9, ICD10, SNOMED, ICPC and DEAP specific adaptations). This is combined with the BRIDGE metadata file that defines risk windows, look-back periods, and algorithms for each study variable [19].

D3: Study variables

D3 datasets are interim data sets with information on study variables for each study participant, the unit may be a person, a medicine, or episode of time. The design of these datasets is described in codebooks. Examples of D3 datasets are the outputs of the ConcePTION pregnancy algorithm (https://github.com/IMI-ConcePTION/ConceptionTools/wiki#conception-pregnancy-algorithm), and outputs of functions that define smoking. Multiple functions/packages exist within the VAC4EU, for different study variables.

T3: Application of epidemiological design

In the T3 step epidemiological designs are applied such as sampling, matching (on specific variables and/or propensity scores), and selection based on inclusion and exclusion criteria using the study variables in the D3 datasets. The designs will be implemented for the various study objectives using R-scripts, and these may use the existing functions (R-cran) or functions that have been developed in the VAC4EU community (e.g. matching).

D4: Analytical data set

D4 is an analytical dataset, and multiple D4 data sets may be produced based on the objectives of the study. The format is described initially in a code book for communication between programmers and statisticians.

T4: Statistical analysis

This step in the data transformation pipeline will produce statistical estimates such as descriptives (counts, percentages), distributions (mean, percentiles), rates (prevalence, incidence), regression coefficients, or other relevant estimates. This will be conducted using R.

D5: Results

D5 is the set of estimates, tables or aggregate data that is transferred from the DEAP to the Digital Research Environment (DRE). The aggregated results produced by these scripts at the DEAP's site will be uploaded to the UMCU DRE for post-processing, pooling and visualisation (Figure 2). The DRE is a cloud-based, globally available research environment where data are stored and organised securely and where researchers can collaborate. The DRE is made available through UMCU. The DRE applies double authentication where researchers can collaborate using data that are stored and organised securely [20]. UMCU is responsible for data processing and data security.

All researchers who need access to the DRE will be granted access to study-specific secure workspaces by UMCU. Access to the workspaces will be possible only after double authentication using an identification code and password together with the user's mobile phone for authentication.

Uploading files will be possible for all researchers with access to the workspace within the DRE. Downloading of files will be possible only after requesting and receiving permission from a workspace member with an "owner" role, who will be a UMCU team member.

T5: Post-processing/pooling

In this step, the result from the DEAP is pooled and converted into tables and figures for reporting.

7.9. Quality control

For quality control of the data instance from the Dutch Cancer Registry, we will use the OHDSI DataQualityDashboard, available at https://github.com/ohdsi/DataQualityDashboard. This tool performs more than 3,000 standardized checks on a populated OMOP Common Data Model (CDM) instance. Its objective is to evaluate the quality of observational data in a systematic, reproducible, and transparent manner.
The quality checks are structured according to the Kahn Framework, which defines categories and contexts representing different strategies for assessing data quality. The DataQualityDashboard implements 24 core check types within this framework, which are systematically applied across all relevant tables and fields of the OMOP CDM (see https://doi.org/10.1093/jamia/ocab132).

Code Quality

These coding practices define how the TARGET programming team collaborates to write clean, reliable, and reproducible code for the VAC4EU Real-World Evidence (RWE) Analytical Pipeline. They aim to ensure clarity, consistency, and maintainability across all case studies within the project.

Coding conventions

To ensure clarity, consistency, and maintainability across the project, the following conventions will be applied to all codebases within the project:

- Consistent style: Code follows a consistent and readable style (see the tidy verse style guide for R).
- Meaningful names: Use clear, descriptive names for variables, functions, and files to convey their purpose.
- Modular code: Break down code into small, reusable functions where possible.
- No hardcoded paths: Use configuration files or relative paths to ensure portability.

Following these conventions makes the code easier to understand, test, and reuse across case studies and teams.

Documenting Code

Code documentation is used to promote good coding practices and ensure our work is understandable, maintainable, and reproducible. To achieve this, we will:

- Use descriptive comments that explain the purpose and rationale behind code sections, focusing on why something is done, not just what.
- Clearly document function inputs, outputs, and side effects, using standardized formats (e.g., roxygen2 in R) where appropriate and supported.
- Write meaningful variable and function names to make the code as self-explanatory as possible.

Version Control

We use Git and GitHub to manage version control. These tools support good coding practices by enabling collaboration, tracking changes, accessing a project's history, and ensuring code quality through review and documentation.

A dedicated GitHub organisation has been created for the project (https://github.com/target-roc19). Each case study is managed in its own repository within this organisation. Repositories are structured consistently across case studies, to reinforce modularity. Access to repositories is controlled through teams.

During development, all repositories remain private to ensure confidentiality. Once the project is finalised, relevant repositories will be made public and assigned a digital object identifier (DOI) via Zenodo to support transparency, reproducibility, and reuse by the wider research community.

To maintain code quality and clarity, we follow the git and GitHub guidelines below.

- Always use pull requests (PRs): never push directly to the main branch.
- Open an issue before creating a new branch. Ideally, one PR resolves one issue to keep changes focused and reviewable.
- Every PR must be reviewed by at least one other person before merging.
- The PR author merges the PR after it has been reviewed and approved.
- Write clear, descriptive commit messages.
- Write informative PR descriptions, including:
    - A concise title
    - Links to related issues
    - A summary of the changes

Continuous Integration

Continuous Integration (CI) is set up to automatically check code quality and run tests whenever changes are pushed to the repository or submitted through a pull request (PR). The CI workflow ensures that the package adheres to predefined style guidelines and that all automated tests pass before changes are merged.

Coding Template

Every case study follows the general coding template used across all code in the TARGET project. The folder structure is organised as follows:

```
case-study-template
|____data
| |____D2_cdm
| |____D3_study_variables
| |____D4_analytic_datasets
| |____D5_results
```

```
| |____D6_report
|____docs
|____logs
|____run
|____tests
|____transformations
| |____T2_semantic_harmonization
| |____T3_study_design
| |____T4_statistical_analysis
| |____T5_processing_results
|____CHANGELOG.md
|____LICENSE
|____README.md
```

Project Data Structure and Storage

The data folder follows the Real-World Evidence pipeline structure. Data conforming to the common data model is stored in the D2_cdm folder.

Results from transformations T2, T3, T4, and T5 are saved in the respective folders:

- D3_study_variables
- D4_analytic_datasets
- D5_results
- D6_report

Each dataset is associated with a codebook, explained in more detail below.

All data remain securely stored on the DEAPs servers and are never transferred externally. For testing purposes, dummy datasets are created. These fall into two categories:

- Unit test data: Small, predefined input and output pairs used to test individual transformation steps. These are stored in the tests folder, not in data, and can support automated testing.
- Pipeline test data: Larger, more complex dummy datasets used to test whether the full pipeline runs as expected. These may be included in the repository only if they remain below GitHub’s 100 MiB file size limit and will otherwise be shared via SharePoint.

Logging System

When the pipeline is executed, log files are saved in the logs folder. These logs are especially helpful when running the code in the DEAPs environment, as they help trace and diagnose potential errors. We recommend using the logger R package to handle logging throughout the pipeline. A sample logging setup can be found in the logger. R script located at the root of the project directory.

Executing the Analytical Pipeline

The run folder contains scripts used to execute each transformation step in the pipeline.

- A central script, run_pipeline.R, orchestrates the full pipeline from start to finish.
- Subscripts (e.g., run_T2.R or similar) are available to run individual transformation steps separately.

Typically, the run_pipeline.R script is the main entry point used by a DEAP to execute the full pipeline. Before running it in the DEAP environment, the pipeline may need to be adapted to local settings. This can be done using a configuration file that defines variables required to tailor the pipeline to a specific DEAP. Please note that configuration files should not include sensitive information.

Such a file might include variables like:

- The name of the DEAP
- The path to the local data instance
- The path to any required external resources

Testing and Quality Assurance

The tests folder contains scripts to test the analytical pipeline. Tests will be used to ensure code behaves as expected and remains stable over time. By systematically checking inputs, outputs, and edge cases, tests help catch errors early and make future changes safer. We use the test that R package to structure and run unit tests.

Continuous integration (CI) is used to automate testing. With CI, tests are automatically run each time code is pushed to the repository (e.g., via GitHub Actions). This helps identify issues immediately, ensures that new changes do not break existing functionality, and supports better collaboration by enforcing consistent code quality across contributors.

Modular Data Transformation Workflow

The transformations folder follows the Real-World Evidence pipeline structure. It contains the source code for all transformation steps, which is typically written in R. Each subfolder corresponds to a specific step in the pipeline (e.g., T2_semantic_harmonization, T3_study_design, T4_statistical_analysis, T5_processing_results) and includes the relevant scripts and helper functions for that step.

During the T2 step, a database is usually created (e.g., using DuckDB). This database can be queried using SQL, and it is recommended that all SQL queries be saved as clearly named, standalone SQL script files to ensure readability and reusability.

The purpose of the transformations folder is to structure and modularise the processing logic, making it easier to maintain, test, and reuse across different case studies. By organising code by transformation step, teams can work in parallel, increasing efficiency.

Changelog

A changelog will be kept for all notable changes in the project. Changelogs help track the evolution of the project over time, making it easier for collaborators to understand what has changed between versions. We follow the structure and best practices outlined in Keep a Changelog.

Codebooks

Before developing code, codebooks are created to describe each dataset (D) within the pipeline. A codebook is a comprehensive document that outlines the structure, contents, and metadata of a dataset. It serves as a detailed reference guide for anyone working with the data and plays a crucial role in guiding the development of the analytical pipeline by clearly defining both the inputs and expected outputs.

All codebooks are summarized in a central index file, which provides a high-level overview of the pipeline's structure. For each codebook, the index file includes:

- A brief description of its purpose,
- A list of the scripts used to generate the corresponding dataset,
- A description of the input datasets and input parameters required.

The datasets D2, D3, D4, and D5 are typically subdivided into multiple smaller transformation steps, each detailed within their respective codebooks. These smaller transformation steps ensure that each part of the pipeline is clearly scoped and well-documented.

In addition to supporting development, codebooks help ensure quality control by making transformation logic transparent and verifiable, and they enhance reproducibility by documenting exactly how data is structured and used throughout the analytical pipeline.

Deployment

The analytical pipeline is delivered to DEAPs as a GitHub release, tagged with a version number. Versioning follows the format: YYYYYMMDD.XX, where the date indicates the release date and XX denotes the sub-version or revision number.

Any deployment issues can be reported via the GitHub repository using the issues feature, where the programming team responsible for the R code will collaborate with the local DEAP to resolve them as needed.

Reproducibility

It is recommended to locally use the renv R package to maintain the R version and version of packages for reproducibility purposes.

At this time, however, using renv reliably across different systems and environments remains challenging. For this reason, we currently recommend its use only in local development setups.

We are actively monitoring developments in the R ecosystem related to cross-platform reproducibility. As soon as a more stable and portable solution becomes available, we will revisit this guidance and promote broader adoption.

Open Source Licensing

The code will be made available under an open source license.

README Guidelines

Each case study repository includes a README that covers the following points:

- Project Overview: brief summary of the study goals and key research questions.
- Background: context and rationale for the study.
- Repository Structure: Outline of main folders and their contents.
- Data Overview: Description of data sources, formats, and data privacy considerations.
- How to Run: Instructions for running the pipeline and key scripts, plus where outputs are saved.
- Testing: How to run tests to verify code functionality.
- Contributing: Guidelines for code contributions and issue tracking.
- License: Information about the code license.
- Contact: Who to reach out to for help or questions.

7.10. Study size

All patients in the NCR meeting the in- and exclusion criteria will be included in the study dataset, although there is a theoretical chance that some may be excluded from the final analysis due to non-overlapping propensity scores. It is already known that the number of patients who started the index treatment will be much smaller than the number of patients who started with the comparator treatment. The reason for this is that not all clinics in the Netherlands allowed to administer nivolumab + ipilimumab following reimbursement in 2022 as a national policy required a minimum number of patients on the particular treatment yearly. Nevertheless, conducting the study is of interest as a use case for the TARGET EU project. The anticipated sample size (extrapolation from Verschueren et al. 2024 and personal communication with 12 clinics within the NVALT network) is around 200 patients treated with nivolumab + ipilimumab and over 1000 patients treated with pembrolizumab.

A limitation of this study is the relative short follow up times to compare differences in OS after 3 years as nivolumab + ipilimumab is only reimbursement since 2022. Stage IV NSCLC patients have an overall risk of death within 2 years of 65% and within 3 years of 75% (Carbone et al. 2025), so many events will already occur within a shorter follow up time. Nevertheless, we anticipate a proportion of administratively censored patients amongst those diagnosed in 2023 and 2024.

Estimation

The level of precision for the estimated HR that is achievable with a fixed sample size can be estimated by the expected width of the 95% confidence interval for the effect estimate.

Assumptions: 200 patients receive nivolumab + ipilimumab vs 1000 patients receive comparator
Risk of death over 3 years for the comparator treatment = 0.75 (Carbone et al. 2025) (varied to 0.65 in scenario 2 to determine the impact of a lower event rate)
Hazard ratio to detect = 0.65

Scenario 1

HR=0.65
Estimate of expected number of outcomes if risk of death over 3 years in comparator group is 75% and with 200 patients on nivolumab + ipilimumab and 1000 on comparator

Estimate number of outcomes to be (200*0.75*0.65) =98 for nivolumab + ipilimumab group + (1000*0.75) =750 for comparator = 98 + 750 = 848
SE = sqrt (1/98 + 1/750) = sqrt(0.011537) = 0.1074
ln(HR) = ln(0.65) = -0.43078
Margin of error = 1.96 x SE = 1.96 x 0.1074 = 0.2105

lower bound = -0.43078 – 0.2105 = -0.64128
upper bound = -0.43078 + 0.2105 = -0.22028
lower CI = exp(-0.64128) = 0.53
upper CI = exp(-0.22028) = 0.80

Scenario 2
HR=0.65
Estimate of expected number of outcomes if risk of death over 3 years in comparator group is 65% and with 200 patients on nivolumab + ipilimumab and 1000 on comparator

Estimate number of outcomes to be (200*0.65*0.65) = 85 for nivolumab + ipilimumab group + (1000*0.65) = 650 for comparator = 85 + 650 = 735
SE = sqrt (1/85 + 1/650) = sqrt(0.013303) = 0.1153
ln(HR) = ln(0.65) = -0.43078
Margin of error = 1.96 x SE = 1.96 x 0.1153 = 0.2260
lower bound = -0.43078 – 0.2260 = -0.65678
upper bound = -0.43078 + 0.2260 = -0.20478
lower CI = exp(-0.65678) = 0.52
upper CI = exp(-0.20478) = 0.81

Table 15. Power and sample size

Scenario	Number of Events	Hazard Ratio (HR)	log(HR)	Standard Error (SE)	Margin of Error	Lower CI (HR)	Upper CI (HR)	% precision
Scenario 1	848	0.65	-0.43078	0.1074	0.2105	0.53	0.80	23%
Scenario 2	735	0.65	-0.43078	0.1153	0.2260	0.52	0.81	25%

CI= confidence interval

8. Limitation of the methods

Follow up: The most important limitation of the study is the relatively short follow-up times at the time of analyses in Q1 2026. Follow up time was 5 years in the hypothetical target trial and 3 years in this target trial emulation. This impairs the possibility to detect differences in long-term survival. Recently available long term (up to six years) follow-up data from clinical trials conducted with the index and comparator treatment suggest that benefit of dual immunotherapy in no PD-L1 expressing tumours is most evident in the tail of the survival curve where with nivolumab + ipilimumab the curve becomes flat most likely caused by a group of patients with a durable complete response. The present study could be followed up by an additional analysis in 2028 (out of scope of this EMA project).

In addition, in the emulation follow up begins at the date of first treatment rather than at randomisation. This aligns the start of follow up with treatment initiation but differs from the conditions of the HTT. In our study, treatment allocation reflects real-world prescribing decisions based on patient characteristics rather than random assignment, which may introduce confounding.

Residual Confounding: In addition, the data source used for the emulation does not allow adjustment for unmeasured confounding (e.g. smoking history, comorbidities and co-medications) that could lead to residual confounding, which would not be the case in a prospective randomised clinical trial. QBA is performed to assess potential confounding by corticosteroids.

Stage of disease: Inclusion of stage IIIb patients was foreseen in the HTT because the index treatment nivolumab + ipilimumab is sometimes off-label offered to patients with stage IIIb disease as well. Stage IIIb has a better prognosis than stage IV disease and requires a stratified analysis. Because the percentage of stage IIIb will be less than 10% in the study population together with the foreseen limited total sample size (n~200 index treatment patients, and thus very few stage IIIb patients n~20) the focus the TTE is on stage IV solely. By doing so we reach a homogeneous population and also match with the reference pivotal 9LA clinical trial eligibility (stage IV only). The limitation is that the results will not be generalisable to the whole patient population receiving the index or comparator treatment.

Treatment exposure: in the TTE treatment initiation is based on the treating physician's and patient's decision and is identified using codes instead of the treatment protocol being prespecified and treatment randomly allocated as in the HTT. Moreover, the number of chemotherapies is not defined in the TTE, as this information is not included in the data source. This represents a departure from the controlled treatment procedure of the HTT likely to result in undertreatment in the TTE in comparison to the HTT and might introduce exposure misclassification. This would be most likely to lead to an underestimation of any treatment effect in comparison to the HTT and the reference trial. As we cannot correct for number of chemotherapies in the TTE, it will remain unknown whether it is different between treatment groups.

In addition, in the TTE chemotherapy can be started 28 days before the start of immunotherapy, reflecting real world practice, and which is not included as an option (run-in period) in the current HTT.

Study population: might be different in the TTE due to an additional inclusion criterion around initiating treatment with the therapies investigated, whereas in the HTT eligibility is assessed prior to initiation of treatment (and randomisation). However, it is assumed that between randomisation and treatment initiation the risk of drop out is very little, and if any that would be non-differential. **Randomisation**: cannot be emulated. Instead, IPTW is used to approximate exchangeability. While IPTW balances measured baseline covariates, it cannot account for unmeasured confounding. A key limitation of emulating a target trial using real-world data is that it may not be possible to fully reproduce all features of the target trial and data are captured in different settings. These limitations are inherent to the TTE approach.

Unlike the HTT, which relied on randomisation to achieve balance of baseline participant characteristics, the emulation employs IPTW in Cox and AFT models. This approach requires unverifiable assumptions: no unmeasured confounding, correct model specification, and positivity. Departures from these assumptions could bias results. In addition, the Cox proportional hazards model assumes that the relative treatment effect is constant over time, which may not be appropriate. This assumption is explicitly tested using Schoenfeld residuals and log-log survival plots.

Potential exclusion: some subgroups of patients may be very unlikely to receive one of the treatments, leading to limited overlap in covariate distributions between groups. Subjects might be excluded at the analysis stage due to non-overlapping propensity scores. This restricts the analysis to a subset of patients, meaning the estimated effects may not generalise to the entire eligible population. The propensity score distribution is examined for evidence of positivity violations by assessing the overlap of propensity scores between treatment groups.

Time of death: is in the TTE identified by linkage to the GBA. As follow up of all patients will end at 1-2-2026, the endpoint data are complete and no bias is expected in the main outcome due to this deviation.

Missing data: are assumed to occur more often in HTT compared to TTE. In the TTE no missingness is expected in the outcome (death of any cause through linkage with GBA), nor in covariates selected for the study (see Table 9 in section 7.4.4). Regarding the ECOG being missing in the NCR in some patients, this is not expected to have impact on the foreseen analyses because ECOG performance status is an inclusion criterium and thus cannot be missing to be able to enter the study. Regarding selection of patients for entering the study population, it is expected the ECOG being missing is very rare in patients who start either index or reference treatment because ECOG 0-1 is a condition for reimbursement and therefore should have been assessed (see also 7.6.1. data quality and text with Table 13) and for the rare case of missing ECOG values, there is no reason to expect that this will be different between treatment arms, because both treatments share similar treatment indications (and thus patients characteristics and clinical work-up) thus will not introduce a risk of selection bias.

CNAR: is not applicable in the TTE because of administrative censoring only and will not introduce bias.

9. Protection of human subjects

This is a non-interventional study using secondary data collection and does not pose any risks for individuals. The data source research partner holds data contracts with all Dutch hospitals and applies an opt-out procedure for patients who are not willing to be included in the NCR. This study involves data that exists in a pseudonymized structured format and contains minimal identifying information (e.g. age at diagnosis instead of birth date, survival times instead of date of diagnosis and date of death). Data protection and privacy regulations will be observed in collecting, forwarding, processing, and storing data from study participants. All parties will comply with all applicable laws, including laws regarding the implementation of organisational and technical measures to ensure the protection of data.

As this study does not involve data subject to privacy laws according to applicable legal requirements, obtaining informed consent from individuals is not required.

10. Reporting of adverse events

Adverse events are not incorporated in the NCR, so not applicable for this study.

11. References

Brown JP et al 2024 doi: 10.1002/pds.70026. PMID: 39375940.

Gandhi 2018 10.1056/NEJMoa1801005

Inno 2025 10.1007/s00262-025-04125-w

Matsumoto 2024 10.1007/s00262-023-03583-4

Paz-Ares 2018 10.1056/NEJMoa1810865

Reck 2019 10.2217/fon-2019-0031

Reck 2024 10.1016/j.ejca.2024.114296

Shiraishi 2024 10.1016/S2213-2600(24)00185-1

Smit 2021 10.24078/onco.2021.10.128011

Verschueren 2025 10.1016/j.cllc.2025.01.009